Міністерство освіти та науки України
Національна академія наук України
Люблінський відділ Польської Академії Наук
Представництво „Польська академія наук” у Києві
Харківський національний університет радіоелектроніки
Харківський національний університет імені В.Н.Каразіна
AGH науково-технологічний університет ім. Ст. Сташіца в Кракові
Прикарпатський національний університет ім. В. Стефаника
Національний університет кораблебудування імені адмірала Макарова
Одеський національний університет імені І.І. Мечникова
Національний університет Запорізька політехніка

Інформаційні системи та технології ІСТ-2024

Матеріали
13-ї Міжнародної науково-технічної конференції

Частина 1

26-28 листопада 2024 р.
Харків, Україна

Харків 2024

Ministry of education and science of Ukraine
The National Academy of Sciences of Ukraine
Polish Academy of Science, branch in Lublin
Representative office „Polish Academy of Science” in Kyiv
Kharkiv National University of Radio Electronics
V.N. Karazin Kharkiv National University
Science and Technological University AGH after St. Staszic in Krakow
V. Stefanik Precarpathian National University
Admiral Makarov National University of Shipbuilding
Odesa I.I. Mechnikov National University
National University "Zaporizhzhia Polytechnic"

Information Systems and Technologies IST-2024

Proceedings

of the 13th International Scientific and Technical Conference

Part 1

November 26 – 28, 2024
Kharkiv, Ukraine

Kharkiv 2024

УДК: 004.9

Наукові редактори:

Ю.О. Романенков – доктор технічних наук, професор, проректор з наукової роботи (Харківський національний університет радіоелектроніки)

В.В. Безкоровайний - доктор технічних наук, професор (Харківський національний університет радіоелектроніки)

S. Yakovlev – доктор фізико-математичних наук, професор, член-кореспондент НАН України (Харківський національний університет ім. В. Каразіна, Lodz University of Technology, Poland);

L. Petryshyn - доктор технічних наук, професор (AGH University of Science and Technology Poland; Прикарпатський національний університет ім. В. Стефаника);

З.В. Дудар – кандидат технічних наук, професор (Харківський національний університет радіоелектроніки)

В.Г. Кобзєв – кандидат технічних наук, с.н.с. (Харківський національний університет радіоелектроніки)

Матеріали опубліковано в авторській редакції.

Інформаційні системи та технології: матеріали 13-ї Міжнародної науково-технічної конференції. Частина 1. [Електронний ресурс], Харків, 26 - 28 листопада 2024 року / наук. ред. Ю.О. Романенков, В.В. Безкоровайний, С.В. Яковлев, Л.Б. Петришин, З.В. Дудар, В.Г. Кобзєв. – Х.: ХНУРЕ, 2024. – 198с.

ISBN 978-966-659-354-5
DOI: 10.30837/IST-2024.P1

Збірка містить матеріали доповідей науковців Міжнародної науково-технічної конференції з проблем сучасних інформаційних систем та технологій.

Роботи представляють інтерес для фахівців, науковців, аспірантів, студентів, діяльність яких пов’язана з розробкою та впровадженням сучасних інформаційних систем і технологій.

ISBN 978-966-659-354-5

Голова конференції

Романенков Ю.О. – проректор з наукової роботи ХНУРЕ, д.т.н., проф.

Співголови

Дудар З.В. - к.т.н., проф., ХНУРЕ
Petryshyn L. - д.т.н., проф., AGH UST, Польща, ПНУ (Івано-Франківськ)
Sobczuk H. - д.т.н., проф., Польська академія наук
Yakovlev S. - член-кореспондент НАН України, д.ф-м.н., проф., Lodz University of Technology, Польща
Кобзєв В.Г. - к.т.н., с.н.с., ХНУРЕ

Відповідальні секретарі

Міщеряков Ю.В., к.т.н., доц., ХНУРЕ
Ховрат А.В., асп. ХНУРЕ

Робоча група організаційного комітету

ХНУРЕ: Груздо І.В., Олійник О.О., Широкопетлєва М.С., Бірюкова Н.С., Волоховський В.Є., Міщеряков А.Ю.
ПНУ – Петришин М.Л., Ізмайлов А.В.
ОНУ – Фразе-Фразенко О.О.

Програмний комітет конференції

- **Безкоровайний В.В.**, ХНУРЕ, Україна – голова
- **Антощук С.Г.**, ОНПУ, Україна
- **Безсонов О.О.**, ХНУРЕ, Україна
- **Бісікало О.В.**, ВНТУ, Україна
- **Блінцов В.С.**, НУК, Україна
- **Бодянський Є.В.**, ХНУРЕ, Україна
- **Бондаренко В.Г.**, НТТУ «КПІ», Україна
- **Голуб С.В.**, ЧДТУ, Україна
- **Гоменюк С.І.**, ЗНУ, Україна
- **Горбенко І.Д.**, ХНУ, Україна
- **Гребеннік І.В.**, ХНУРЕ, Україна
- **Гунченко Ю.О.**, ОНУ, Україна
- **Дорошенко В.О.**, ХНУРЕ, Україна
- **Євланов М.В.**, ХНУРЕ, Україна
- **Єрохін А.Л.**, ХНУРЕ, Україна
- **Захарченко О.О.**, ННЦ ХФТІ НАН, Україна
- **Іванченко С.О.**, ІСЗЗІ НТУУ «КПІ», Україна
- **Іляш Ю.Ю.**, ПНУ, Україна
- **Іохов О.Ю.**, НАНГУ, Україна
- **Казакова Н.Ф.**, ОНУ, Україна
- **Карасюк В.В.**, НЮУ, Україна
- **Карташов В.М.**, ХНУРЕ, Україна
- **Кіріченко Л.О.**, ХНУРЕ, Україна, Lodz University of Technology, Польща
- **Кобозева А.А.**, ОНУ, Україна
- **Кобилін О.А.**, ХНУРЕ, Україна
- **Косенко В.В.**, НУПП, Україна
- **Лемешко О.В.**, ХНУРЕ, Україна
- **Литвин О.О.**, ХНУ, Україна
- **Лужецький В.А.**, ВНТУ, Україна
- **Машталір В.П.**, ХНУРЕ, Україна
- **Мінухін С.В.**, ХНЕУ, Україна
- **Нефьодов Л.І.**, ХНАДУ, Україна
- **Нужний С.М.**, НУК, Україна
- **Пасічник В.В.**, НУЛП, Україна
- **Петров К.Е.**, ХНУРЕ, Україна
- **Полозова Т.В.**, ХНУРЕ, Україна
- **Пономарьов Ю.В.**, АТ «Укртрансгаз», Україна
- **Пушкар О.І.**, ХНЕУ, Україна
- **Руденко О.Г.**, ХНУРЕ, Україна
- **Руткас А.Г.**, ХНУРЕ, Україна
- **Свид І.В.**, ПНУ, Україна
- **Снитюк В.Є.**, КНУ, Україна
- **Сидоров М.В.**, ХНУРЕ, Україна
- **Синєглазов В.М.**, НАУ, Україна
- **Смеляков К.С.**, ХНУРЕ, Україна
- **Стоян Ю.Г.**, ІПМаш НАН, Україна
- **Субботін С.О.**, НУЗП, Україна
- **Тевяшев А.Д.**, ХНУРЕ, Україна
- **Удовенко С.Г.**, ХНЕУ, Україна
- **Узлов Д.Ю.**, ХНУ, Україна
- **Філатов В.О.**, ХНУРЕ, Україна
- **Хажмурадов М.А.**, ННЦ ХФТІ НАН, Україна
- **Хлевна Ю.Л.**, КНУ, Україна
- **Чалий С.Ф.**, ХНУРЕ, Україна
- **Четверіков Г.Г.**, ХНУРЕ, Україна
- **Чопоров С.В.**, ЗНУ, Україна
- **Чумаченко Д.І.**, НАКУ «ХАІ», Україна, University of Waterloo, Канада
- **Чумаченко І.В.**, ХНУМГ, Україна
- **Чумаченко О.І.**, НТТУ «КПІ», Україна
- **Шаронова Н.В.**, НТУ «ХПІ», Україна
- **Шаховська Н.Б.**, НУЛП, Україна
- **Шевченко І.В.**, КрНУ, Україна
- **Шубін І.Ю.**, ХНУРЕ, Україна
- **Бондаренко А.**, РТУ, Латвія
- **Chamot M.**, WSG, Польща
- **Dudek M.**, AGH UST, Польща
- **Kadyrov A.**, KhSU, Таджикістан
- **Kusz A.**, ПАН, Польща
- **Lebkovski P.**, AGH UST, Польща
- **Milczarski P.**, Лодзький ун-т, Польща
- **Nykyforchyn O.**, Kasimir the Great University, Польща
- **Petryshyn L.**, AGH UST, Польща, ПНУ, Україна
- **Sakalys P.**, VK, Литва
- **Schumm E.**, HSHL, Німеччина
- **Yehorchenkova N.**, STU, Словаччина

Секція 1.
Сучасні інформаційні системи та технології: проблеми, методи, моделі. Управління проектами та програмами

DOI: https://doi.org/10.30837/IST-2024.P1.006

Концепція побудови інформаційної системи підприємства з використанням ШІ в рамках Індустрії 4.0

Дмитро Костарев[1†], Андрій Тевяшев[1†], Наталія Сизова[2†] та Володимир Ткаченко[1†]

[1] Харківський національний університет радіоелектроніки, пр. Науки 14, Харків, 61166, Україна

[2] Харківський національний університет міського господарства, вул. Маршала Бажанова, 17, Харків, 61002, Україна

Анотація

Подається концепція створення, використання та впровадження інтелектуальних інформаційних технологій управління промисловим підприємством шляхом використання штучного інтелекту.

Ключові слова

інформаційне моделювання, системи управління, промислове підприємство, штучний інтелект, Індустрія 4.0

1. Вступ

Сучасний розвиток економіки, використання цифрових технологій, ускладнення та удосконалення економічних моделей призвели до необхідності запровадження, пристосування до сучасних умов та зміни існуючих технологій управління підприємством, оскільки існуючі моделі управління підприємством стають мало ефективними. На даний час одним із ключових підходів розвитку економіки є підхід, що базується на Індустрії 4.0/5.0, характерні риси якого – розвиток інформаційно-комунікаційних технологій, автоматизація та роботизація виробничих процесів [1, 2].

Актуальними постають питання створення та розвитку сучасних систем управління промисловими підприємствами, оскільки підприємства є рушійною силою економіки будь-якої держави. Одним із інноваційних методів управління промисловим підприємством є проактивне управління.

Проактивне управління ґрунтується на передбаченні можливих несприятливих процесів у функціонуванні та розвитку підприємств, що дозволяє підприємству бути готовим до можливих проблем, запобігати їх появі та вирішувати їх до того, як вони виникнуть [1, 2].

Одним із підходів проактивного управління є питання створення, використання та впровадження інтелектуальних інформаційних технологій управління промисловим підприємством шляхом використання нових технологій, таких як штучний інтелект, Інтернет промов, великі дані тощо. і створення інформаційної платформи [3]. Ефективне розв'язання завдань проактивного управління підприємством може бути реалізовано у складі інформаційних систем на основі штучного інтелекту [4, 5].

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine

* Автор-кореспондент.

† Ці автори мають рівний внесок.

dmytro.kostariev@nure.ua (Д. Костарев); andrew.teviashev@nure.ua (А. Тевяшев); sizovnataliad11@gmail.com (Н. Сізова) ; volodymyr.tkachenko@nure.ua (В. Ткаченко)

0000 –0002–9928-8571 (Д. Костарев); 0000-0001-5261-9874 (А. Тевяшев);
0000 –0002– 0103–1939 (Н. Сізова) ; 0000-0002-5076-0724 (В. Ткаченко)

Метою роботи є розробка концепції інформаційної платформи для проактивного управління промисловим підприємством, що включає створення інформаційно-аналітичної системи на основі штучного інтелекту.

Для досягнення цієї мети передбачено вирішення наступних завдань дослідження::

- Аналіз існуючих підходів до проактивного управління підприємствами з акцентом на новітні інформаційно-комунікаційні технології.
- Розробка інформаційно-аналітичної системи на основі штучного інтелекту, здатної збирати, обробляти та аналізувати великі обсяги даних, отримані з виробничих ліній та датчиків.
- Впровадження передових рішень для прогнозування витрат ресурсів та оптимізації виробничих процесів з використанням алгоритмів машинного навчання та IoT.
- Моделювання автоматизованих процесів управління ресурсами на основі даних про споживання електроенергії, води, газу та сировини з можливістю передбачення відхилень та оперативного реагування на них.
- Розробка модулів для передиктивного обслуговування обладнання з використанням інтелектуальних алгоритмів аналізу стану технічного парку підприємства, що дозволить мінімізувати ризики поломок та знизити витрати на технічне обслуговування.
- Оптимізація процесів прийняття рішень керівниками підприємств на основі даних аналітики та прогнозів, наданих системою, з метою підвищення гнучкості та конкурентоспроможності підприємства.

2. Концепція інформаційної системи проактивного управління

Концепція інформаційної системи передбачає комплексний підхід до проактивного управління підприємством, починаючи від збору та обробки даних і закінчуючи підготовкою документів для впровадження їх у процес управління.

Концепція системи управління підприємствам у рамках Індустрія 4.0 передбачає:

- збір даних про витрати ресурсів на тих чи інших виробничих лініях (електроенергія, вода, газ, сировина);
- мати дані норми витрат у базі даних для встановлені на виробництво тих чи інших видів продукції, що випускаються на підприємстві;
- отримувати дані про кількість та якість сировини, необхідну для виготовлення продукції;
- дані про продаж тих чи інших видів продукції у попередній період, а також аналогічний період попередніх років (з CRM підприємства).

Інформаційна система на підставі аналізу даних система повинна:

- мати можливість адаптації лімітів сировини та витрат ресурсів на виробничих лініях залежно від виробленої продукції (встановлювати ліміт витрати ресурсів на ту чи іншу продукцію та у разі відхилення від прийнятих норм інформувати оператора);
- мати можливість прогнозувати витрати ресурсів залежно від встановленого виробничого плану;
- система повинна мати портфель пропозицій з виробництва тієї чи іншої продукції залежно від запланованого обсягу продажів на підставі інформації про вхідну сировину.

Використання передових технологій штучного інтелекту дозволяє не лише прискорити процеси, а й значно підвищити ефективність функціонування підприємства та його конкурентоспроможність.

3. Структура інформаційної системи

Структура інформаційної системи складається з інтегрованих підсистем – модулів, які мають такі компоненти.

1. Модуль збору даних про витрати ресурсів на виробничих лініях, що інтегрується з різними датчиками та вимірювальними пристроями на виробничих лініях для збору даних про споживання ресурсів (електроенергія, вода, газ). Він також пов'язаний із базою даних, де зберігаються нормативи витрати цих ресурсів для різних видів продукції.
2. Модуль повідомлення про відхилення від лімітів, що модуль дозволяє підприємству оперативно реагувати на будь-які відхилення у витраті ресурсів чи виробничих параметрах від встановлених норм. Наприклад, якщо витрата води або електроенергії на будь-якій виробничій лінії перевищує встановлений ліміт, система миттєво надсилає повідомлення відповідальним особам. Це дозволяє швидко ідентифікувати та усунути причини неефективності, запобігаючи додатковим витратам і втратам.
3. Модуль сповіщення про необхідність обслуговування обладнання. Система також включає модуль передиктивного обслуговування, який відстежує відпрацьовані мотогодини та інші параметри роботи обладнання. На основі цих даних система прогнозує оптимальні інтервали для проведення технічного обслуговування, що допомагає запобігти раптовим поломкам і простоям у роботі. Крім того, систематичне обслуговування обладнання продовжує його термін служби та підвищує загальну надійність виробництва.
4. Модуль контролю якості та кількості сировини (наприклад, молока). Цей модуль дозволяє отримувати дані про кількість і якість сировини, включаючи компоненти сировини та інші важливі параметри. Ця інформація використовується для планування виробничих процесів.
5. Модуль аналізу даних продажів з CRM. Система інтегрована з CRM підприємства для отримання даних про продаж різних видів продукції. Це дозволяє аналізувати сезонні тренди та переваги споживачів, наприклад, збільшення продажів йогуртів влітку та сметани взимку.
6. Модуль пропозицій щодо виробництва продукції. Використовуючи дані про передбачуваний обсяг продажів та інформацію про якість та кількість вхідної сировини, система пропонує оптимальний асортимент продукції для виробництва. Це допомагає підвищити ефективність виробництва та задовольнити попит споживачів.

Зв'язок між модулями системи проактивного управління підприємством представлений на рис. 1.

Інтегрована інформаційна система підприємства складається з взаємозалежних модулів, які спільно забезпечують оптимізацію виробничих процесів та ефективне управління ресурсами.

1. Модуль збору даних про витрати ресурсів: Цей модуль відповідає за автоматизований збір фактичних показників споживання ресурсів, таких як електроенергія, вода та газ, безпосередньо з датчиків і вимірювальних пристроїв

на виробничих лініях. Отримані дані співставляються з нормативними значеннями, що зберігаються в базі даних, для виявлення можливих відхилень.

Рисунок 1: Взаємозв'язок між модулями у системі управління підприємством

2. Модуль повідомлення про відхилення від лімітів: Отримуючи інформацію від першого модуля, цей компонент здійснює аналіз на предмет перевищення встановлених лімітів споживання ресурсів. У разі виявлення аномалій, модуль оперативно генерує сповіщення для відповідальних осіб, забезпечуючи швидку реакцію на потенційні проблеми.
3. Модуль сповіщення про необхідність обслуговування обладнання: Використовуючи дані про мотогодини та інші експлуатаційні параметри від модуля збору даних, цей модуль прогнозує оптимальні інтервали для проведення технічного обслуговування обладнання. Це дозволяє запобігти несподіваним поломкам та мінімізувати простої у виробництві.
4. Модуль контролю якості та кількості сировини: Цей компонент забезпечує моніторинг кількості та якості вхідної сировини, наприклад, молока. Отримані дані використовуються для точного планування виробничих процесів та гарантування відповідності продукції встановленим стандартам якості.
5. Модуль аналізу даних продажів з CRM: Інтегрований з системою CRM підприємства, цей модуль аналізує дані про продажі, враховуючи сезонні тренди та споживчі вподобання. Це дозволяє прогнозувати попит на різні види продукції та адаптувати виробництво відповідно до ринкових умов.
6. Модуль пропозицій щодо виробництва продукції: На основі інформації від модулів контролю сировини та аналізу продажів, цей модуль формує науково обґрунтовані пропозиції щодо оптимального асортименту та обсягів продукції для виробництва. Це сприяє підвищенню ефективності виробництва та максимальному задоволенню потреб споживачів.

Взаємодія модулів у системі:

- Модуль збору даних про витрати ресурсів тісно інтегрується з модулем повідомлення про відхилення від лімітів та модулем сповіщення про необхідність обслуговування обладнання. Така взаємодія забезпечує постійний контроль за споживанням ресурсів та технічним станом обладнання, що є критичним для безперервного та ефективного виробничого процесу.

- Модуль контролю якості та кількості сировини спільно з модулем аналізу даних продажів з CRM надають ключові дані для модуля пропозицій щодо виробництва

продукції. Ця взаємодія дозволяє синхронізувати виробничі плани з реальним станом ресурсів та ринковим попитом, забезпечуючи гнучкість та адаптивність виробництва.

- Модуль пропозицій щодо виробництва продукції виступає як центральний елемент, який інтегрує інформацію з різних джерел та генерує оптимальні виробничі стратегії. Це сприяє раціональному використанню ресурсів та підвищенню конкурентоспроможності підприємства.

Інтеграція цих модулів базується на принципах системного аналізу та управління, використовуючи сучасні інформаційні технології для забезпечення високої точності та оперативності в прийнятті рішень. Застосування передових методів збору та обробки даних дозволяє підприємству здійснювати проактивний контроль за всіма аспектами виробництва, від забезпечення ресурсами та сировиною, через контроль якості і планування до задоволення кінцевого споживача.

Такий системний підхід забезпечує синергію між різними функціональними областями підприємства, підвищуючи ефективність управління та сприяючи сталому розвитку в умовах динамічного ринкового середовища.

Автоматизація та підтримка прийняття рішень при такому подході дозволяє розробляти рекомендації з управління виробництвом, що базуються на аналізі поточних даних, історичних тенденцій та прогнозів.

4. Огляд інформаційної системи в концепції Індустрія 4.0

Можна відзначити, що на сьогоднішній день є розроблені інформаційні системи, серед яких інтелектуальна система ЕКОФЛЕКС [6], що є IoT системою і впроваджується відповідно до концепцій Industry 4.0/5.0 [7]. Ця система вирішує завдання моніторингу витрат споживання ресурсів, здійснює автоматизований збір, облік, зберігання даних та надає аналітичну інформацію щодо споживання та витрат ресурсів як для всього підприємства, так і окремих його підрозділів. Система ЕКОФЛЕКС постійно удосконалюється у напрямку створення та реалізації системи управління підприємством, це стосується таких потань як

1. Моніторинг витрат та споживання ресурсів: Система автоматично збирає дані про витрати ресурсів (електроенергії, води, газу) на виробничих лініях, забезпечуючи детальний аналіз та оптимізацію використання ресурсів для всього підприємства.
2. Автоматизований збір та обробка даних: ЕКОФЛЕКС здійснює збір, облік та зберігання великих обсягів даних, що дозволяє в режимі реального часу отримувати інформацію про споживання ресурсів на різних етапах виробництва.
3. Аналітика та надання звітів: Система надає аналітичну інформацію як для всього підприємства, так і для окремих підрозділів, дозволяючи керівникам приймати обґрунтовані рішення на основі даних.
4. Проактивне управління ресурсами: Використовуючи технології штучного інтелекту та великі дані, система здатна прогнозувати майбутні витрати ресурсів та заздалегідь попереджати про можливі відхилення або нестачі ресурсів, запобігаючи негативним наслідкам.
5. Автоматизація виробничих процесів: ЕКОФЛЕКС сприяє роботизації виробничих процесів, забезпечуючи автоматичне налаштування та адаптацію систем до змін у виробничих умовах.
6. Предиктивне обслуговування обладнання: Завдяки автоматизованому моніторингу роботи обладнання система може передбачати несправності, оптимізувати графіки технічного обслуговування та зменшувати ризик раптових поломок.

7. Оптимізація управлінських рішень: Система надає керівництву підприємства можливість оперативно реагувати на зміни у виробничих процесах, впроваджуючи проактивні рішення для покращення ефективності та конкурентоспроможності.

5. Наукові та технічні рішення в системі Екофлекс

При побудові системи було застосовано низку математичних моделей і методів для забезпечення її ефективності та надійності [8-10]. У модулі збору даних про витрати ресурсів використовувалися методи сигналового аналізу та статистичної обробки для точного вимірювання та фільтрації даних від датчиків. Модуль повідомлення про відхилення від лімітів базується на статистичному процесному контролі, зокрема на контрольних картах Шухарта, що дозволяє своєчасно виявляти аномалії у виробничих процесах.

Модуль сповіщення про необхідність обслуговування обладнання застосовує методи машинного навчання, такі як регресійні моделі та дерева рішень, для прогнозування технічного стану обладнання на основі аналізу часових рядів. У модулі контролю якості та кількості сировини використовуються статистичний контроль якості та класифікаційні алгоритми для оцінки відповідності сировини встановленим стандартам.

Модуль аналізу даних продажів з CRM використовує моделі прогнозування часових рядів, зокрема ARIMA та нейронні мережі LSTM, для прогнозування попиту з урахуванням сезонних трендів. Модуль пропозицій щодо виробництва продукції базується на методах оптимізації, таких як лінійне програмування та генетичні алгоритми, для визначення оптимального плану виробництва з метою максимізації прибутку.

Для побудови інформаційної системи Екофлекс використовуються сучасні технології. Система базується на модульній та мікросервісній архітектурі, з використанням Інтернету речей (IoT) для збору даних у реальному часі та систем реального часу для негайної обробки. Дані зберігаються в реляційних та нереляційних базах даних. Хмарні платформи забезпечують масштабованість.

Застосовуються технології штучного інтелекту: машинне навчання та предиктивна аналітика (TensorFlow, PyTorch) для прогнозування технічного обслуговування, аналізу продажів та оптимізації виробництва.

Штучний інтелект був технічно інтегрований у систему шляхом впровадження алгоритмів машинного навчання та аналітики даних у ключові модулі. У Модулі сповіщення про необхідність обслуговування обладнання були реалізовані предиктивні моделі, які на основі даних про мотогодини та інші параметри прогнозують оптимальний час для технічного обслуговування. Модуль аналізу даних продажів з CRM використовує алгоритми глибинного навчання для аналізу історичних даних продажів та прогнозування майбутнього попиту, враховуючи сезонні тренди. У Модулі пропозицій щодо виробництва продукції застосовані алгоритми оптимізації та ШІ для формування оптимального асортименту продукції, базуючись на даних про якість сировини та прогнозований попит. Модуль контролю якості та кількості сировини використовує технології ШІ для аналізу параметрів сировини в реальному часі та виявлення відхилень від норм. ШІ інтегрований у систему через ці модулі, підвищуючи ефективність, точність прогнозування та автоматизуючи процеси прийняття рішень.

Таким чином, система інтегрує моделі, архітектурні, технічні та програмні рішення з галузей статистики, машинного навчання, оптимізації та системного аналізу. Це забезпечує можливість ефективно обробляти великі обсяги даних, точно прогнозувати та оптимізувати виробничі процеси, що сприяє прийняттю обґрунтованих управлінських рішень і підвищенню загальної ефективності роботи підприємства.

6. Висновки

Впровадження технологій штучного інтелекту на підприємстві дозволяє досягти більшої гнучкості та адаптивності до змін ринкових умов і потреб споживачів. Автоматизація та прискорення бізнес-процесів сприяють підвищенню ефективності діяльності та забезпечують додаткову економію ресурсів. Ефективна інтеграція та застосування цих технологій і моделей не тільки покращують поточні операційні показники, але й гарантують довгострокову стійкість і конкурентоспроможність підприємства на ринку.

Описана концепція та наведений приклад представляють собою наступний крок у побудові подібних систем та їх впровадженні у виробничі процеси. Це демонструє можливості сучасних технологій у інтеграції з існуючими системами для досягнення більш високого рівня автоматизації та інтелектуалізації виробництва. Такий підхід є важливим етапом у реалізації концепції "Індустрія 4.0", де технології та дані стають ключовими елементами в управлінні підприємством.

Література

[1] Schwab, K. "The Fourth Industrial Revolution". World Economic Forum, 2017. https://www.weforum.org/about/the-fourth-industrial-revolution-by-klaus-schwab/.

[2] Frank, AG, Dalenogare, LS, Ayala, NF "Industry 4.0 technologies: Implementation patterns in manufacturing companies". International Journal of Production Economics, 2019. International o Production Economics V 210. - p.15-26.

[3] Ковтуненко Ю.В. Застосування штучного інтелекту в системі управління підприємством: проблеми та переваги / Ю.В. Ковтуненко // Економічний журнал Одеського політехнічного університету. - 2019. - № 2 (8). - С. 93-99. DOI: 10.5281/zenodo.4171114.

[4] Фещур Р. В., Шишковський С. В., Якимів А. І. Інструменти управління проактивним розвитком підприємств // Бізнесінформ – 2018. № 2. – С.283-289.

[5] Чорноус Г. Інформаційне забезпечення проактивного управління // Вісник КНТЕУ, 2012. – №5. – С.102-114.

[6] https://flexsys.com.ua/o-nas/

[7] Задорожний Г.В., Дуна Н.Г., Задорожна О.Г., Особливості та перспективи індустрії 4.0 в економіці України (науковий огляд) // Вісник економічної науки України, 2021. - № 1. - с.159-179.

[8] Баган Т.Г. Методи машинного навчання при проектуванні автоматизованих систем керування [Електронний ресурс] /Т. Г. Баган - Київ: КПІ ім. Ігоря Сікорського, 2021. 28 с.

[9] Литвин В. В. Методи та засоби інженерії даних та знань / В.В. Литвин, Ю.В.Нікольський, В.В. Пасічник – Львів: Магнолія 2006, 2021. - 252 с.

[10] D. B. Kostarev, A. D. Tevyashev, N. D. Sizova, V. P. Tkachenko. Analysis of the Problem of Forecasting in Proactive Enterprise Management / Intelligent Sustainable Systems. Selected Papers of WorldS4 2023, Volume 1. - 2024. -PP. 539-549. https://doi.org/10.1007/978-981-99-8031-4.

Децентралізація в основі систем афілійованого маркетингу

Микола Маленко[1†]

[1] *Київський національний університет будівництва і архітектури, пр-т Повітряних Сил, 31, 030371, Київ, Україна*

Анотація

У роботі проводиться аналіз централізованої афіліат мережі та пропонується розглянути впровадження децентралізованого підходу до побудови систем афілійованого маркетингу на основі Web3 технологій. Використання блокчейну та смарт контрактів дозволить зменшити роль афіліат платформи як посередника між рекламодавцем та афіліатом, знизить операційні витрати, та забезпечити прозорість і довіру між учасниками. Це сприятиме підвищенню ефективності, надійності та прибутковості систем афілійованого маркетингу, а також зменшить ризики шахрайства та конфліктів між учасниками мережі. Робота показує актуальність та перспективність дослідження підходів використання Web3 технологій у сфері афілійованого маркетингу.

Ключові слова

блокчейн, смарт-контракти, децентралізація, афіліат маркетинг

1. Вступ

Бізнес та підприємництво в сучасній економіці України є фундаментальною складовою [1]. Одним з ключових рис успішного та правильно вибодуваного бізнесу є здатність швидко адаптуватись до економічних умов та змін ринку, це допомагає економіці бути гнучкою, стійкою до негативних чинників та швидко відновлюватись від економічних шоків і криз [2]. З розвитком та популяризацією інтернет технологій підприємництво активно почало переміщуватись в онлайн простір, наразі, це настільки масове явище, що можна сміливо стверджувати, якщо вашого бізнесу немає в інтернеті, то вас немає в бізнесі [2]. Через велику конкуренцію та перенасичення інтернет-користувачів інформацією, інтернет маркетинг став невід'ємною та життєзабезпечуючою складовою будь-якого бізнеса представленого в інтернеті. Чим краще налагоджені процеси онлайн-просування, тим більше можливостей для зростання бізнесу [2]. Таким чином, вдосконалення маркетингових інструментів та підходів, в тому числі за рахунок впровадження інноваційних технологій, сприяє стабільному підвищенню прибутковості, конкурентоспроможності на ринку та в подальшому покращенню економічної ситуації в країні.

2. Основна частина

Одним з найефективніших підходів у інтернет-маркетингу є перфоманс маркетинг [3]. Такий підхід у порівнянні з класичними маркетингом, де створюються довготривалі рекламні кампанії і закладаються великі рекламні бюджети без будь-якої гарантії успішності таких кампаній, базується на досягненні вимірюваних результатів у відносно короткий термін [3]. Ключовим елементом цього підходу є плата за дію. Компанія як

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine

[2*] Маленко Микола, аспірант кафедри кібербезпеки та комп'ютерної інженерії, Київського національний університету будівництва і архітектури.

mm.inftech@outlook.com (М. Маленко)

0000-0002-7360-7749 (М. Маленко)

споживач рекламних послуг оплачує не всі рекламні витрати, а лише ті, які призвели до бажаного результату або дії з боку інтернет-користувача, а в подальшому можливого клієнта компанії. Серед таких дій можуть бути реєстрація користувача на сайті компанії, створення заявки на отримання послуг компанії, додавання товару в кошик інтернет-магазину, внесення оплати за товар або послугу компанії, та в цілому будь-яка дія за яку компанія готова платити рекламному підряднику.

Рекламні підрядники спеціалізуються та використовують різні канали залучення потенційних клієнтів, або ж лідів. Одні канали можуть бути більш ефективними для компанії інші менш, те саме стосується і загальної специфіки роботи підрядника з каналом просування на якому він спеціалізується [4].

З метою управління та оптимізації процесів взаємодії з рекламними підрядниками на більш масштабному рівні, компанії часто звертаються до систем афілійованого маркетингу або партнерських мереж, які об'єднують рекламних підрядників і забезпечують інфраструктуру для автоматизації всіх етапів співпраці [5]. Системи афілійованого маркетингу є платформами посередниками між бізнесом, компанією, власником продукту, рекламодавцем, що потребує послуг онлайн просування та рекламним посередником, афіліатом, партнером, спеціалістом який вміє ефективно взаємодіяти з конкретними рекламними каналами [5]. Партнерські мережі забезпечують інструменти для моніторингу результатів співпраці між рекламодавцем і афіліатом, механізми та інструменти відстеження дій користувачів, потенційних клієнтів, залучених рекламними підрядниками–партнерами, афіліатами до взаємодії з продуктами або сервісами компанії–рекламодавця [5].

Також системи афілійованого маркетингу керують фінансовими потоками, що виникають в результаті співпраці між рекламодавцем та партнером. Окрім цього партнерські мережі, виконуючи роль посередника, регулюють конфліктні ситуації та спори, які можуть виникати в процесі співпраці між учасниками мережі. Останні можуть виникати через шахрайські дії, здійснені однією зі сторін конфлікту [6]. Прикладом шахрайських дій може бути залучення невалідних лідів або фактично не існуючих користувачів за які будуть списані кошти з балансу рекламодавця [6]. Прикладом шахрайства з боку рекламодавця може бути приховування факту здійснення цільової дії, за яку афіліат повинен отримувати винагороду. Оскільки партнерські мережі є централізованими системами, то є високі ризи того, що конфліктні ситуації не будуть вирішені за принципами чесності та справедливості. Окрім цього, через жорстку централізацію виникає ризик шахрайства з боку системи афілійованого маркетингу, адже її прибуток формується на основі комісій за посередницькі послуги, що зумовлює певну зацікавленість у отриманні більшою суми від рекламодавця та зменшенні виплати рекламному підряднику. До частини процесів, з фінансовими процесами включно, в партнерській мережі, можуть бути залучені працівники, що створює додаткові ризики, а саме допущення помилок через людський фактор.

Децентралізований підхід побудови систем афілійованого маркетингу, з використанням Web3 технологій, міг би мінімізувати роль посередника та автоматизувати посередницькі задачі, які зазвичай виконуються співробітниками. Зокрема валідацію кліків або інших цільових дій, обробку транзакцій, розподіл та нарахування винагород партнерам. Впровадження децентралізованих реєстрів [7] та застосування смарт-контрактів [8], що працюють на блокчейні може забезпечити формалізацію відносин між учасниками системи у вигляді незмінних, відкритих, за потреби гнучких, правил. Ці правила виконуються автоматично в мережі [8]. Такий підхід може дозволити знизити операційні витрати за рахунок зменшення традиційних комісій посередників, а також може зменшити, або взагалі виключити ймовірність помилок та прихованих маніпуляцій компонентами системи спричинених співробітниками афіліат мережі. Більше того, впровадження технології блокчейн може забезпечити прозору, надійну та доступну для всіх сторін інформацію про фінансові потоки, це сприятиме

більшій довірі між учасниками системи, скоротить кількість спорів і непорозумінь, а у разі їх виникнення дозволить вирішувати такі ситуації за принципами чесності та справедливості.

3. Висновки

Децентралізований підхід і використання Web3 технологій, таких як блокчейн та смарт-контракти, відкривають нові можливості для модернізації систем афілійованого маркетингу, встановлюючи стандарти прозорості, надійності та ефективності. Впровадження цих технологій сприяє розвитку інноваційних бізнес-моделей для ефективного управління трафіком, відстеження результатів рекламних кампаній та захисту даних учасників. Таким чином, дослідження використання Web3 технологій у сфері афілійованого маркетингу є актуальним напрямом, що підвищує ефективність і надійність взаємодії між учасниками мережі та покращує прибутковість партнерських систем.

Посилання

[1] Петриченко В. П., Розвиток малого і середнього бізнесу в Україні: проблеми і перспективи [Електронний ресурс] // Вісник Національного університету "Львівська політехніка". Режим доступу: https://science.lpnu.ua/uk/semi/vsi-vypusky/volume-4-number-1-2020/rozvytok-malogo-i-serednogo-biznesu-v-ukrayini-problemy-i.

[2] Гноєвий В.Г., Корень О.М. Сучасні тенденції цифрового маркетингу та їх вплив на формування маркетингової стратегії [Електронний ресурс] // Академічний огляд. – 2021. – №1. Режим доступу: https://acadrev.duan.edu.ua/images/PDF/2021/1/6.pdf.

[3] Корж Н. Що таке перформанс-маркетинг [Електронний ресурс] // Admixer Academy Blog. Режим доступу: https://blog.admixer.academy/ua/shcho-take-performance-marketing/

[4] Вдовічена О. Г. Digital-маркетинг: особливості використання різних каналів для залучення клієнтів [Електронний ресурс] // Науковий вісник ЧНУ ім. Петра Могили. Режим доступу: https://dspace.chmnu.edu.ua/jspui/bitstream/123456789/782/1/%D0%92%D0%B4%D0%BE%D0%B2%D1%96%D1%87%D0%B5%D0%BD%D0%B0%20%D0%9E.%20%D0%93.%20Digital-%D0%BC%D0%B0%D1%80%D0%BA%D0%B5%D1%82%D0%B8%D0%BD%D0%B3.pdf

[5] Sareen P., Rani P. Affiliate Marketing: Strategies, Trends, and Challenges in the Digital Landscape [Електронний ресурс] // International Journal of Creative Research Thoughts (IJCRT). – 2018. – Т. 6, № 1. – Лютий. – ISSN 2320-2882. Режим доступу: https://ijcrt.org/papers/IJCRT1135580.pdf

[6] 7 Types of Affiliate Marketing Fraud to Be Aware of to Protect Your Affiliate Program [Електронний ресурс] // SpiderAF. Режим доступу: https://spideraf.com/articles/7-types-of-affiliate-marketing-fraud-to-be-aware-of-to-protect-your-affiliate-program.

[7] Hayes A. Blockchain Facts: What Is It, How It Works, and How It Can Be Used [Електронний ресурс] // Investopedia. Режим доступу: https://www.investopedia.com/terms/b/blockchain.asp

[8] Taherdoost H. Smart Contracts in Blockchain Technology: A Critical Review [Електронний ресурс] // Information. – 2023. – Т. 14, № 2. – С. 117. DOI: 10.3390/info14020117. Режим доступу: https://www.mdpi.com/2078-2489/14/2/117

DOI: https://doi.org/10.30837/IST-2024.P1.016

Підвищення ефективності логістичних процесів виробничого підприємства шляхом формування раціональних каналів та оптимізації запасів

Ольга Малєєва[1] та Юрій Полупан[1]

[1] Національний аерокосмічний університет ім. М. Є. Жуковського "Харківський авіаційний інститут," вул. Вадима Манька, 17, 61070, м. Харків, Україна

Анотація

У роботі розглянуто основні проблеми оптимізації логістичних процесів на виробничих підприємствах. Приділено увагу вибору структури логістичних каналів, методам управління запасами та оцінці ефективності логістичних рішень. Розглядаються різні моделі управління логістикою з порівнянням їх переваг та недоліків. Також, наведений аналіз використання інформаційних технологій для підвищення ефективності логістичних процесів, мінімізації витрат і ризиків.

Ключові слова

Логістичні канали, логістичні ланцюги, управління запасами, конкуренція, оптимізація, модель EOQ, економічна ефективність, виробниче підприємство

1. Вступ

В сучасному світі глобалізації та великої конкуренції успішність виробничих підприємств визначається ефективністю логістичних процесів. Особливо важливим для стабільності випуску продукції та забезпечення її конкурентоспроможності є вчасна поставка ресурсів та оптимізація процесів збуту продукції. До переліку компонентів логістичних каналів, на які треба звертати увагу, можна віднести матеріальні ресурси, фінансові та інформаційні потоки, що задають швидкість та ефективність виробничих процесів. У зв'язку з постійним зростанням виробництва та змінами ринкових умов, логістичні ланцюги потребують вдосконалення як в організації окремих процесів, так і в їх інтеграції з використанням сучасних інформаційних технологій.

Задача формування раціональних логістичних каналів для постачання ресурсів і збуту продукції на виробничих підприємствах потребує вирішення комплексу завдань:

- вибір моделі структури логістичних каналів;
- оптимізація запасів виробничого підприємства;
- оптимізація логістичних процесів для забезпечення визначених запасів;
- управління логістичними ризиками;
- оцінка ефективності логістичних каналів.

2. Моделі логістичних каналів

Розглянемо існуючі методи, які забезпечують вирішення вказаних завдань.

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
o.malyeyeva@khai.edu (Малєєва О.В.); y.v.polupan@student.khai.edu (Полупан Ю.В.);
0000-0002-9336-4182 (Малєєва О.В.); 0009-0009-3030-8448 (Полупан Ю.В.)

Існують кілька основних *моделей логістичних каналів*, які допомагають ефективно організувати процеси постачання та збуту. До них належать централізовані, децентралізовані та змішані моделі. У централізованих моделях всі логістичні операції зосереджені в одному центрі. Виробниче підприємство має контроль над усіма етапами постачання і збуту з одного місця. Централізація дає можливість забезпечити високу ступінь контролю та узгодженість дій. Однак це може призвести до перевантаження центрального складу та затримок у випадку непередбачених обставин. У децентралізованій моделі логістичні операції розподілені між кількома центрами. Це зменшує навантаження на окремі склади і дозволяє їм гнучко реагувати на зміни попиту. Такі моделі краще адаптуються до ринкових змін, але потребують складнішої координації та управління. Змішані моделі поєднують елементи централізованого та децентралізованого підходів. Наприклад, компанія може мати центральний склад для зберігання основних товарів і кілька регіональних складів для швидкої доставки. Це дозволяє користуватися перевагами обох моделей, але вимагає детального планування та управління процесами.

Порівняння переваг і недоліків різних підходів та методів допомагають вибрати найбільш ефективні з них для конкретного виробничого підприємства. Змішані моделі поєднують переваги централізованого та децентралізованого управління, але потребують ретельного планування. Використання методів лінійного програмування, теорії ігор та моделювання на основі подій дозволяє оптимізувати логістичні процеси.

Коли мова йде про *управління запасами*, виділяють декілька підходів до цього процесу. Одним з них є модель EOQ (економічний розмір замовлення), яка дозволяє визначити оптимальний обсяг замовлення для зменшення загальних витрат на зберігання товарів та організацію постачання [1]. ABC-аналіз використовується для класифікації товарних запасів за їх важливістю: товари категорії A – найбільш значущі для бізнесу та мають найбільший вплив на загальні витрати; продукти категорій B і C є менш значущими і суттєвими для облаштування складського приміщення [2]. Ще однією моделлю є метод управління запасами на основі системи Just-In-Time. Цей підхід спрямований на мінімізацію запасів шляхом доставки матеріалів та товарів у потрібний час і в потрібній кількості, що дозволяє знизити витрати на зберігання та підвищити ефективність процесів [3].

3. Централізована модель логістичних каналів

Розглянемо задачу оптимізації запасів виробничого підприємства на основі централізованої моделі логістичних каналів.

Наприклад, підприємство приладобудування має центральний склад, куди надходять усі ресурси для виробництва. За центральним складом закріплено функції зберігання і своєчасного постачання матеріалів у виробничі підрозділи та дистрибуцію готової продукції. Основними елементами цієї структури є центральний склад, виробничі цехи та транспортна система. На центральному складі зберігається сировина та готова продукція. Виробничі цехи отримують ресурси з центрального складу для певних технологічних операцій з виготовлення приладів. Виробничий процес розділений на етапи, організовані для оптимальної обробки ресурсів, які надходять у вигляді матеріалів та комплектуючих і перетворюються у готову продукцію. Транспортна система в контексті оптимізації логістичних процесів на виробничому підприємстві є важливим компонентом, що забезпечує вчасне переміщення ресурсів, комплектуючих і готової продукції. В даному випадку транспортна система є невід'ємною частиною раціональних логістичних каналів постачання та збуту.

Основними параметрами для оптимізації в цьому прикладі є:

- розмір замовлення,

- частота поповнення запасів,
- рівень запасів.

При оптимізації цих параметрів підприємство отримає оптимальні обсяги замовлення сировини для мінімізації загальних витрат, визначить оптимальні часові інтервали для доставки матеріалів на склад та знизить ризик дефіциту за допомогою розрахунку мінімально необхідного запасу для безперервного виробництва.

Для оптимізації розміру замовлення будемо використовувати модель EOQ:

$$EOQ = \sqrt{\frac{2DS}{H}}, \tag{1}$$

де D – це попит на ресурс, тобто кількість матеріалів та комплектуючих, необхідна протягом певного періоду, S – постійні витрати на оформлення та отримання однієї партії приладів, H – витрати на зберігання одиниці ресурсу в центральному складі за певний період.

Далі можемо визначити оптимальну частоту поповнення запасів, використовуючи модель EOQ в поєднанні з інтервалами замовлень. На основі розрахованого обсягу EOQ й річного попиту можна розрахувати часовий інтервал, за яким має оновлюватися запас для уникнення дефіциту:

$$IT = \frac{EOQ}{D} * T, \tag{2}$$

де D – річний попит, T – період виробництва.

Рівень запасу передбачає підтримку певного обсягу ресурсів для мінімізації ризику дефіциту. Для цього можна використовувати модель, яка враховує рівень замовлень і середній час їх виконання.

Визначити мінімальний запас можна за формулою:

$$MINQ = TZ * DZ, \tag{3}$$

де TZ – середній час виконання замовлення, DZ – середній обсяг реалізації.

Модель EOQ дозволяє підприємству ефективно контролювати запаси ресурсів, мінімізуючи витрати, зберігаючи достатній рівень запасів для стабільного виробництва. В рамках централізованої структури логістичних каналів, застосування цієї моделі допомагає уникнути затримок у виробництві через недостатній рівень запасів і оптимізувати розподіл ресурсів.

*Оптимізація л*огістичних процесів має важливе значення для підтримки ефективного постачання ресурсів і збуту продукції. Основні методи включають моделі лінійного програмування для визначення оптимального рішення при плануванні та розподілі ресурсів, що дозволяє зменшити витрати або підвищити продуктивність, враховуючи обмеження та доступні ресурси. Крім того, теорія ігор використовується для оцінки та прийняття рішень у конкурентному середовищі, імітуючи взаємодію між різними учасниками логістичних ланцюгів і вибираючи найкращі стратегії для досягнення оптимальних результатів. Моделювання на основі подій полегшує аналіз і моделювання логістичних процесів шляхом урахування різноманітних подій, які можуть вплинути на ланцюжок поставок, допомагаючи визначити потенційні проблеми та сформулювати стратегії їх вирішення. Такі методи, як AHP і TOPSIS, які включають численні критерії, полегшують прийняття рішень, враховуючи важливість кожного критерію, оцінюючи та порівнюючи різні альтернативи для визначення найкращих рішень для логістичних операцій. У ситуаціях, що характеризуються невизначеністю та неповними даними, методи нечіткої логіки використовуються для допомоги в прийнятті рішень, моделюючи

складні системи, одночасно враховуючи різні елементи, які впливають на логістичні процеси.

Для пом’якшення *ризиків* у логістичних ланцюгах використовуються різноманітні методи управління ризиками, включаючи ідентифікацію ризиків, оцінку їхньої ймовірності та впливу, формулювання планів реагування та постійний моніторинг ризиків [4]. Вони сприяють стабільності логістичних процесів та допомагають мінімізувати несприятливі наслідки від несподіваних ситуацій.

Оцінка економічної *ефективності логістичних каналів* має вирішальне значення для підтримки їх стабільності та конкурентоспроможності. Ключові методи оцінки включають ROI (повернення інвестицій), який оцінює віддачу від інвестицій у логістичні процеси, допомагаючи визначити, наскільки ефективно використовуються ресурси та результати, отримані від інвестицій у логістику. Метод TCO (загальна вартість володіння) дає змогу оцінити загальні витрати, пов’язані з володінням і експлуатацією логістичних каналів, охоплюючи не лише прямі витрати, пов’язані із закупівлею та обслуговуванням, але й непрямі витрати, такі як утримання запасів, транспортування, управління ризиками тощо. За допомогою аналізу витрат можна оцінити співвідношення між витратами та вигодами від впровадження логістичних рішень, допомагаючи в оцінці економічної життєздатності різних стратегій і підходів у логістиці.

Розгортання сучасних інформаційних систем сприяє автоматизації та оптимізації логістичних операцій, забезпечуючи точність і своєчасність даних, одночасно покращуючи комунікацію та ефективне управління ресурсами, мінімізуючи ризики, пов'язані з помилками і затримками в логістичних ланцюгах.

Список літератури

[1] M. R. Asadabadi, A revision on cost elements of the EOQ model, Studies in Business and Economics, Vol. 11: Issue 1, 2016, pp. 5-14. doi: 10.1515/sbe-2016-0001.

[2] P. Mieczysław, P. Zbigniew, Customer profitability calculation based on time-driven abc model, wholesaler case, International Journal of Contemporary Hospitality Management, Vol. 22, Issue 5, 2014, pp. 91-120. doi: 10.1108/09596111011053774.

[3] D. Stojkanović, Z. Petković and L. Ivanov, Just in time as a modern principle of inventory management, KNOWLEDGE – International Journal, Vol. 45, Issue 1, pp. 91-120. URL: https://ikm.mk/ojs/index.php/kij/article/view/5012/4991.

[4] Ю. Полупан, О. Малєєва, Системна модель ризиків та дерева альтернативних рішень з удосконалення логістичного ланцюга виробничого підприємства, Сучасний стан наукових досліджень та технологій в промисловості, (2(28), 2024, С. 133–142. doi: 10.30837/2522-9818.2024.2.133.

Переваги та недоліки використання гібридної роботи

Світлана Назарова[1*†] та Микола Слісаренко[1*†]

[1] *Simon Kuznets Kharkiv National University of Economics, Kharkiv, ave. Nauki, 9-A, 61166 Ukraine*

Аннотація

У роботі визначено переваги та недоліки гібридної роботи (як поєднання дистанційної роботи та традиційної роботи на робочому місці в офісі) для роботодавця в цілому та працівника за результатами морфологічного аналізу даних широкомасштабних опитувань, що виконані світовими аналітичними агенціями.

Ключові слова

Гібридна робота, дистанційна робота, децентралізація виконання роботи у просторі; децентралізація виконання роботи у часі; виконання роботи в звичайному режимі (в офісі)

1. Вступ

Гнучкість діяльності як суб'єктів господарювання так і окремих працівників, є визначальною компетентністю, яка залишається незмінною протягом останніх чотирьох років.

За результатами аналітичних досліджень, що проводять провідні світові агенції [1-4], у 2024 р. у США 62% респондентів працюють повний робочий день в офісі, порівняно з 66% у 2023 році - незначне зниження. Натомість кількість технічних фахівців, що виконують гібридну роботу (2022 р. – 25%, 2023 р. – 26%, 2024 р. – 27%) та працюють повністю дистанційно (2022 р. – 34%, 2023 р. – 7%, 2024 р. – 11%) щороку збільшується.

В Україні 27% працівників (технічних фахівців) наразі працюють у гібридному форматі, що на 1% більше, ніж у 2023 році, а 11% працівників – повністю віддалено, що на 7% більше, ніж у 2023 році [1-2].

У 2017-2020 роках основною формою роботи була робота в офісі. У 2020-2022 роках світ брав участь у вимушеному експерименті з роботи з дому, який познайомив багатьох з дистанційною роботою. Сьогодні панує гнучкість, а гібридна робота - це те, чого прагнуть працівники, незважаючи на те, що більшість працівників у США вже повернулися до офісу на повний робочий день.

На основі вивченого досвіду застосування гібридної роботи, результатів теоретико-практичних досліджень сутності цього явища [1-4], автори дослідження пропонують таке визначення цього поняття (як синтезу двох форм організації праці): «гібридна робота» – це форма організації праці, за якої виконання дистанційної роботи поєднується з виконанням працівником роботи на робочому місці у приміщенні чи на території роботодавця.

Отже, гібридна робота поєднує в собі ознаки дистанційної та традиційної роботи на робочому місці в офісі, а саме передбачає часткову децентралізацію виконання працівником своєї роботи у часі та/або просторі та звичайну роботу в офісі.

Information Systems and Technologies (ISTs-2024), November 26-28, 2024, Kharkiv, Ukraine

* Corresponding author.

† These authors contributed equally.

svitlana.nazarova@hneu.net (С. О. Назарова); slisarenko.mykola@hneu.net (М. В. Слісаренко)

0009-0007-2229-423X (С. О. Назарова); 0009-0000-3874-2112 (М. В. Слісаренко)

2. Методологія дослідження

Останні аналітичні дослідження [1-4] показали, що поєднання двох форм організації праці надає гібридній роботі переваг та недоліків обох форм організації праці, зумовлює посилення деяких з них та появу нових переваг та недоліків такої форми.

Для виявлення всієї множини переваг і недоліків гібридної роботи у даній роботі проаналізовано результати опитувань виконавців гібридної роботи та їх роботодавців та систематизовано отримані провідними міжнародними агенціями результати спостережень[1-4] за допомогою методу морфологічного аналізу. Морфологічний аналіз, на відміну від інших загальнонаукових методів, дозволяє розбиття досліджуваного об'єкта на характеристики та атрибути (рядки морфологічної скриньки). При цьому характеристики повинні бути незалежними одна від одної [5].

У даному досліджені ознаками (атрибутами, характеристиками) гібридної роботи визначено часткову:

- децентралізацію виконання роботи у просторі [6].;
- децентралізацію виконання роботи у часі [6].;
- виконання роботи в звичайному режимі (в офісі).

На рівні роботодавця (компанії) застосування гібридної роботи доцільно розглядати через призму ресурсів, що забезпечують цей процес, а саме: обладнання (робоче місце, виробнича техніка та оснащення); менеджменту працівників.

3. Результати

Морфологічна скринька, за якою визначено переваги компанії та працівників від застосування гібридної роботи наведено в табл. 1.

Таблиця 1
Переваги застосування гібридної роботи працівників

Ознаки гібридної роботи	Обладнання	Менеджмент	Працівники
децентралізація виконання роботи у просторі	Вивільнення певного числа робочих місць, їх площі, обладнання та обслуговування, а отже зменшення відповідних витрат	Більший вибір виконавців з високим рівнем професіоналізму; Застосування ІКТ для комунікацій з виконавцями	Можливість обрати зручний час (день/дні в тижні) для роботи в офісі та/або вдома; Можливість працювати без постійного контролю з боку керівництва та співучасників команди; Можливість обмежити відволікання, спричинене іншими працівниками Можливість самостійно організувати робочий простір
децентралізація виконання роботи у часі		Збільшена продуктивність праці та/або якість результату; Самоорганізація команди і	Участь у різних проектах (у складі різних команд) та гнучкість виборі роботи (проекту, роботодавця); Можливість самостійно організувати робочий процес;

	самоменджмент виконавців;	Поєднання посадових обов'язків, соціальних та особистих потреб; Збільшення задоволеності праці;
виконання роботи в звичайному режимі (в офісі)	Можливість безпосереднього контролю та комунікації з виконавцями в офісі	Можливість зустрітися та комунікувати з керівництвом і колегами в офісі

Дані досліджень [1-2] вкотре показали, що працівники почуваються більш продуктивними, збалансованими та лояльними до своїх компаній, коли вони виконують гібридну роботу. Багато працівників готові йти на жертви, щоб досягти гнучкості у виборі місця роботи: 62% погодилися б на зниження зарплати на 10% і більше, а 4% звільнилися б з роботи, якби не мали змоги працювати повністю віддалено або в гібридному режимі. Кожен четвертий працівник (25%), який працює в звичайному режимі в офісі готовий пожертвувати 15% своєї річної зарплати за гібридний режим роботи.

Разом з зазначеними вище перевагами застосування гібридної роботи для компанії та працівників існують й недоліки такої форми організації праці, які обумовлюють ризик отримання незадовільного результату (табл. 2).

Таблиця 2

Недоліки застосування гібридної роботи для працівників та компанії (роботодавця)

Ознаки гібридної роботи	Обладнання	Менеджмент	Працівники
децентралізація виконання роботи у просторі	Потреба у забезпеченні дистанційних робочих місць працівників, у т.ч. ІКТ та їх обслуговування (підтримка у стані експлуатації)	Складнощі підбору виконавців для робіт	Високий рівень професіоналізму
		Складнощі комунікацій між керівництвом та виконавцями за допомогою ІКТ;	Вміння застосовувати сучасні ІКТ для комунікацій з керівництвом та колегами
		Слабша корпоративна культура: зниження рівня згуртованості та прийняття цінностей компанії (роботодавця), прив'язаності до компанії;	Потреба у ізольованому від інших мешканців дистанційному робочому місці
			Потреба у забезпеченні дистанційного робочого місця
			Відволікання іншими мешканцями
децентралізація виконання роботи у часі		Складність управління дистанційною роботою виконавців, у т.ч. підтримка	Складність взаємодії з колегами та керівництвом;
			Високий рівень самостійності праці,

	вмотивованості працівників	самоменеджменту та відповідальності
	Складнощі контролю за виконанням роботи та продуктивністю виконавців	Інформаційне перевантаження
	Потреба у високому рівні самоорганізація команди і самоменджмент окремих виконавців;	Ненормований графік роботи Поєднання посадових обов'язків, соціальних та особистих потреб;
виконання роботи в звичайному режимі (в офісі)	Зміна методів та інструментів управління	Зміна методів та інструментів виконання роботи

У випадку гібридної роботи усуваються або максимально мінімізуються такі недоліки роботи на традиційному робочому місці в офісі (або на території роботодавця) як: неможливість поєднувати виконання професійних та соціальних обов'язків, задоволення особистих потреб; витрати часу та коштів на проїзд до офісу, неможливість самостійно організовувати робоче місце та робочий процес тощо. Водночас за такої форми організації праці усуваються або максимально мінімізуються такі недоліки дистанційної роботи як: самотність, відсутність спілкування віч-на-віч з іншими співучасниками та керівництвом, відсутність зміни робочого місця, відсутність доступу до службової інформації, яка не розповсюджується мережею, тощо.

4. Висновки

Слід зауважити, що результати морфологічного аналізу вказав на те, що певні обставини можуть бути як перевагами, так і недоліками гібридної роботи (наприклад, застосування ІКТ для комунікацій з колегами та керівництвом, або самоменеджмент). Якщо об'єкт аналізу (працівник, або інший ресурс роботодавця) володіє цією ознакою у повній мірі, то її використання є перевагою для нього, у протилежному випадку – це недолік, бо потребує додаткових зусиль на нарощування цієї ознаки. Тобто подана у роботі класифікація переваг та недоліків гібридної роботи, що виконана за методом морфологічного аналізу, виконана, в першу чергу, для того, щоб окреслити все коло можливих особливостей застосування такої форми організації праці як гібридна, що потребують врахування в процесі її впровадження та управління нею для досягнення запланованого результату.

Література

[1] State of Hybrid Work 2024. Електронний ресурс: режим доступу: https://owllabs.com/state-of-hybrid-work/2023.

[2] State of Hybrid Work 2024. Електронний ресурс: режим доступу: https://owllabs.com/state-of-hybrid-work/2024#current-conditions.

[3] Smith M. 90% of companies say they'll return to the office by the end of 2024—but the 5-day commute is 'dead,' experts say. Електронний ресурс: режим доступу:

https://www.cnbc.com/2023/09/11/90percent-of-companies-say-theyll-return-to-the-office-by-the-end-of-2024.html.
[4] Turner J. 9 Future of Work Trends for 2024. Електронний ресурс: режим доступу: https://www.gartner.com/en/articles/9-future-of-work-trends-for-2024.
[5] Панкратова, Н. Д., Савченко, І. О., Стратегія застосування методу морфологічного аналізу в процесі технологічного передбачення // Наукові вісті НТУУ «КПІ»: науково-технічний журнал, 2009. № 2(64), с. 35-44.
[6] Пушкар О. І. Назарова С. О. Соціально-економічні аспекти дистанційної трудової діяльності // Економіка розвитку. – №4(36).– 2005. С. 331-335.

DOI: https://doi.org/10.30837/IST-2024.P1.025

Intelligent Decision Support System in the Face of Uncertainty and Risks in Project Management

Marina Grinchenko[1] and Mykyta Rogovyi[1]

[1] *National Technical University "Kharkiv Polytechnic Institute," Kyrpychova str. 2, Kharkiv, 61002, Ukraine*

Abstract

In today's world, Scrum is the most popular approach to managing teams in software development. However, there are factors that influence the emergence of project risks, namely: inaccuracies in the formulation of tasks that the team evaluates and a low level of communication culture, as well as shortcomings in the procedures for distributing tasks among team members. An intelligent decision support system under conditions of uncertainty and risk in project management allows the project manager to provide mechanisms to reduce the impact of risks. The purpose of this study is to reduce the impact of risks associated with the perception of tasks from the tracking system and the distribution of tasks among team members within the project sprint to ensure the successful completion of the project. Key factors affecting the effectiveness of teamwork are analyzed and identified. An intelligent decision support system has been developed that integrates data from the tracking system and, using LLM and a stable matching algorithm, provides recommendations for describing project sprint tasks and distributing these tasks among performers.

Keywords

project, risk, project team, intelligent system

1. Introduction

Among the popular approaches to team management in the software development process, the Scrum methodology occupies a special place.

The main tasks of Scrum-based project management are: controlling the execution of tasks within one or more sprints (terms of reference), efficient resource allocation, and prompt coordination of actions in case of unforeseen circumstances to ensure stable team performance [1]. Given the limited time horizon of management, it is of particular importance to study the factors that affect the effectiveness of teamwork. However, even the use of agile methodologies does not completely eliminate the problems associated with the impact of risks on the project team's activities.

The effectiveness of a project team is influenced by a significant number of factors. Among the most important are the following [2]:

- team qualification, which determines the professional level of the participants, their ability to adequately evaluate the sprint tasks, and perform them efficiently and on time;
- errors in the assessment of sprint tasks can affect the allocation of resources, timing, and the final result. The consequences can extend beyond a single sprint, affecting future phases;
- Inaccurate task descriptions or misunderstanding of tasks by developers can lead to incorrect estimates and complications in execution;
- Reduced resources, i.e., planned or unexpected losses of team members, for example, due to illness or vacation, reduce the amount of work performed and may cause delays;

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
marinagrunchenko@gmail.com (M. Grinchenko); nikrogovoy@gmail.com (M. Rohovyi);
0000-0002-8383-2675 (M. Grinchenko); 0000-0002-7902-3592 (M. Rohovyi)

- The internal culture of the team, which includes established rules, task distribution processes, decision-making strategy and internal communication, determines the team's ability to adapt to crisis situations and overcome unforeseen circumstances.

These factors significantly affect the quality of project implementation, so it is important to create a flexible mechanism by which the team can reduce the impact of these risk factors on the effectiveness of achieving project goals.

The purpose of the study is to reduce the impact of risks associated with the perception of tasks from the tracking system and the distribution of tasks among team members within the project sprint to ensure the successful completion of the project.

2. Proposal of conception

An analysis of approaches to managing a project team in the face of risks has revealed that the key factors that increase the risk of project failure [3] are:

- inaccuracies in the wording of the tasks that the team evaluates;
- low level of communication culture;
- shortcomings in the procedures for distributing tasks among team members.

In this paper, the team is considered within a specific project and consists of a project manager, developers, a technical lead who is also involved in the project implementation, and a tester. The paper considers projects that are executed iteratively by the project team, i.e., agreed and approved tasks are broken down into clearly defined tasks to be completed during the project sprint. The project manager is responsible for managing, coordinating the team's work, and communicating with the customer, receiving tasks and clarifying technical requirements. These tasks are recorded in the tracking system and the project team uses these descriptions to select and assign tasks to the sprint. Failure to complete the project sprint tasks is due to inaccurate task descriptions or lack of understanding of the tasks by the developers, which can lead to incorrect estimates and difficulties in implementation.

The developed intelligent decision support system for uncertainty and risk in project management allows the project manager to provide mechanisms to reduce the impact of project risks by improving the task formulations stored in tracking systems. The intelligent system uses artificial intelligence methods, in particular, large language models (LLM) to support decision-making, which are capable of processing textual data, including technical documentation, risk reports, project status reports, etc.

The proposed system is based on a methodology for researching linguistic characteristics for project task management. With this methodology, the project manager chooses the wording options for the task description, which the LLM evaluates and assigns points based on its clarity and comprehensibility. The project manager can use these recommendations to improve the task description of the project team [2]. When creating a project product backlog or project sprint, the project manager has two options for choosing the best task description. To do this, they need to review and reformulate the task descriptions of the sprint in order to reduce project risks and increase the quality of project execution.

The distribution of tasks in a project sprint between performers also affects the failure to complete project sprint tasks, which may be due to a lack of qualified performers or switching performers to other specific tasks. Thus, the task of task distribution arises as a problem of finding stable pairings.

In the proposed intelligent decision support system, to reduce the impact of the risk associated with the unstable distribution of tasks in the project sprint, a technology based on the stable matching algorithm (SOSM) [3] and the efficiency-adjusted deferred decision-making algorithm (EADAM) [4] is proposed. According to the authors' research, the EADAM and SOSM

algorithms are the most efficient, they ensure the distribution of tasks in a sprint, meet the requirements of developers, and are characterized by low computational complexity. Based on these algorithms, the project manager receives a stable comparison based on the prioritization of tasks and the qualifications of performers, which allows him to successfully complete the project sprint tasks.

To assess the impact of the risks of failing to meet the project sprint objectives, the failure rate per unit of time is used. This indicator determines the number of failures per hour, day, or a certain period (sprint). It is used to assess the reliability and productivity of the workflow [5]. Failure rate can be calculated as the ratio of the number of failures that occurred during a certain sprint to the total number of attempts that may fail. The failure rate of an IT project is a key indicator of the quality of sprint or project management. A high failure rate indicates the presence of significant risks that can be caused by insufficient team preparation, inaccurate task description and planning, and the influence of external factors. The use of recommendations by the project manager helps to reduce the failure rate and increase the stability, accuracy of tasks, and ensure the successful completion of the project.

The use of the failure rate also helps to track the dynamics of risk exposure and evaluate the effectiveness of the implemented management measures. With the help of the failure rate, it is possible to assess the effectiveness of the implementation of the proposed intelligent decision support system in the face of uncertainty and risks in project management.

3. Conclusion

The developed intelligent system provides the project team with the necessary information to effectively manage the project under conditions of uncertainty. The system integrates data from the tracking system and, using LLM and a stable matching algorithm, provides recommendations to the project manager on how to describe the project sprint tasks and distribute these tasks among the performers.

The implementation of an intelligent system allows you to create tasks with greater clarity, which will contribute to better team understanding and project execution. By utilizing LLM in the project management workflow, an intelligent system provides a more efficient and effective task management process that increases the productivity of the project team.

References

[1] The Scrum Guide, The Definitive Guide to Scrum: The Rules of the Game, 2024. URL: http://www.scrumguides.org/.

[2] M. Grinchenko, M. Rohovyi, A model for identifying project sprint tasks based on their description, Innovative technologies and scientific solutions for industries, Kharkiv, 2023, no. 4 (26), pp. 33-44. DOI: https://doi.org/10.3 0 837/ITSSI.2023.26.033.

[3] Diebold, F., Aziz, H., Bichler, M. *et al.* Course Allocation via Stable Matching. Bus Inf Syst Eng 6, 97-110 (2014). https://doi.org/10.1007/s12599-014-0316-6

[4] Zhenhua Jiao, Ziyang Shen. School choice with priority-based affirmative action: A responsive solution. Journal of Mathematical Economics. Vol. 92, January 2021, Pages 1-9. https://doi.org/10.1016/j.jmateco.2020.10.007

[5] Jens Schmidt Mitigating risk of failure in information technology projects: Causes and mechanisms Project Leadership and Society Volume 4, 2023, 100097, pp. 1-13. https://doi.org/10.1016/j.plas.2023.100097.

Секція 2.
Математичне та комп’ютерне моделювання у інформаційних системах

DOI: https://doi.org/10.30837/IST-2024.P1.029

Normalized Difference Model of the Descriptor Control System

Andrew Rutkas[1]

[1] *Kharkiv National University of Radio Electronics, Nauky Ave. 14, Kharkiv, 61166, Ukraine*

Abstract

A difference model approximating a semi-linear descriptor control system with continuous time is studied to enable further application of the results in developing neural networks. An algorithm has been developed to transform the implicit difference equation so that, in its linear part, the original regular characteristic pencil of two matrices is converted into a normalized pencil with an identity matrix as the free term and a certain "informational" matrix T associated with the spectral parameter. One of the key steps in the algorithm involves reducing the matrix T to its Jordan form with a special arrangement of Jordan blocks. An example of a descriptor dynamic system in an electrical network is considered.

Keywords

neural network, control system, dynamic system, descriptor system, electrical network

The article examines a semi-linear descriptor dynamic control system. The analysis includes the original system and its evolution law, as well as a transformation to another system whose state and evolution significantly simplify the application of neural networks for predicting the states of the original system. As an example of application, a descriptor system in an electrical network is considered.

1. Normalized difference approximation

A semi-linear descriptor control system is considered, where the state vector function $x(t) \in \mathbb{R}^n, t \in [0, \theta)$ satisfies the implicit differential-algebraic equation

$$\frac{d}{dt}\big(A_0 x(t)\big) + B_0 x(t) = F\big(x(t)\big) + L_0 u(t)\,. \tag{1}$$

For the theory of differential-algebraic equations and their applications, we refer to [1,2]. In equation (1) $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is assumed to be a nonlinear mapping, $u(t) \in \mathbb{R}^l(\forall t)$, L_0 is a matrix of size $n \times l$, $\lambda A_0 + B_0$ is a regular pencil of $n \times n$-matrices, $A_0 \neq \mathbf{0}$. The finite spectrum of the pencil $\sigma_0 = \sigma(A_0, B_0) = \{\lambda_j\}$ consists of its eigenvalues λ_j — the roots of the characteristic polynomial $p_0(\lambda) = det(\lambda A_0 + B_0)$. All other points in the plane $\mathbb{C}$ form the set of regular points $\rho = \rho(A_0, B_0) = \mathbb{C} \setminus \sigma\,(A_0, B_0)$, at each of which the inverse matrix $(\lambda A_0 + B_0)^{-1}$, $\lambda \in \rho$ exists.

For the difference approximation of equation (1) with a constant time step $\Delta > 0$, a set of isolated points $\{\mathrm{t_k}\}$ is chosen in the interval $[0, \theta)$: $t_k = k \cdot \Delta$, $k = 0,1,2, \ldots$ Replacing the differentials in (1) with finite differences at points t_k and denoting $x_k = x(t_k), u_k = u(t_k)$, we obtain the difference equation:

$$\Delta^{-1} A_0 x_{k+1} + B_0 x_k - \Delta^{-1} A_0 x_k = F(x_k) + L_0 u_k. \tag{2}$$

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
andrii.rutkas@nure.ua (Andrew A. Rutkas)
0009-0005-9684-3380 (Andrew A. Rutkas)

The time step $\Delta > 0$ can be chosen small enough such that $\Delta^{-1} > \max\{|\lambda_j|, \forall \lambda_j \in \sigma_0\}$. Then $\pm\Delta^{-1} \in \rho(A_0, B_0)$. Rewrite equation (2) in the form

$$Ax_{k+1} + Bx_k - Ax_k = F(x_k) + L_0 u_k\,, \tag{3}$$

where $A = \Delta^{-1}A_0, B = B_0$. Neural networks [3] are proposed for solving the difference equation (3), which approximates the differential-algebraic equation (1).

The spectrum σ of the characteristic pencil $S(\lambda) = \lambda A + B$ lies inside the unit circle $|\lambda| < 1$. Therefore $\pm 1 \in \rho(A, B)$ and the inverse matrices $D = (B - A)^{-1}$, $(B + A)^{-1}$ exist.
Equation (3) is equivalent to the following equation, normalized to the identity matrix E associated with x_k and the informational matrix $T = DA$ associated with x_{k+1}:

$$Tx_{k+1} + x_k = \Psi(x_k) + Lu_k, \tag{4}$$

where $\Psi(x_k) = DF(x_k), L = DL_0$.

2. Jordan form of the informational matrix *T*

If the matrix T is non-invertible, the permissible initial data x_0 in equation (4) cannot be arbitrary. They must satisfy a specific algebraic constraint linking the components of the vectors x_0, u_0. To derive this constraint, the Jordan form J of matrix T is employed. A Python program was developed to compute an invertible matrix Q that transforms T into its Jordan form with a specific ordering of Jordan blocks:

$$J = QTQ^{-1} = \begin{pmatrix} G & 0 \\ 0 & H \end{pmatrix},\ T = Q^{-1}JQ. \tag{5}$$

The block-diagonal matrix G consists of invertible blocks J_i of dimensions $p_i \times p_i$, while matrix H consists of nilpotent blocks N_j of dimensions $s_j \times s_j$:

$$G = diag\{J_1, J_2, \dots, J_r\}, \sum_{i=1}^{r} p_i = m, \tag{6}$$

$$H = diag\{N_1, N_2, \dots, N_q\}, \sum_{j=1}^{q} s_j = n - m, \tag{7}$$

$$J_i = \begin{pmatrix} \lambda_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \cdots & \lambda_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_i \end{pmatrix},\quad N_j = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}. \tag{8}$$

The special ordering of the blocks is such that: $s_1 \geq s_2 \geq \cdots \geq s_q \geq 1$,

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_r| > 0,\ |\lambda_i| = |\lambda_{i+1}| \Rightarrow p_i \geq p_{i+1},\ \ i \leq r - 1.$$

3. Analysis of the normalized model

Let us introduce projection $n \times n$-matrices V_i with elements v_{kj}^i $(i = 1,2, \dots, n)$ such that each matrix V_i has a single nonzero element $v_{ii}{}^i = 1$ on the main diagonal:

$$v_{kj}^i = \begin{cases} 1, k = j = i \\ 0, |k - i| + |j - i| \neq 0;\ k, j = 1,2, \dots, n \end{cases}$$

The action of the matrix V_i on an n-dimensional state vector preserves the component with index i and nullifies all other components.

Define the projection $n \times n$-matrices

$$P_G = \sum_{i=1}^{m} V_i, \quad P_H = \sum_{i=m+1}^{n} V_i. \tag{9}$$

Using Jordan form (5) equation (4) can be rewritten in the form of the following equation with respect to $y_k = Qx_k$:

$$Jy_{k+1} + y_k = Q\Psi(Q^{-1}y_k) + QLu_k. \tag{10}$$

With the help of projectors (9), equation (10) splits into two equations:

$$Gy_{k+1}^G + y_k^G = P_G Q\Psi(Q^{-1}y_k) + P_G QLu_k, \tag{11}$$

$$Hy_{k+1}^H + y_k^H = P_H Q\Psi(Q^{-1}y_k) + P_H QLu_k, \tag{12}$$

where $y_k^G = P_G y_k, \;\; y_k^H = P_H y_k$.

Due to the invertibility of the $m \times m$-matrix G equation (11) is transformed into an explicit difference expression for the vector y_{k+1}^G in terms of the full vector y_k at the previous time step k :

$$y_{k+1}^G = -G^{-1}y_k^G + G^{-1}P_G Q\Psi(Q^{-1}y_k) + G^{-1}P_G QLu_k. \tag{13}$$

To analyze equation (12), we construct a self-adjoint projection $n \times n$-matrix $P_0 = P_0^*$ onto the kernel of the matrix J^* such that $J^*P_0 = 0 \;\sim\; P_0 J = 0$.

From the structure of the nilpotent blocks N_j, it follows that

$$P_0 = \sum_{j=1}^{q} V_{\alpha_j}, \quad \alpha_j = m + s_1 + s_2 + \cdots + s_j. \tag{14}$$

Introducing the additional projection matrix $P_1 = P_H - P_0$, and applying the projectors P_1, P_0 to equation (12), we split (12) into two equations. It can be verified that $P_1 H = H, \;\; P_1 P_H = P_1,$ $P_0 H = 0, \;\; P_0 P_H = P_0$. Consequently, equation (12) is equivalent to the equations:

$$P_1 H P_H y_{k+1} = -P_1 y_k + P_1 Q\Psi(Q^{-1}y_k) + P_1 QLu_k \tag{15}$$

$$0 = -P_0 y_k + P_0 Q\Psi(Q^{-1}y_k) + P_0 QLu_k. \tag{16}$$

For a given control u_k, the set of admissible n-dimensional vectors y_k, satisfying the algebraic relationship (16), forms a manifold $\Lambda_k = \{y_k\}$ in the space $\mathbb{R}^n$ or in the complex space $\mathbb{C}^n$ if there exist complex spectrum points λ_i of the matrix T. By construction, the matrices P_G, P_1, P_0 are mutually orthogonal projection matrices and $P_G + P_1 + P_0 = E$.

Let $y_k \in \Lambda_k$ be a known state. Is it possible, based on y_k and the controls u_k, u_{k+1} to obtain the state y_{k+1}, that satisfies equation (10)? It is sufficient to find the three projections in the decomposition of the vector y_{k+1}:

$$y_{k+1} = P_G y_{k+1} + P_1 y_{k+1} + P_0 y_{k+1}. \tag{17}$$

The first two projections are uniquely determined through the difference relations (13), (15). Let us represent the algebraic relation (16) for the moment *k+1* as:

$$P_0 y_{k+1} - P_0 Q\Psi(Q^{-1}y_{k+1}) = P_0 QLu_{k+1}. \tag{18}$$

In the right-hand side of equality (17), the first two terms at the moment *k+1* are replaced by their already known representations (13), (15) through the data from the preceding moment k, and the obtained representation for y_{k+1} is substituted into the term with nonlinearity Ψ in (18). As a result, an algebraic equation is obtained for the projection $P_0 y_{k+1}$.

If the original mapping F in (1), (2), (3) is linear, then Ψ (4) is also linear, and $P_0 y_{k+1}$ is uniquely determined from the relation (18).

For the nonlinear mapping Ψ, restrictions depending on the type of nonlinearity must be imposed. Under 'favorable' properties of the mapping Ψ the outlined procedure guarantees the existence and uniqueness of the desired projection $P_0 y_{k+1}$, and thus the existence of a recurrent mapping

$$\Phi_k \colon \Lambda_k \times \mathbb{R}^l \times \mathbb{R}^l \to \Lambda_{k+1}, \;\; y_{k+1} = \Phi_k(y_k, u_k, u_{k+1}). \tag{19}$$

It follows from (19) and the relation $y_k = Qx_k$ that for the sequence of states $\{x_0, x_1, \dots, x_k, x_{k+1}, \dots\}$ of difference equation (2) or (3) or (4) there exist the recurrent mappings

$$x_{k+1} = M_k(y_k, u_k, u_{k+1}) = Q^{-1}\Phi_k(Qx_k, u_k, u_{k+1}). \tag{20}$$

Next, we will demonstrate how the proposed approach to the study of nonlinear differential-algebraic equations is applied to the controlled radio engineering system from Section 4. This approach can also be applied to other classes of radio engineering systems, for example, from works [4,5].

4. Example. Transient state equations of an electrical network

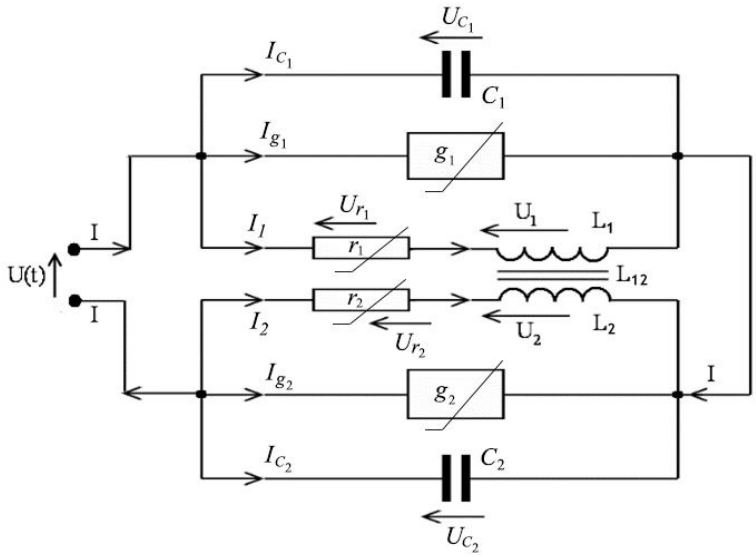


Figure 1. A two-terminal network with nonlinear resistances r_j and conductances $g_j, j = 1,2$.

The electrical network in Fig. 1 represents a two-terminal network with a given input voltage $U(t)$ on the pair of input terminals, two capacitances C_j, two nonlinear conductances g_j, inductances L_1, L_2, mutual inductance L_{12}, and two nonlinear resistances $r_j, j = 1,2$.

The currents, voltages, and element parameters of the network satisfy the relationships:

$$\left.\begin{aligned} I_{C_j} = C_j \frac{dU_{C_j}}{dt}, \quad I_{g_j} = g_j(U_{C_j}); \quad U_{r_j} = r_j(I_{r_j}) \\ U_1 = L_1 \frac{dI_1}{dt} + L_{12} \frac{dI_2}{dt}; \quad L_j > 0, \quad L_{12} > 0 \\ U_2 = L_{12} \frac{dI_1}{dt} + L_2 \frac{dI_2}{dt}; \quad L_{12}^2 = L_1 L_2; \quad L_1 \neq L_2 \end{aligned}\right\} \tag{21}$$

The currents through the inductances and *the voltages across the capacitors* are chosen as the dynamic state variables of the network. For the network in Fig. 1 the state vector $x(t)$ is four-dimensional:

$$x(t) = \left(I_1(t), I_2(t), U_{C_1}(t), U_{C_2}(t)\right)^{tr} \tag{22}$$

while the total number of currents and voltages across all nine branches of the circuit equals 18. In general, the complete system of Kirchhoff's equations can be written, for instance, using the fundamental cycle and cut-set matrices of the network graph. For the model network, this results in nine Kirchhoff equations. By eliminating the external current I, the currents I_{r_j}, and the voltages U_{g_j}, the following four conservation law equations are obtained ($j = 1,2$):

$$U_j - U_{C_j} = -U_{r_j}, U_{C_1} - U_{C_2} = U(t),\ I_{C_1} + I_{C_2} + I_1 + I_2 = -I_{g_1} - I_{g_2}. \tag{23}$$

Substituting the expressions for $U_j, I_{C_j}, U_{r_j}, I_{g_j}$ from (21), we obtain the desired system of differential-algebraic equations with respect to the dynamic variables (22):

$$\left.\begin{aligned} &L_1\frac{dI_1}{dt} + L_{12}\frac{dI_2}{dt} - U_{C_1} = -r_1(I_1)I_1 \\ &L_{12}\frac{dI_1}{dt} + L_2\frac{dI_2}{dt} - U_{C_2} = -r_2(I_2)I_2 \\ &U_{C_1} - U_{C_2} = U(t) \\ &C_1\frac{dU_{C_1}}{dt} + C_2\frac{dU_{C_2}}{dt} + I_1 + I_2 = -g_1(U_{C_1})U_{C_1} - g_2(U_{C_2})U_{C_2} \end{aligned}\right\} \tag{24}$$

The system of differential-algebraic equations (24) takes the vector form (1) with respect to the state vector $x(t)$ (22), where:

$$A_0 = \begin{bmatrix} L_1 & L_{12} & 0 & 0 \\ L_{12} & L_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & C_1 & C_2 \end{bmatrix}, B_0 = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & 0 & 0 \end{bmatrix}, F(x) = \begin{bmatrix} -r_1(x_1)x_1 \\ -r_2(x_2)x_2 \\ 0 \\ -g_1(x_3)x_3 - g_2(x_4)x_4 \end{bmatrix}, L_0u = \begin{bmatrix} 0 \\ 0 \\ U(t) \\ 0 \end{bmatrix}.$$

Both matrices A_0, B_0 are singular: $det\, A_0 = 0,\ det\, B_0 = 0$. At the same time $rang A_0 = 2$ due to the parameter relationships in (21). The characteristic polynomial in this case is

$$p_0(\lambda) = det(\lambda A_0 + B_0) = a \cdot \lambda,\ a = \left(\sqrt{L_1} - \sqrt{L_2}\right)^2. \tag{25}$$

Thus, under the condition $L_1 \neq L_2$ the pencil $\lambda A_0 + B_0$ is regular with a single eigenvalue of 0, and for any $\lambda \neq 0$ the inverse matrix $(\lambda A_0 + B_0)^{-1}$ exists.

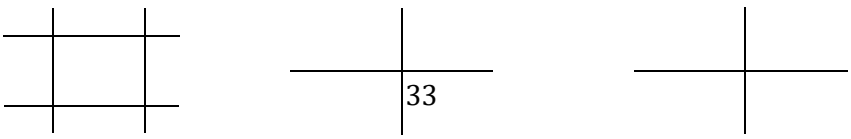
To illustrate, let us assign numerical values to the parameters of the circuit's inertial elements—inductances and capacitances:

$$L_1 = 4, L_{12} = 2, L_2 = 1, C_1 = 1, C_2 = 2.$$

The matrix $S = B_0 - A_0$ is invertible, with the inverse $S^{-1} = D$. Since $\pm 1 \in \rho(A_0 B_0)$, it is convenient to choose a time step $\Delta = 1$ when discretizing equation (1). Thus, the difference model of type (2) for equation (1) immediately takes the form (3), where $A = A_0, B = B_0$. The matrices $D = S^{-1}, T$ in the normalized form (4) of equation (1) are expressed as follows:

$$D = \begin{bmatrix} -4 & 7 & -5 & -1 \\ 7 & -13 & 9 & 2 \\ 1 & -2 & 2 & 0 \\ 1 & -2 & 1 & 0 \end{bmatrix}, T = \begin{bmatrix} -2 & -1 & -1 & -2 \\ 2 & 1 & 2 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Calculations using our Python program yielded the following results for transforming the matrix T into its Jordan form $J = QTQ^{-1}$:



$$J = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, Q = \begin{bmatrix} -2 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}, Q^{-1} = \begin{bmatrix} -1 & -1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The specific order of Jordan block arrangement in (6), (7) is preserved:

$$G = J_1 = -1, N_1 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, s_1 = 2, N_2 = 0, s_2 = 1.$$

Thanks to this, the required projectors take the form of the following diagonal matrices:

$$P_G = P_{J_1} = diag\{1\ 0\ 0\ 0\}, P_H = diag\{0\ 1\ 1\ 1\}, P_0 = diag\{0\ 0\ 1\ 1\}, P_1 = diag\{0\ 1\ 0\ 0\}.$$

Conclusion

The traditional approach to studying the descriptor system described by a linear or semi-linear differential-algebraic equation depends on the specific fixed index of the characteristic matrix pencil of the linear part of the equation (1). A key feature and versatility of our method is its independence from this specified index. It is also worth noting the potential use of the method in problems of conflict control and pursuit games.

Acknowledgement

The author would like to thank Associate Professor Oleh Storchak for his attention to this article.

References

[1] R. Riaza, Differential-Algebraic Systems. Analytical Aspects and Circuit Applications, World Scientific Publishing, Basel, 2008.

[2] S. Campbell, A. Ilchmann, V. Mehrmann, T. Reis, Applications of Differential-Algebraic Equations: Examples and Benchmarks, Springer, Dordrecht, 2019.

[3] C.C. Aggarwal, Neural Networks and Deep Learning, Springer Cham, 2018.

[4] L.A. Vlasenko, A.A. Rutkas, A.G. Rutkas, A.O. Chikrii, Stochastic descriptor pursuit game, Cybernetics and Systems Analysis 60, 2024, 433–441. https://doi.org/10.1007/s10559-024-00684-5

[5] A. Chikrii, A. Rutkas, L. Vlasenko, On descriptor control impulsive delay systems that arise in lumped-distributed circuits, Advanced Control Systems: Theory and Applications, River Publishers, 2024, pp. 3–19. ISBN 978-877022341-6

DOI: https://doi.org/10.30837/IST-2024.P1.035

Нечітка модель прийняття рішень для сортування пошти

Ігор Гребеннік[1†] та Олексій Коваленко[1†]

[1] *Харківський Національний Університет Радіоелектроніки, 14, просп. Науки., Харків, 61166, Україна*

Анотація

У роботі представлено дослідження можливості реалізації заданої логіки сортування посилок, що обробляються автоматизованими сортувальними лініями. Пропонується використовувати нечітку модель прийняття рішень, що дозволяє класифікувати посилки за їх параметрами ваги та габаритів. Класифіковані посилки розподіляються із заданою логікою за дверями терміналів центрів сортування посилок, для яких визначено діапазони значень ваги та габаритів.

Ключові слова

Модель прийняття рішень, автоматизована сортувальна лінія, транспортування, нечітка логіка.

1. Вступ

Глобальний розвиток електронної комерції обумовлює постійно зростаючу складність логістики, зв’язаної з доставкою пошти. Одним зі шляхів підвищення ефективності обробки посилок є використання промислової автоматизації у вигляді автоматизованих сортувальних ліній (ASL – Automated Sorting Line) зі стрічковими конвеєрами-транспортерами [1]. Такими лініями оснащуються центри сортування посилок (PSC – Parcel Sorting Centres), які є складовими пунктами мереж доставки логістичних компаній США та Європи [2, 3].

Автоматизована сортувальна лінія (АСЛ) – це стаціонарна автоматизована комп'ютерна система, що здійснює переміщення та сортування посилок до дверей терміналів PSC у відповідності до адреси доставки, яка зберігається в QR-коді ярлика. ASL не вимагає контролю з боку людини, лише її обслуговування.

Недоліки ASL пов’язані з їхньою «вузькою» групою завдань – це завдання сортування та переміщення посилок для подальшого завантаження. Одним з недоліків застосування ASL є обмеження її моделі прийняття рішень, пов'язаних з невизначеністю умов сортування. Ця невизначеність обумовлена тим, що стандарти визначення категорій посилок у різних фірмах відрізняються. Інший недолік обумовлений тим, що сортування посилок проводиться без урахування їх ваги та габаритів. Це може призвести до неефективного використання об'єму кузова вантажівок при їх завантаженні, а також ризику пошкодження посилок, коли вони розміщуються одна на одній.

Для усунення зазначених недоліків модель прийняття рішень ASL потребує управління сортуванням посилок на принципах нечіткої логіки. Для цього потрібно розробити модель прийняття рішень ASL, яка вирішує завдання нечіткої класифікації та сортування посилок з урахуванням їх параметрів ваги та габаритів з метою компактного завантаження та збереження цілісності об'єктів поштових відправлень.

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
† These authors contributed equally.
igor.grebennik@nure.ua (І. Гребеннік); oleksii.kovalenko3@nure.ua (О. Коваленко)
0000-0003-3716-9638 (І. Гребеннік); 0009-0008-4779-6161 (О. Коваленко)

2. Постановка задачі

Розроблювана нечітка модель прийняття рішень повинна реалізувати задану логіку сортування посилок за дверима терміналів PSC. З метою збереження посилок логіка їх сортування має проводитися з урахуванням ваги та габаритів (ширини, висоти, глибини) і забезпечувати визначені варіанти компактного завантаження.

3. Визначення варіантів компактного завантаження та логіки сортування

Будемо вважати, що PSC, оснащений одним терміналом розвантаження вантажівок і двома терміналами завантаження посилок, які позначаються літерами "A" і "B". Кожен з них обладнується завантажувальними дверима "L1", "L2", ..., "LN". Розвантажені посилки подаються до ASL. Модель прийняття рішення ASL отримує адресу посилки за ярликом і визначає, на який термінал вона транспортується. Мета сортування – транспортувати посилку до однієї з завантажувальних дверей терміналу з певним номером.

Для завантаження визначені два варіанти:

- За першим варіантом реалізується послідовне завантаження вантажівок на дверях "LN", "LN-1", ..., "L1" відповідно до зменшення значень діапазонів ваги та габаритів посилок. Тим самим забезпечується можливість компактного завантаження кузова вантажівки та знижується ризик пошкодження посилок при їх розміщенні одна на одній.
- За другим варіантом реалізується одночасне завантаження N вантажівок на дверях "L1", "L2", ..., "LN". Це забезпечує їх компактне завантаження посилками одного діапазону ваги і теж знижує ризик їх пошкодження.

Для забезпечення визначених варіантів завантаження модель прийняття рішень має реалізувати задану логіку сортування, що визначається такими умовами:

- Для завантажувальних дверей терміналів визначаються діапазони значень ваги та габаритів посилок, які збільшуються відповідно до нумерації завантажувальних дверей "L1", "L2", ..., "LN".
- Завдання моделі прийняття рішень ASL провести нечітку класифікацію посилок за їх вагою та габаритами та визначити номер завантажувальних дверей, на яку вони повинні транспортуватися.

4. Розробка нечіткої моделі прийняття рішень ASL

Нечітка модель прийняття рішень має п'ять входів та один вихід. Вхідними чіткими параметрами є: e* – вага h* – висота, w* – ширина, d* – глибина посилки. Вихідним чітким параметром моделі є номер завантажувальної двері y*, що визначається результатом нечіткої класифікації. Визначимо нечіткі лінгвістичні змінні {E, H, W, D} у загальному вигляді. Нечітка лінгвістична змінна X задається у вигляді кортежу

$$X =< I_X, R(X), U >=\{< x_1, U, X_1 >, < x_2, U, X_2 >, ..., < x_N, U, X_N >\}, \quad (1)$$

де в якості X розглядається нечітка лінгвістична змінна з {H, W, G, V}; I_X – найменування X; x* – вхідний параметр з {h*,w*,g*,v*}; U – область визначення (універсум) всіх нечітких змінних, значення яких набуває X; $R(X)=\{R_{Xi}\}$, i=1,2,...,N – терм-множина X; кортежі $<x_i,U,X_i>$ – нечіткі змінні, значення яких приймає X. Кортеж нечіткої змінної можна подати у вигляді

$$< x_i, U, X_i > =< R_{xi}, \{x_i \mid \min x_i < x_i < \max x_i\}, X_i \in U, \ x_i = (x^*, \mu_{Xi}(x^*)) >, \quad (2)$$

де X_i - нечітка множина; min x_i, max x_i – граничні значення нечіткої множини X_i; $\mu_{Xi}(x^*)$ – функція приналежності чіткої змінної x* до нечіткої множини X_i; R_{Xi} - терм з індексом "i".

Для кожної завантажувальної двері з індексом n=1,2,...,N задаються діапазони значень ваги, висоти, ширини і глибини у вигляді нечітких множин $\{E_q, H_j, W_m, D_s\}$ з термами $\{R_{Ejq}=\{R_{Hj}\}=\{R_{Wm}\}=\{R_{Ds}\}=\{"1","2", ...,"N"\}$ і пов'язаних з ними індексами q, j, m, s=1,2,...,N нечітких множин.

Для нечіткої класифікації використовується алгоритм Такагі-Сугено-Канга [3]. З урахуванням того, що нечітка база правил є синглтонною, визначимо чіткі числові значення вихідної лінгвістичної змінної Y набором

$$Y =< I_Y, R(Y), U >= \{< y_1, U, 1 >, < y_2, U, 2 >, ..., < y_N, U, N >\}, \quad (3)$$

де: I_Y – найменування Y; $R_{Yi}=\{"1","2", ...,"N"\}$ – терм-множина Y; $<y_i,U,i>$ – кортежі чітких змінних, значення яких відповідають номерам дверей $\{y_1, y_2,..., y_N\}= \{1,2,...,N\}$.

Представимо базу правил системи нечіткого виведення моделі у вигляді K правил із чотирма умовами для кожної:

$$\left\{[\underset{k=1}{\overset{N}{\text{Rule}}}\ k]: \text{IF}(e^* \in E_q) \text{ and IF}(h^* \in H_n) \text{ and IF}(w^* \in W_m) \text{ and IF}(d^* \in D_s) \rightarrow Y = y_k, \right. \quad (4)$$

де k = 1, 2, ..., K – індекс правила. Для нечітких множин індекси в правилах вибираються незалежно від індексу k. Під операцією визначення приналежності чіткої змінної x* до множини X_i в умовах правил (4) розуміється знаходження значення функції приналежності $\mu_{Xi}(x^*)$. Для нечітких множин операція AND відповідає операції знаходження перетину або мінімізації. Перепишемо правила (4) в іншому вигляді

$$[\underset{k=1}{\overset{N}{\text{Rule}}}\ k]: \min(\mu_{Eq}(e^*), \mu_{Hj}(h^*), \mu_{Wm}(w^*), \mu_{Ds}(d^*)) \rightarrow Y = y_k \quad (5)$$

або

$$[\underset{k=1}{\overset{N}{\text{Rule}}}\ k]: \mu_{Y_k}(y_k) \rightarrow Y = y_k, \quad (6)$$

де $\mu_{Yk}(y_k)=\min(\mu_{Eq}(e^*), \mu_{Hj}(h^*), \mu_{Wm}(w^*), \mu_{Di}(d^*))$ – узагальнене позначення результату активізації k-го правила. Для розрахунку чіткого значення y* використовується вираз для розрахунку центру тяжіння для синглтонних множин.

$$y^* = \left(\sum_{k=1}^{K} y_k \cdot \mu_{Yk}(y_k)\right) / \left(\sum_{k=1}^{K} \mu_{Yk}(y_k)\right). \quad (6)$$

Логіка сортування посилок реалізується завданням індексу k консеквенту y_k в правилах (4) та (5) залежно від умов приналежності параметрів {e*, h*,w*, d*} до нечітких множин $\{E_q, H_j, W_m, D_s\}$ з індексами q, j, m, s. Алгоритм завдання консеквенту складається з двох критеріїв. За першим критерієм контролюється вага. Для цього визначається індекс q за умовою $e^* \in E_q$. Якщо в інших умовах не більше одного індексу з {j, m, s}, які перевищують q, то індекс консеквенту y_k визначається як "k=q" і посилка транспортується до завантажувальних дверей з номером "q". Другий критерій застосовується для сумісного контролю ваги та габаритів. Якщо в умовах два і більше індексів з {j, m, s} перевищують індекс q, то індекс консеквенту "k=q+1", і посилка транспортується до завантажувальних дверей з номером "q+1".

Для тестування розробленої моделі використовувалось програмне забезпечення середовища Matlab. Розроблено модель з чотирма входами (рис. 1), які відповідають чітким значенням параметрів {e*, h*, w*, d*} і одним виходом – чітким значенням номерів дверей терміналів. Тестування проводилося для терміналів із трьома

завантажувальними дверями. Значення діапазонів нечітких множин, заданих для моделювання, подані в табл. 1.

Таблиця 2
Діапазони значень нечітких множин

Лінгвістичні змінні	Нечіткі множини	Діапазони значень
E	E_1, E_2, E_3	$E_1 \in [0,30]$, $E_2 \in [15,45]$, $E_3 \in [35,60]$ кг
H	H_1, H_2, H_3	$H_1 \in [0,80]$, $H_2 \in [40,120]$, $H_3 \in [80,160]$ см
W	W_1, W_2, W_3	$W_1 \in [0,80]$, $W_2 \in [40,120]$, $W_3 \in [80,160]$ см
D	D_1, D_2, D_3	$D_1 \in [0,80]$, $D_2 \in [40,120]$, $D_3 \in [80,160]$ см

Результати моделювання представлені у вигляді графіків функції відгуку моделі для двох варіантів. Для першого варіанта на рис. 2 представлена функція відгуку, що визначає номер завантажувальних дверей (y*) для двох вхідних параметрів висоти (h*) та ваги (e*), а на другого – для параметрів ваги (e*) та ширини (w*).

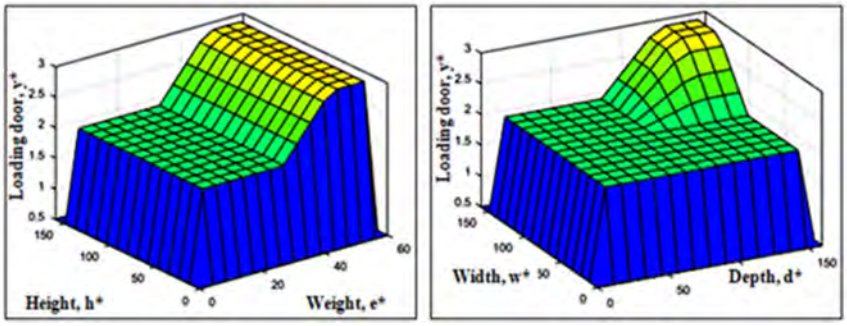


Рисунок 1. Функції відгуку нечіткої моделі

Аналіз отриманих функцій відгуку нечіткої системи виведення моделі дозволяє зробити висновок, що усі критерії сортування реалізовані, а розроблена нечітка модель прийняття рішень ASL повністю відповідає поставленій задачі.

5. Висновки

Розроблена нечітка модель прийняття рішень ASL повністю відповідає поставленій задачі, реалізує нечітку класифікації та сортування посилок за завантажувальними дверями терміналів. Реалізоване сортування забезпечує два варіанти компактного завантаження посилок.

Література

[1] ICONVEY® Sortation Conveyor, 2024. URL: www.iconveytech.com/products/sortation-conveyor/.
[2] D.L. McWilliams, P.M. Stanfield, C.D. Geiger, "The parcel hub scheduling problem: A simulation-based solution approach." Computers & Industrial Engineering, 49(3): 393–412, 2005.
[3] Гребеннік І.В., Коваленко О.А. "Нечітка модель прийняття рішень для автоматичної сортувальної лінії пошти." АСУ і прилади автоматики, 180: 16–26, 2024.

DOI: https://doi.org/10.30837/IST-2024.P1.039

Використання нечітких систем логічного виводу для математичних моделей оцінки надійності програмного забезпечення

Юрій Здоренко[1], Аліна Янко[1], Олена Гайтан[1], Сергій Любарський[2] та Анатолій Мартиненко[2]

[1] Національний університет «Полтавська політехніка імені Юрія Кондратюка», Першотравневий проспект, 24, Полтава, 36011, Україна

[2] Військовий інститут телекомунікацій та інформатизації імені Героїв Крут, вул. Князів Острозьких, 45/1, м. Київ, 01011, Україна

Анотація

Розвиток та інтеграція інформаційних систем в усі сфери діяльності сучасного суспільства потребує забезпечення надійності відповідного програмного забезпечення. Задачею дослідження є знаходження шляхів для доповнення математичних моделей прогнозування дефектів програмного забезпечення на основі використання нечіткої логіки.

Ключові слова

Надійність, математична модель, дефекти, нечіткі системи

1. Вступ

Надійність програмного забезпечення є важливою характеристикою та регламентується вимогами міжнародного стандарту ISO 25010 [1]. Дані про виявлені дефекти можуть характеризуватися різною частотою розподілу на різних етапах життєвого циклу [2], [3]. Так, під час етапу верифікації ПЗ кількість виявлених дефектів може мати найбільшу частоту, поступово зменшуючись на подальших етапах функціонування ПЗ. Вартість усунення дефектів переважно більша при виявленні їх на пізніх етапах життєвого циклу. Існуючі підходи оцінювання характеристик надійності програмного забезпечення здебільшого базуються на поєднанні різних математичних моделей оцінки дефектів, а також, на сучасному етапі розвитку інформаційних технологій, можуть бути доповнені з використанням підходів на основі нечіткої логіки [4]. Тому існує потреба в адекватних математичних моделях прогнозування можливих дефектів в ПЗ, які будуть забезпечувати необхідний рівень точності.

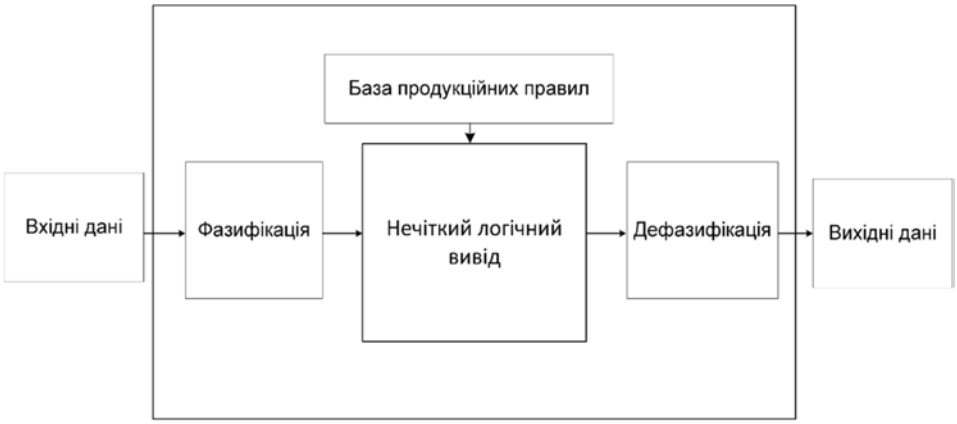
2. Постановка задачі

Забезпечення надійності ПЗ на основі завчасної оцінки кількості можливих дефектів в ПЗ можливе з використанням систем прогнозування. Існуючі математичні моделі призначені для вирішення цієї задачі не завжди дозволяють врахувати складні причинно-наслідкові зв'язки виникнення нових дефектів та потребують доповнення [5-9].

Одним з перспективних напрямків реалізації систем прогнозування дефектів в ПЗ є використання нечітких систем логічного виводу. Такі системи показали свою

zdorenkoviti@gmail.com (Ю. Здоренко); al9_yanko@ukr.net (А. Янко); azalie@ukr.net (О. Гайтан)
0000-000L-5649-771X (Ю. Здоренко); 0000-0003-2876-9316 (А. Янко); 0000-0002-7228-9937 (О. Гайтан); 0000-0001-8068-1106 (С. Любарський); 0000-0002-9576-0138 (А. Мартиненко)

ефективність при неповноті та неточності вхідних даних [5-9]. Для налаштування таких систем можуть використовуватися різні підходи. Узагальнена структура нечіткої системи логічного виводу представлена на рис.1 [5].



Малюнок 1. Основні етапи нечіткого виводу

Математична модель на її основі буде мати такі етапи функціонування: збір даних про виявлені дефекти ПЗ, фазифікація, нечіткий логічний вивід, дефазифікація, отримання вихідних даних [5].

3. Прогнозування дефектів ПЗ з використанням моделі на основі нечіткої логіки

Проведений аналіз літературних джерел показав необхідність в удосконаленні моделей прогнозування дефектів ПЗ. Одним з перспективних шляхів для цього є використання нечітких систем логічного виводу. Навчання таких систем пропонується здійснювати на основі статистичних даних моніторингу всіх етапів життєвого циклу зразка ПЗ.

Процес моніторингу пропонується здійснювати шляхом підрахунку кількості дефектів ПЗ на дискретній ділянках часу певної тривалості, Дані про кількість дефектів на таких ділянках складають статистичний набір, який буде використаний для подальшого налаштування. Моніторинг для виявлення дефектів необхідно здійснювати протягом всього життєвого циклу ПЗ. Такий підхід дозволить зібрати необхідні дані для подальшої роботи моделі.

Так, в якості вхідних величин пропонується використати значення кількості дефектів ПЗ протягом певної кількості попередніх часових інтервалів. Кількість таких вхідних величин добирається виходячи з вимог до точності прогнозу та може бути різною. Також, додатково, в якості вхідних величин можуть бути застосовані додаткові параметри або усереднюючі характеристики. Вихідною величиною прийнято – кількість дефектів ПЗ протягом прогнозного часового інтервалу такої ж тривалості, як і для вхідних величин. В загальному на модель на основі нечіткої логіки впливає: вибір алгоритму нечіткого виводу (Mamdani, Tsukamoto, Sugeno, Larsen та інші); вибір вхідних величин та їх кількості; вибір кількості функцій належності для кожної вхідної величини та їх форми; вибір вихідної величини; розробка бази правил та визначення значущості (ваги) кожного правила; вибір алгоритму налаштування (навчання) моделі на попередньо зібраних даних.

Для отримання необхідного ступеня точності прогнозу пропонується здійснювати певну множину операцій направлених на знаходження балансу для досягнення необхідної точності прогнозу з однієї сторони та забезпечення вимог по продуктивності моделі з іншої. Виходячи з цього пропонується здійснювати дослідження при варіації такими параметрами моделі: зміна алгоритму нечіткого логічного виводу; різна кількість

вхідних величин; різна тривалість інтервалу прогнозування; різна кількість та форма функцій належності вхідних величин.

Так, коректному добору підлягають значення тривалості дискретних часових інтервалів. Також важливою складовою є кількість нечітких правил та їхня вага, яка по замовчуванню може бути однаковою. Для автоматизованого налаштування нечітких систем пропонується використовувати підхід з використанням ANFIS [10-15], що дозволить автоматизувати процес створення бази правил та визначення ваги кожного правила. Такі системи є симбіозом нечітких систем логічного виводу та нейронних мереж. Налаштування таких систем зводиться до використання класичних алгоритмів навчання нейронних мереж [16]. Для проведення необхідних досліджень пропонується використовувати середовище Matlab, а саме призначений для створення нечітких моделей пакет Fuzzy Logic Toolbox [17].

4. Висновки

В роботі запропоновано використовувати підходи на основі нечіткої логіки з автоматизованим налаштуванням для доповнення математичних моделей оцінки надійності ПЗ. Для цього пропонується використовувати структуру на основі ANFIS. Визначено шляхи досягнення необхідної точності прогнозу дефектів ПЗ з використанням запропонованих підходів.

Наукова новизна отриманих результатів полягає в удосконаленні математичних підходів для створення моделей прогнозування, в яких на відміну від існуючих пропонується змінювати необхідну кількість вхідних параметрів для досягнення балансу між точністю та продуктивністю, що дозволяє отримати необхідне рішення для різних умов та обмежень. Моделі які можуть бути реалізовані з використанням вищезазначеного підходу, підлягають легкому масштабуванню для досягнення необхідної точності прогнозу та продуктивності.

Практична значимість отриманих результатів полягає в тому, що з використанням запропонованих підходів є можливість отримати необхідну модель на основі FIS для знаходження значення показника надійності ПЗ з необхідною точністю. Це в кінцевому результаті дозволяє забезпечити кращі значення якості ПЗ.

Література

[1] ISO/IEC 25010:2024, Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Software Product and System Quality, ISO/IEC JTC1/SC7/WG6 (2024).

[2] ISO/IEC 12207:2017, Systems and software engineering – Software life cycle processes, ISO/IEC/IEEE (2017).

[3] ISO/IEC/IEEE 26531:2023, Systems and software engineering – Content management for product life cycle, user and service management information for users, ISO/IEC/IEEE (2023).

[4] Takagi, T, Sugeno, M.: Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Systems, Man, and Cybernetics 15(1), 116–132 (1985). https://doi.org/10.1109/TSMC.1985.6313399

[5] Memon, A. M., Mujeeb-u-Rehman, M. B., Memon, M., Musavi, S. H.: A Regression Analysis Based Model for Defect Learning and Prediction in Software Development. Mehran University Research Journal of Engineering and Technology 40 (3), 617-629 (2021). https://doi.org/10.22581/muet1982.2103.15

[6] Oyo-Ita, E. U., Edim E. A., Otiko, A., Izuki, D. E.: Improving model performance for software defect detection and prediction using ensemble methods and cross-validation technique.

International Journal of Science and Research Archive 12(02), 2363–2373 (2024). https://doi.org/10.30574/ijsra.2024.12.2.1518
[7] Ponochovniy, Y., Bulba, E., Yanko, A., Hozbenko, E.: Influence of diagnostics errors on safety: Indicators and requirements. In: 9th International Conference on Dependable Systems, Services and Technologies (DESSERT), pp. 53–57. IEEE, Kyiv, Ukraine (2018). https://doi.org/10.1109/DESSERT.2018.8409098
[8] Jailani, M. A., Bhakta, K.: Advancements in Predictive Models for Software Defects: A Comprehensive Exploration. International Journal of Scientific Research and Engineering Trends 10(4), 1231–1236 (2024). https://doi.org/10.61137/ijsret.vol.10.issue4.186
[9] Shebl, K. S., Afify, Y. M., Badr, N.: Software defect prediction approaches revisited. International Journal of Intelligent Computing and Information Sciences 23(3), 31–58 (2023). https://doi.org/10.21608/IJICIS.2023.200737.1261
[10] Elsabagh, M. A., Emam, O. E., Gafar, M. G., Medhat T.: Handling uncertainty issue in software defect prediction utilizing a hybrid of ANFIS and turbulent flow of water optimization algorithm. Neural Computing and Applications 36, 4583–4602 (2024). https://doi.org/10.1007/s00521-023-09315-0
[11] Kakkar, M., Jain, S., Bansal, A., Grover, P.: An optimized Software defect prediction model based on PSO - ANFIS. Recent Advances in Computer Science and Communications 14(9), 2732–2741 (2020). https://doi.org/10.2174/2666255813999200818130606
[12] Vashisht, V., Lal, M., Sureshchandar, G. S.: Defect Prediction Framework Using Adaptive Neuro-Fuzzy Inference System (ANFIS) for Software Enhancement Projects. Journal of Advances in Mathematics and Computer Science 19(2), 1–12 (2016). https://doi.org/10.9734/BJMCS/2016/29644
[13] Singh, S. K., Shree, R.: Smart Prediction Method of Software Defect Using Neuro-Fuzzy Approach. Asian Journal of Computer Science and Technology 7(2), 6–10 (2018). https://doi.org/10.51983/ajcst-2018.7.2.1878
[14] Zdorenko, Y., Lavrut, O., Lavrut, T., Lytvyn, V., Burov, Y., Vysotska V.: Route Selection Method in Military Information and Telecommunication Networks Based on ANFIS. In: 3rd International Workshop on Modern Machine Learning Technologies and Data Science (MoMLeT+DS), pp. 514–524. CEUR, Lviv-Shatsk, Ukraine (2021). https://ceur-ws.org/Vol-2917/paper36.pdf
[15] Onyshchenko, S., Haitan, O., Yanko, A., Zdorenko,Y, Rudenko, O.: Method for detection of the modified DDoS cyber attacks on a web resource of an Information and Telecommunication Network based on the use of intelligent systems. In: Proceedings of the Modern Data Science Technologies Workshop (MoDaST), pp. 219–235. CEUR, Lviv, Ukraine, (2024). https://ceur-ws.org/Vol-3723/paper12.pdf
[16] Doremaele, E., Stevens, T., Ringeling, S., Spolaor, S., Fattori, M., Van de Burgt, Y.: Hardware implementation of backpropagation using progressive gradient descent for in situ training of multilayer neural networks. Science Advances 10(28), eado8999 (2024). https://doi.org/10.1126/sciadv.ado8999
[17] Ramya, T., Kannan, A. C., Balasenthil, R. S., Bagirathi, B. A.: Fuzzy Logic modeling for decision making processes using MATLAB. Advanced Materials Research (984–985), 425–430 (2014). https://doi.org/10.4028/www.scientific.net/AMR.984-985.425

DOI: https://doi.org/10.30837/IST-2024.P1.043

Quadrotor Simulation Packages for Reinforcement Learning

Vadym Honcharenko[1] та Valentyn Yesilevskyi[1]

[1] *Kharkiv National University of Radio Electronics, Kharkiv, Ukraine*

Abstract

This publication reviews the most popular quadrotor simulations applied in reinforcement learning for UAV control, obstacle avoidance, path planning, and swarm behavior. We emphasize the importance of high-quality simulations that integrate visual and physical models to mimic real-world dynamics and sensor data accurately. Several popular quadrotor simulation platforms - AirSim, gym-pybullet-drones, Flightmare, Aerial Gym, RotorPy, and RotorS, are evaluated for their different capabilities and supporting reinforcement learning applications. Each package is analyzed for its strengths in rendering, physical accuracy, parallelization, and compatibility with reinforcement learning frameworks. These simulators provide valuable tools for developing and testing reinforcement learning algorithms.

Keywords

quadrotor; reinforcement learning; simulation AirSim; gym-pybullet-drones; Flightmare; Aerial Gym; RotorPy; RotorS

1. Introduction

Over the last decade, quadrotors have become valuable in many practical applications in agriculture, civil engineering, environmental monitoring [1], search and rescue operations, the military sphere, and many other fields.

Reinforcement learning, a fundamental machine learning component, is the cornerstone of quadrotor control. It involves iterative training of agents to make decisions by interacting with an environment via a feedback loop with rewards and penalties as a loss function.

For the last decade, reinforcement learning achieved significant advancements for UAV in many fields of its autonomy [9]: UAV control, obstacle avoidance, path planning, and flocking behaviors. UAV control optimizes flight stability and responsiveness. Obstacle avoidance helps dynamically navigate through space to avoid collisions with objects. Path planning algorithms enable drones to find the most efficient routes subject to the conditions, adjusting to real-time environmental changes. Flocking is applied to control multiple UAVs in cooperative scenarios, allowing swarms of quadrotors to fly in formation, coordinate on missions, and avoid inter-agent collisions.

2. Quadrotor simulation in reinforcement learning

The training of reinforcement learning agents uses numerical simulation, which necessitates a robust mathematical model of interaction between an agent and an environment and a reward function. The quality of these elements is pivotal in transferring agent policymaking from simulation to the real world. Over the last decade, numerous works have been dedicated to solving quadrotor control problems with reinforcement learning [2].

Simulations are used to imitate real-world agent-environment feedback in reinforcement learning. Quadrotor simulation has two main parts - visual and physical. Visual simulation

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
vadym.honcharenko@nure.ua (V. Honcharenko); valentyn.yesilevskyi@nure.ua (V. Yesilevskyi)
0009-0002-0370-3361 (V. Honcharenko); 0000-0002-5935-1505 (V. Yesilevskyi)

imitates optical-flow data (visual RGB – input, segmentation, depth maps, etc.) for sensors like cameras and LiDARs. They differ from rendering engines (Unreal Engine, Gazebo, Unity, and others) and provide data quality (resolution, world realism) and quantity (RGB, segmentation maps, depth maps, etc.). Visual simulation is essential for obstacle avoidance, object tracking, and visual navigation. Physical simulation imitates quadrotor kinematics, interactions with other objects, and different space conditions (wind, air pressure, collisions). They differ with several implemented effects (like some not commonly implemented effects – downwash, parasitic aerial drug, wind, and others), physics simplifications, and assumptions (linearized motion equations, quadratic drag and torque, constant mass distribution, and others). Physical simulations are essential for UAV control, navigation, and path planning.

3. Simulation packages

Many simulation packages have been developed and published since the increasing adoption of machine learning-based techniques for UAVs in the mid-2010s. They help to make fast, precise, and realistic UAV simulations that are needed to test UAVs and train machine learning and control algorithms. Besides physical and visual fidelity, simulation packages have other significant properties: speed, compatibility with programming languages and software (like different ROS and libraries like OpenAI Gym), licensing, maintenance status, community support, and development entrance level. [9]

3.1. AirSim

AirSim [3] is a Microsoft UAV simulation product built on the Unreal rendering engine that utilizes Fast Physics and PhysX physics engines developed in C++ and C#. It is focused on visual fidelity; because of this, it is used mainly for autonomous navigation and computer vision. AirSim offers a large number of sensors and compatibility with open-source controllers (like PX4 and ArduPilot), which can be helpful for pilot training applications.

AirSim uses NVIDIA's physics engine, PhysX, which is not specialized for UAVs, and it is tightly coupled with the rendering engine to allow simulating environment dynamics; for this reason, it can achieve only limited simulation speeds. This limitation makes it difficult to use for reinforcement learning tasks in control and navigation areas.

3.2. gym-pybullet-drones

gym-pybullet-drones [4] is an open-source package for simulating multiple quadrotors with PyBullet physical and visual engines suited for Python development. It is mostly used for swarm-controlling tasks like line trajectory tracking, takeoff, and hovering.

Gym-pybullet-drones is based on the open-source Bullet Physics engine, which is notable for its realistic collisions and aerodynamic effects (like a drag, ground effect, and downwash). PyBullet engine also allows the extraction of synchronized rendering and kinematics. Both CPU- and GPU-based rendering are also available. Notably, it is compatible with the Unified Robot Description Format (URDF).

This package has interfaces for multi-agent and vision-based reinforcement learning applications utilizing the Gymnasium APIs. It includes examples of reinforcement learning workflows for single-agent and multi-agent scenarios.

3.3. Flightmare

Flightmare [5] is an open-source simulator based on the Unity rendering engine and supports different physical engines suited for both Python and C++ development. Flightmare is used for machine learning applications like path-planning in a complex 3D environment and visual-inertial odometry and is a popular tool for autonomous drone racing. The rendering engine can

generate realistic visual information and simulate sensor noise, lens distortions, and extract the 3D point-cloud of the scene.

Flightmare supports three physical engines: a Gazebo-based quadrotor dynamics, real-world dynamics, and a parallelized implementation of classical quadrotor dynamics. Gazebo-based dynamics are slower but more realistic thanks to high-fidelity physics engines, e.g., Bullet. That allows users to control robot dynamics, ranging from simplified UAV models to advanced ones with friction and rotor drag.

In contrast to AirSim, we decouple the rendering module from the physics engine, which offers fast and accurate physics simulation when rendering is not required.

3.4. Aerial Gym

Aerial Gym [6] is an open-source simulator extension to Isaac Gym developed by NVIDIA that can simulate millions of multirotor vehicles parallelly, taking advantage of GPU parallelization. It provides customizable obstacle randomization functionality for navigation tasks, uses Universal Robot Description Format (URDF) files with an extensive set of objects in the environment, and interfaces for efficiently managing the environment. Aerial Gym provides sample environments containing robots equipped with simulated cameras capable of capturing RGB, depth, segmentation, and optical flow data at thousands of frames per second. It also integrates with PX4 and ROS 2 and additional sensors (magnetometer, GPS, and barometer).

3.5. RotorPy

RotorPy [7] is a lightweight open-source Python simulator focused on providing a comprehensive quadrotor model. The simulator's quadrotor model is extensively detailed, including 6-DoF dynamics, aerodynamic wrenches, actuator dynamics, sensors, and wind models. RotorPy includes a Gymnasium environment for RL applications. However, despite having an accurate quadrotor physics model and aerodynamic effects, it lacks complicated environments.

3.6. RotorS

RotorS [10] is built on top of Gazebo Classic and offers a modular framework for designing UAVs and developing control algorithms, mainly focusing on simulating vehicle dynamics. It provides several multi-rotor models. RotorS has been extensively used in robotics to develop algorithms on MAVs, such as drone racing, exploration, path-planning, or mapping. Nevertheless, the Gazebo engine has limited rendering capabilities and is not designed for efficient parallel dynamics simulation, which makes it challenging to use in reinforcement learning. Also, it does not come with ready-to-use reinforcement learning interfaces, and its dependency on Gazebo makes it ill-advised for parallel execution or vision-based learning applications.

4. Conclusions

While reinforcement learning has significantly advanced quadrotor autonomy in many areas, the choice of simulation environment plays a critical role in the success of these applications. The simulation packages reviewed in this paper offer distinct strengths and limitations, making it essential to select a simulator that best suits the specific application carefully. For example, AirSim excels in photorealistic visual environments but may need more speed for large-scale reinforcement training. At the same time, gym-pybullet-drones prioritize fast, multi-agent simulations with realistic physics but more straightforward visual rendering. Flightmare offers a balance with decoupled physics and rendering, providing flexibility for high-speed simulations, whereas Aerial Gym specializes in large-scale, GPU-accelerated simulations for multi-agent

learning. RotorPy, on the other hand, is lightweight and focuses on detailed quadrotor dynamics but lacks complex environments.

Each simulator has trade-offs regarding computational speed, physical fidelity, visual realism, and compatibility with RL frameworks. Therefore, it is crucial to interrogate the pros and cons of each package relative to the specific requirements of the task - whether the focus is on visual navigation, high-fidelity physics, or large-scale multi-agent coordination -before applying it. Users can maximize efficiency, realism, and effectiveness in their reinforcement learning-based quadrotor applications by aligning the simulation environment with the research or development project goals.

References

[1] Haeri, S. P., Razi, A., Khoshdel, S., Afghah, F., Coen, J., Fule, P., Watts, A., Kokolakis, N.-M., & Vamvoudakis, K. G. (2024). A comprehensive survey of research towards AI-enabled unmanned aerial systems in pre-, active-, and post-wildfire management. https://doi.org/10.48550/arXiv.2401.02456

[2] Buşoniu, Lucian, et al. "Reinforcement learning for control: Performance, stability, and deep approximators." *Annual Reviews in Control* 46 (2018): 8-28.

[3] Shah, Shital, et al. "Airsim: High-fidelity visual and physical simulation for autonomous vehicles." *Field and Service Robotics: Results of the 11th International Conference*. Springer International Publishing, 2018.

[4] Panerati, Jacopo, et al. "Learning to fly—a gym environment with pybullet physics for reinforcement learning of multi-agent quadcopter control." *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021.

[5] Song, Yunlong, et al. "Flightmare: A flexible quadrotor simulator." *Conference on Robot Learning*. PMLR, 2021.

[6] Kulkarni, Mihir, Theodor JL Forgaard, and Kostas Alexis. "Aerial Gym--Isaac Gym Simulator for Aerial Robots." *arXiv preprint arXiv:2305.16510* (2023).

[7] Folk, Spencer, James Paulos, and Vijay Kumar. "Rotorpy: A python-based multirotor simulator with aerodynamics for education and research." *arXiv preprint arXiv:2306.04485* (2023).

[8] Dimmig, Cora A., et al. Survey of Simulators for Aerial Robots: An Overview and In-Depth Systematic Comparisons. *IEEE Robotics & Automation Magazine*, 2024.

[9] AlMahamid, Fadi, and Katarina Grolinger. "Autonomous unmanned aerial vehicle navigation using reinforcement learning: A systematic review." *Engineering Applications of Artificial Intelligence* 115 (2022): 105321.

[10] Furrer, Fadri, et al. "Rotors—a modular gazebo mav simulator framework." *Robot Operating System (ROS) The Complete Reference (Volume 1)* (2016): 595-625.

DOI: https://doi.org/10.30837/IST-2024.P1.047

Аналіз стійкості розв’язків задач оптимального розміщення розподільчих центрів з територіальним зонуванням

Данило Лубенець[1*†] та Лариса Коряшкіна[1†]

[1] *Національний технічний університет «Дніпровська політехніка», пр. Д. Яворницького, 19, м. Дніпро, 49005, Україна*

Анотація

Розглядається проблема оптимального розміщення розподільчих центрів транспортно-логістичних систем з визначенням зон їх обслуговування. Представлені математичні моделі логістичних задач, зокрема, багатоетапних та частково двоетапних. Наведено результати експериментальних досліджень щодо стійкості розв’язків сформульованих оптимізаційних задач і методів їх розв’язання.

Ключові слова

Оптимальне розміщення-розподіл, логістика, сервісні зони, стійкість розв’язку задачі оптимізації

1. Вступ

Сучасні логістичні системи характеризуються складними зв'язками як між своїми підрозділами, так і з зовнішнім середовищем. Математичні моделі та методи оптимального розміщення об’єктів і розбиття регіону на зони їх обслуговування, які представлені у роботах [1, 2], допомагають аналітикам не тільки раціонально розміщувати елементи логістичних систем, але й оцінювати їх потенціальну спроможність (здатність у повному обсязі надати послугу своїм споживачам або забезпечити їх ресурсом). Розроблено методи розбиття заданого регіону на області, які охоплюють клієнтів, що мають одні й ті самі k найближчі сусідні сервісні центри з N існуючих (або можливих) з оглядом на те, що клієнти кожної області можуть обслуговуватися будь-яким з найближчих k центрів. Окремим випадком є розбиття регіону на монополії, коли споживачі обслуговуються лише одним центром. У роботі [3] описано двоетапні та частково двоетапні процеси розподілу матеріальних потоків, побудовано їх математичні моделі у вигляді неперервних задач оптимального розбиття множин із додатковими зв’язками. Моделі і методи розміщення додаткових підрозділів логістичних систем з перерозподілом сервісних зон наведено в роботі [4].

В усіх зазначених моделях і методах використовуються теоретичні метрики для оцінювання відстані між двома об’єктами регіону, тому часто на практиці у силу різних причин (технічних, географічних) оптимальний розв’язок задачі реалізувати неможливо. Потрібно відхилитися від знайдених координат центрів. Тоді постає питання: як такі зміни вплинуть на потужності центрів, сфери їх обслуговування, значення функціоналу якості розміщення і розбиття. Саме цьому питанню і присвячені дослідження, представлені в даній роботі.

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
* Corresponding author.
† These authors contributed equally.
lubenets.d.y@nmu.one (D. Lubenets); koriashkina.l.s@nmu.one (L. S. Koriashkina)
0009-0000-8563-3760 (D. Lubenets); 0000-0001-6423-092X (L. S. Koriashkina)

2. Постановка задачі оптимального розміщення-розподілу

У прикладному аспекті розглянемо процес евакуації населення з певної території до пунктів первинного збору (ППЗ). Потрібно визначити місця розташування первинних пунктів і розбити територію регіону на зони їх обслуговування задля найшвидшого переміщення населення до цих пунктів. Розглянемо різні випадки задач: коли так звана ємність ППЗ (кількість людського ресурсу, яку він може прийняти) може бути обмеженою, коли населення розподілене в регіоні рівномірно і нерівномірно. Математична модель такого процесу є неперервною задачею оптимального розбиття з розміщенням центрів з обмеженнями.

3. Ідея методу розв'язання задачі оптимального розміщення центрів із розбиттям території на зони обслуговування

Розв'язання неперервних задач оптимального-розміщення-розподілу здійснюється на основі єдиного підходу. В його основі лежить наступна схема. До вихідних задач нескінченновимірної оптимізації з булевими змінними застосовується ЛП-релаксація, далі формулюється критерій оптимальності розв'язку отриманої задачі, заснований на використанні функціоналу Лагранжа. В результаті – характеристичні функції підмножин, які складають розбиття заданого регіону, вдається знайти в аналітичному вигляді. Формула для їх обчислення містить параметри, які є розв'язками допоміжних скінченновимірних негладких задач максимізації або негладких задач максиміну. Для чисельного розв'язання останніх застосовуються різні модифікації r-алгоритму Шора.

4. Експериментальні дослідження стійкості розв'язків

4.1. План експериментів

Експеримент спрямований на перевірку, стабільності алгоритму розв'язання задачі оптимального розміщення-розподілу, оцінки впливу незначних зміщень центрів на їхні зони обслуговування, дотримання обмежень на їхні спроможності, а також на загальні транспортні витрати на доставку ресурсу з кожної точки регіону до відповідного центру.

На початковому етапі визначено поле розміром 10×10 (з кроком дискретизації 0.04), на якому оптимально, з урахуванням мінімізації витрат на доставку ресурсу, розташовані 9 сервісних центрів (рис. 1). Три з них (1й, 4й, 7й) залишаються фіксованими, інші шість зміщуються у випадковому напрямку в межах заданого інтервалу від -ε до +ε, де ε — невелике значення. Зміщення відбуваються поступово, рівномірно на кожній ітерації експерименту, моделюючи можливі реальні умови зміни позицій центрів збору.

Досліджено три окремі випадки розподілу ресурсів (попиту на послугу) і наявності обмежень на спроможність центрів.

1. Рівномірний розподіл щільності ресурсу (рис. 2, а) без обмежень на ємності для центрів збору. У цьому випадку ресурс розподіляється рівномірно по всьому полю, що означає однакову щільність ресурсу на кожній ділянці. Це дозволяє оцінити, як алгоритм працює в умовах рівномірного навантаження на всі центри збору без будь-яких обмежень по їх ємності. Такий розподіл служить базовою точкою відліку, допомагаючи зрозуміти основні характеристики оптимізації при відсутності варіацій у щільності ресурсу.

2. Нерівномірний розподіл ресурсу (рис. 2, б) з точками підвищеної щільності без обмежень на ємності для центрів збору. У цьому випадку на полі присутні три точки скупчення ресурсу. Решта території характеризується меншою щільністю ресурсу. Це дозволяє перевірити алгоритм у випадку нерівномірного навантаження на центри збору, але без обмежень по їхній ємності. Такий розподіл ресурсів більше наближений до реальних умов, де ресурси можуть бути концентровані в певних зонах.

3. Нерівномірний розподіл ресурсу з трьома точками підвищеної щільності та обмеженнями по ємності для деяких центрів збору (початкові координати не змінені). Алгоритм повинен враховувати обмеження по ємності, які не дозволяють окремим центрам збирати більше заданого обсягу ресурсу. Такий вид експерименту дозволяє оцінити, наскільки стабільно алгоритм працює в умовах, коли потрібне дотримання обмежень, і яким чином він вирішує задачу оптимального розподілу ресурсів за наявності фізичних обмежень у системі.

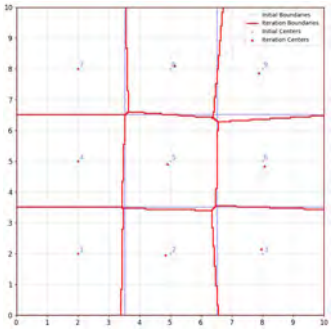


Рисунок 1 – Оптимальне і збурене розміщення центрів

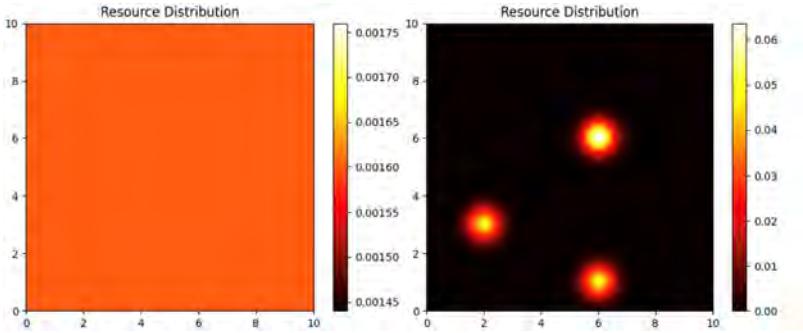


Рисунок 2 – Щільність розподілу ресурсу: а – рівномірна; б – нерівномірна

Програма для експерименту написана мовою Python, а результати зберігаються для подальшого аналізу та візуалізації. Візуалізація результатів дозволяє побачити, як зміщення центрів збору впливає на загальні витрати та ефективність розподілу ресурсу.

4.2. Результати досліджень

Експериментальні дослідження показали стабільність та ефективність розроблених алгоритмів розв'язання задач оптимального розміщення центрів із зонуванням області в різних умовах розподілу щільності ресурсу та при наявності обмежень по ємності центрів збору. У першому випадку з рівномірним розподілом ресурсу алгоритм забезпечує стійкий розподіл, що підтверджує його здатність оптимізувати доставку без відхилень навіть при зміщенні центрів. У другому випадку алгоритм адаптується до змін у щільності, оптимально перерозподіляючи ресурс серед центрів залежно від їхнього розташування щодо зон високої щільності. У третьому випадку результати експериментів свідчать про здатність алгоритмів ефективно управляти розподілом навіть за умов, коли деякі центри швидко досягають своєї максимальної ємності. Це забезпечило збалансоване перерозподілу ресурсу, що сприяє оптимальному використанню обмежених можливостей центрів. На рис. 3 представлено результати трьох експериментів з визначення зон відповідальності 9-ти центрів за їх початковими і зміщеними координатами. Обмеження на ємність центрів задані вектором (100, 5, 5, 100, 5, 3, 100, 5, 5). У таблиці 1 наведено інформацію про зміну (у відсотках) значення цільового функціоналу задачі зі збільшенням сумарної відстані між початковими та зміщеними центрами. Аналіз результатів наведених експериментів дозволяє зробити висновок про те, що по-перше, незначним зміщенням координат центрів відповідають незначні зміни критерія якості розбиття. По-друге, обмеження на кількість ресурсів значно впливають на процес розподілу. У цьому сценарії центри, які знаходяться поблизу зон з високою щільністю ресурсу, могли б зібрати більше, але обмеження ємності перешкоджають цьому, що змушує алгоритм перерозподіляти надлишок ресурсу між іншими центрами.

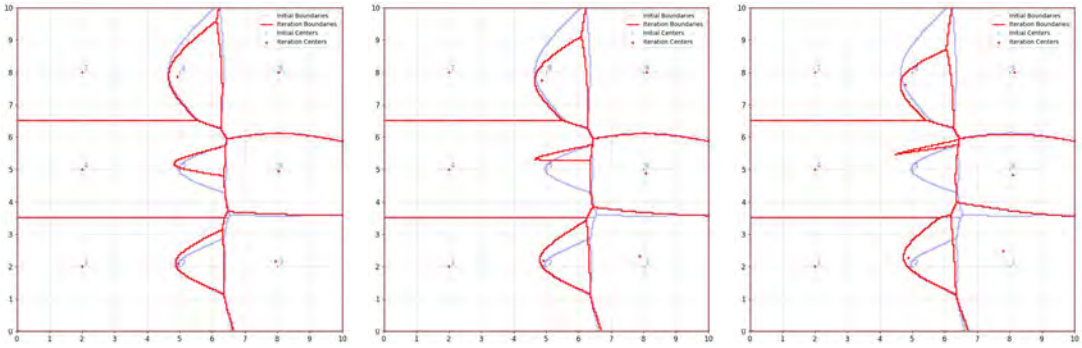


Рисунок 3 - Розподіл ресурсу між початковими та зміщеними центрами збору у випадку нерівномірної щільності ресурсу та наявності обмежень на ємності центрів

Таблиця 3

Зріст функціоналу зі збільшенням збурення координат центрів

№ експерименту	Сума відстаней між зміненими і початковими центрами	Значення цільового функціоналу; ум. од.	Зріст функціоналу відносно його початкового 242,82 ум. од.; %
1	0,751649	243,99	0,481838
2	1,516329	245,42	1,070752
3	2,267544	246,99	1,717321

5. Висновки

Використання побудованих математичних моделей забезпечує швидку наближену оцінку не лише характеристик розподілу і доставки матеріальних ресурсів в ієрархічних транспортно-логістичних системах, а й пов'язаних з ними витрат, сприяє виробленню ефективних управлінських рішень, кількісно обґрунтованих значеннями різних критеріїв оптимальності. Застосування розробленого математичного і програмного забезпечення вказаних задач корисно для отримання додаткової інформації про можливості існуючих сервісних центрів і необхідності відкриття нових задля забезпечення опанування всієї території регіону, або закриття нерентабельних підрозділів, якщо решта центрів здатна надати послугу у повному обсязі усім споживачам.

Список літературних джерел

[1] Л. Коряшкіна, Д. Лубенець, Математичні моделі та методи мультиплексного розбиття і багатократного покриття множин для задач розміщення-розподілу, Information Technology: Computer Science, Software Engineering and Cyber Security, 4(2023): 12 – 25.

[2] Д.Є. Лубенець, Л.С. Коряшкіна. «Системний аналіз і оптимізація розподілу матеріальних ресурсів в ієрархічних транспортно-логістичних системах». Молодь: наука та інновації: матеріали XI Міжнародної науково-технічної конференції студентів, аспірантів та молодих вчених, Дніпро, 22–24.11.2023. НТУ «ДП», 2(2023): 21 – 22. http://ir.nmu.org.ua/handle/123456789/166083.

[3] Л.С. Коряшкіна, С.В. Дзюба, Математичні моделі та методи розміщення об’єктів із зонуванням території в системах екстреної логістики, System technologies, 149(2023): 107 – 122. DOI 10.34185/1562-9945-6-149-2023-09.

[4] M. Sazonova, L. Koriashkina, Optimal location of additional facilities and reallocation of service areas, 27.08.2024. https://doi.org/10.21203/rs.3.rs-4971931/v1.

Методи оптимізації комп'ютерного моделювання у інформаційних системах енергетичного спрямування

Олексій Шапиро[1*†], Ігор Шубін[1†] та Ігор Сотник[1†]

[1] *Харківський національний університет радіоелектроніки, пр. Науки, 14, м. Харків, 61166, Україна*

Анотація

Складність інформаційних систем енергетичного спрямування і потреба в точному комп'ютерному моделюванні реальних процесів вимагають розробки ефективних підходів для оптимізації обчислень. Робота зосереджена на питаннях покращення точності оцінювання та зниження обчислювальних витрат для реалізації таких систем. Пропонується модифікація існуючих метрик складності, що дозволяє більш ефективно виявляти та усувати «вузькі місця» в обчислювальних процесах, зберігаючи необхідний рівень точності результатів.

Ключові слова

оптимізація, комп'ютерне моделювання, інформаційні системи енергетичного спрямування, високоточні обчислення, метрики складності

1. Вступ

Інформаційні системи енергетичного спрямування відіграють важливу роль у дослідженні складних процесів, що відбуваються у фізичному, хімічному та біологічному середовищах. Зокрема, моделювання кліматичних явищ, хімічних реакцій та поведінки біологічних систем вимагає надзвичайно високої точності, адже навіть незначна похибка може призвести до суттєвих відхилень результатів, особливо на великих часових чи просторових масштабах.

Однак, високоточні обчислення для таких моделей [1] потребують значних ресурсів, що призводить до значних обчислювальних витрат, затримок у розрахунках і труднощів із підтриманням високої продуктивності. Це піднімає питання оптимізації процесів комп'ютерного моделювання у інформаційних системах енергетичного спрямування.

Ефективне управління обчислювальними витратами дозволяє скоротити час виконання розрахунків і забезпечити більш швидке отримання результатів без втрати точності, що є критичним для своєчасного аналізу та прийняття рішень.

2. Оптимізація та виявлення «вузьких місць» у обчисленнях

Для виконання оптимізації у процесах комп'ютерного моделювання важливо виявити так звані «вузькі місця», або елементи програми, які найбільше впливають на час виконання та потребують покращення. Відомим методом виявлення таких ділянок є використання метрик оцінки складності програмного забезпечення [2].

Метрики оцінки складності програмного забезпечення, такі як цикломатична складність, ABC-метрика, метрики Холстеда, слугують для оцінки різних аспектів продуктивності програми. Ці метрики дозволяють здійснювати аналіз керуючих структур, кількості операторів, взаємозалежностей і складності обчислень. Наприклад,

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine

* Corresponding author.

† These authors contributed equally.

oleksii.shapyro@nure.ua (О. Шапиро); igor.shubin@nure.ua (І. Шубін); ihor.sotnyk@nure.ua (І. Сотник)

0009-0005-3539-266X (О. Шапиро); 0000-0002-1073-023X (І. Шубін); 0009-0000-5638-6975 (І. Сотник)

цикломатична складність підходить для оцінки кількості незалежних шляхів у коді, що важливо для моделювання різних сценаріїв.

Однак у їхньому оригінальному вигляді ці метрики не завжди підходять для задач, що включають високоточні обчислення процесів енергетичного спрямування [3]. Для таких завдань критично важливо врахувати не тільки структурні характеристики алгоритму, а й обчислювальні витрати на виконання кожної окремої операції.

Тому оптимізація інформаційних систем енергетичного спрямування потребує модифікованих підходів до оцінки складності, які б могли визначати не лише загальну структурну складність, але й специфіку окремих обчислювальних блоків. Завдяки цьому стає можливим не лише оцінити складність обчислень, а й виділити операції, які мають найбільший вплив на загальний час виконання та потребують оптимізації. Виявлення «вузьких місць» дозволяє визначити, де саме слід знижувати точність або перерозподіляти обчислювальні ресурси, що є критично важливим для ефективного використання моделювання у інформаційних системах енергетичного спрямування.

3. Пропозиція вдосконаленої методики оптимізації

Для підвищення ефективності обчислень у інформаційних системах енергетичного спрямування пропонується метод оптимізації комп'ютерного моделювання, що включає модифікацію цикломатичної метрики з урахуванням високоточних обчислень окремих кроків/операцій програми, окрім оцінки загальної алгоритмічної складності програми. Основним інструментом для покращення точності оцінки обчислювальних витрат є використання бінарного дерева формул [4], у якому кожен вузол представляє окрему операцію або змінну з певною точністю.

Запропонована модифікація метрики дає змогу не лише оцінити структурну складність коду, але й розглядати вплив точності на кожному етапі обчислень. Так, операції з високою точністю обчислень можуть бути визначені як «критичні вузли», що потребують особливих обчислювальних ресурсів. Водночас операції з меншою точністю можуть бути адаптовані для досягнення оптимального балансу між точністю результатів та витратами на обчислення. Для досягнення цієї мети бінарне дерево формул дозволяє провести аналіз знизу вгору, де точність змінних у вузлах нижнього рівня визначає витрати обчислень на операції, пов'язані з ними на верхніх рівнях.

Застосування такої методики дозволяє отримати детальнішу оцінку обчислювальних витрат, які стають базою для подальших кроків оптимізації. Наприклад, при моделюванні процесів молекулярної динаміки [5] деякі обчислення потребують значно вищої точності, ніж інші. Модифікована метрика дозволяє відокремити операції з високими вимогами до точності від тих, де можливо застосувати оптимізацію. Цей підхід дозволяє підвищити продуктивність, одночасно зберігаючи високий рівень точності для критично важливих компонентів моделі.

Таким чином, модифікована метрика складності стає не лише засобом аналізу, а й важливим інструментом для адаптивного управління точністю та оптимізації процесів моделювання у інформаційних системах енергетичного спрямування. Застосування вдосконалених методів оптимізації є особливо актуальним у таких галузях, як екологічне моделювання, прогнозування клімату, а також у хімічному та біологічному моделюванні.

4. Практичне застосування вдосконалених методів оптимізації

Прикладом практичного застосування є моделювання гідрологічних процесів. У таких задачах необхідно обробляти великі обсяги інформації, пов’язаної з потоками води, температурними даними та іншими екологічними показниками. Оптимізація процесу обробки цих даних з використанням вдосконаленої метрики дозволяє прискорити моделювання і при цьому зберегти точність критично важливих параметрів. Для

складних обчислень, таких як інтегрування рівнянь руху води, вдосконалена метрика вказує на області, де можна мінімізувати витрати часу на обчислення без зниження якості отриманих результатів.

Ще однією перспективною сферою є біомедичне моделювання, наприклад, в симуляціях взаємодії білків у клітинному середовищі. Вдосконалена метрика допомагає визначити, де саме потрібно зберегти високу точність (наприклад, у розрахунках взаємодій молекул), а де можна знизити її для менш важливих обчислень, щоб не перевантажувати систему.

Таким чином, запропонований метод оптимізації може бути інтегрований в широке коло прикладних задач, що робить його цінним інструментом для розробників та дослідників, які працюють як над створенням складних моделей енергетичних процесів та розробкою інформаційних систем енергетичного спрямування, так і в інших перспективних сферах.

5. Висновки

У роботі запропоновано методику оптимізації процесів комп'ютерного моделювання у інформаційних системах енергетичного спрямування із використанням вдосконалених метрик складності. Вона спрямована на ефективне виявлення та усунення «вузьких місць», що дозволяє значно скоротити обчислювальні витрати та зберегти необхідний рівень точності результатів.

Розроблена методика є універсальною і може бути застосована до широкого спектру завдань, де точність та оптимізація обчислювальних ресурсів є пріоритетними вимогами. Це робить її цінним інструментом для розробників та дослідників у галузі інформаційних систем енергетичного спрямування, сприяючи побудові ефективних моделей для вивчення реальних енергетичних процесів.

Посилання

[1] E. A. Lee, "The Past, Present and Future of Cyber-Physical Systems: A Focus on Models", Sensors, vol. 15, pp. 4837–4869, 2015. DOI: https://doi.org/10.3390/s150304837.

[2] L. Chen and G. Smith, "Beyond lines of code: Advanced software complexity metrics", IEEE Software, vol. 33, no. 2, pp. 42–49, 2016.

[3] D. D. H. Bailey, "High-Precision Arithmetic: Progress and Problems", ACM Transactions on Mathematical Software, vol. 37, no. 1, pp. 1–13, 2013. URL: http://www.davidhbailey.com/dhbpapers/hp-arith.pdf.

[4] G. W. Dueck and T. Scheideler, "Circuit complexity and the Boolean hierarchy", Mathematics of Information and Computational Sciences, vol. 12, pp. 169–182, 2011.

[5] S. Yuan, X. Han and J. Zhang, "Generating High-Precision Force Fields for Molecular Dynamics Simulations to Study Chemical Reaction Mechanisms Using Molecular Configuration Transformer", The Journal of Physical Chemistry A, vol. 128, no. 21, pp. 4378–4390, 2024. DOI: https://doi.org/10.1021/acs.jpca.4c01267

Секція 3.
Системний аналіз. Інформаційні технології сталого розвитку. Геоінформаційні системи та технології

DOI: https://doi.org/10.30837/IST-2024.P1.055

Ensuring Data Confidentiality During Data Analysis in Information Systems

Mirsakhib Miriiev[1†] та Iryna Buchynska[1†]

[1] *Odessa I.I. Mechnikov National University, 2 Vsevolod Zmiienko Street, Odesa, Ukraine*

Abstract

This paper examines scientific methods for guaranteeing data privacy during analysis in information systems. It explores data protection challenges in the digital age, where information is crucial for all sectors. While data growth and analysis advancements allow studying society and business, storage, processing, and analysis pose privacy threats like unauthorized access and leaks. The paper highlights various scientific methods for data privacy in information systems, focusing on combining data anonymization (removing personal details) with l-differential privacy (adding noise) to ensure high privacy while enabling valuable data analysis.

Keywords

data privacy, information systems, data analysis, data anonymization, l-differential privacy, data protection, information security.

1. Introduction

In the era of the digital age, where data becomes a key resource for all spheres of life, privacy protection becomes a primary task for ensuring the privacy and security of users. The increasing volume of data and the development of analysis technologies allow information to be used to study various aspects of society and business. However, the storage, processing, and analysis of data are associated with serious privacy threats, such as unauthorized access, data leaks, and privacy breaches.

2. Scientific Methods for Ensuring Privacy

To ensure data privacy in modern information systems, it is necessary to apply scientific methods that effectively protect information from potential threats. In this context, scientific methods play an important role in the creation and implementation of data protection strategies that provide not only a high level of privacy but also preserve the ability to usefully process and analyze information for informed decision-making.

The paper will consider various scientific methods for ensuring privacy when analyzing data in information systems. Each of them plays an important role in protecting information, ensuring the preservation of privacy and compliance with security requirements in the modern digital environment. When analyzing data in information systems to ensure privacy, there are various scientific methods that can be used (Figure 1).

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine

* Corresponding author.

† These authors contributed equally.

Sahibmiriev4@gmail.com (M. Miriiev); buchinskayaira@gmail.com (I. Buchynska)

0009-0004-7223-1796 (M. Miriiev); 0000-0002-7842-3543 (I. Buchynska)

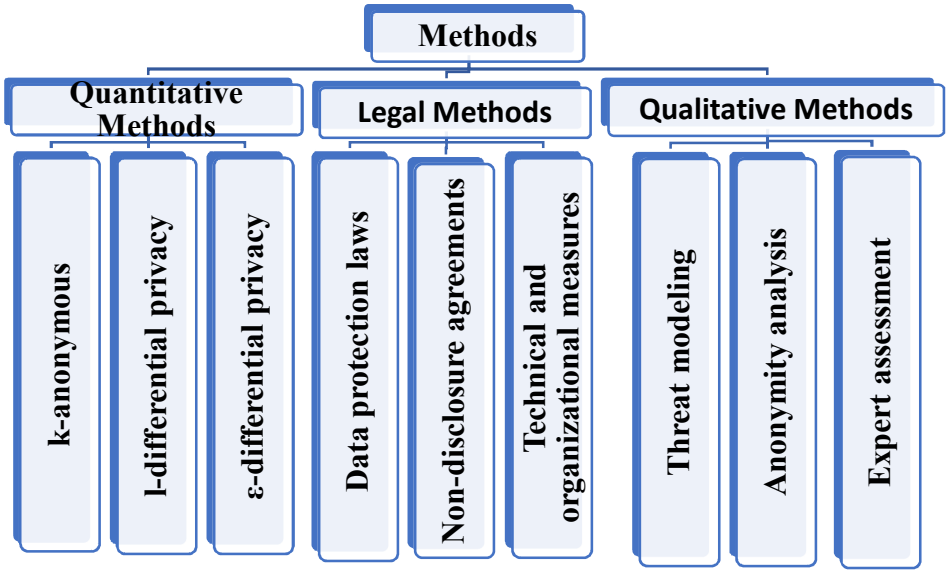


Figure 1: Methods for ensuring privacy when analyzing data in information systems

2.1. Quantitative Methods:

1. Data is considered k-anonymous if each record in the dataset cannot be distinguished from at least k-1 other records based on the values of the attributes used for anonymization.
2. l-differential privacy guarantees that the probability of obtaining any data analysis result for any dataset will be almost the same, regardless of whether a specific record is included in the dataset.
3. ε-differential privacy is similar to l-differential privacy, but uses ε-differential privacy to measure privacy[1].

2.2. Qualitative Methods:

1. Threat modeling involves identifying potential privacy threats and assessing the likelihood and impact of each threat.
2. Anonymity analysis involves analyzing an anonymized dataset to determine whether it is possible to identify records from that dataset.
3. Expert assessment involves involving cybersecurity and privacy experts to assess the risks associated with data analysis.

2.3. Legal Methods:

1. Data protection laws define the rules and regulations that govern the collection, use, and storage of personal data.
2. Non-disclosure agreements (NDAs) oblige parties not to disclose confidential information.
3. Technical and organizational measures include data encryption, access control, and security policies.

3. Combination of Methods

These methods can be used individually or in combination to maximize privacy when analyzing data in information systems. Combining different data protection methods can provide a more effective level of privacy.

Here are some possible combinations:

3.1. Data anonymization with l-differential privacy.

This combination of methods can be used to ensure data privacy when analyzing data that is first anonymized and then l-differential privacy is used to quantitatively assess privacy.

3.2. Threat modeling with k-anonymity.

This combination can be used to identify potential privacy threats associated with data analysis, where k-anonymity is then used to anonymize data to reduce the risk of these threats.

3.3. Data protection laws with data encryption.

The combination of methods can be used to ensure compliance with data protection laws. Data is encrypted to protect it from unauthorized access.

4. Combination of Data Anonymization and l-Differential Privacy

Among the combinations described above, the combination of data anonymization and l-differential privacy is chosen for further examination. Initially, data anonymization is applied to the dataset, which involves removing or replacing personally identifiable information (PII).[2]

Next, considering the anonymized data, l-differential privacy is employed for data analysis. It utilizes algorithms that guarantee that even with minor changes in the input data, the analysis results remain unaltered.

The combination of data anonymization and l-differential privacy can be implemented as follows:

D - represents the dataset.

A - represents the anonymization function that transforms D into an anonymized dataset D'.

Data anonymization can be expressed as:

$$D' = A(D) \tag{1}$$

Let Q represent a database query.

D1 and D2 represent two datasets that differ by only one record.

A(Q, D) represents the response to query Q on database D.

l-differential privacy mandates: for all Q and D1, D2 differing by only one record:

$$\Pr[\mathrm{A}(Q, D1) \in S] \leq e^{l} \times Pr[Q\mathrm{A}(Q, D2) \in S] \tag{2}$$

where Pr[] is the probability that the query Q's result on database D will be within a specific set S.

Formula [2] implies that the probability of obtaining a data analysis result for dataset D1 is almost identical to that for dataset D2, even if they differ by only one record.

Consequently, the combination of data anonymization and l-differential privacy safeguards data privacy by ensuring anonymity and maintaining nearly equal probabilities for data analysis outcomes even with minor dataset changes.

Thus, the combination of data anonymization and l-differential privacy allows protecting the confidentiality of data by ensuring anonymity and nearly equal probability of data analysis results even with small changes in the dataset.

The diagram in Figure 2 illustrates various data protection methods and their interrelationships.

The combination of data anonymization and l-differential privacy simplifies the identification of individuals in the dataset and ensures guaranteed confidentiality during data analysis.

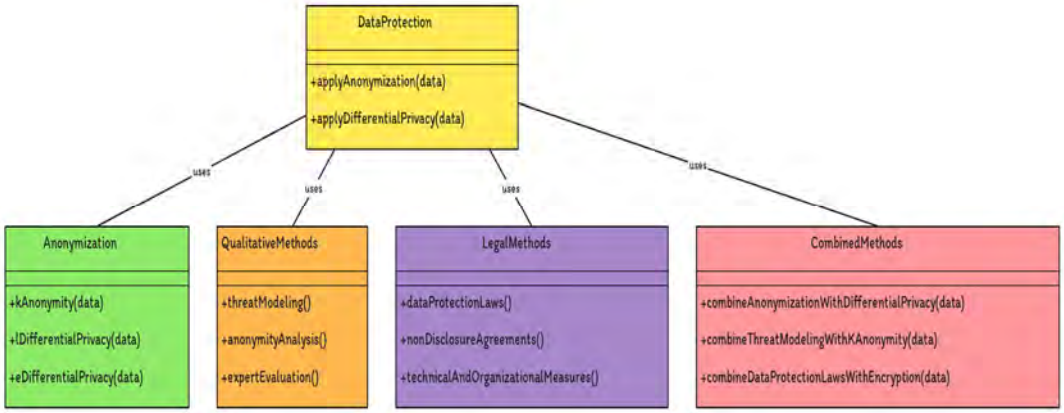


Figure 2: Data Protection Methods

Anonymization transforms data to complicate or prevent the identification of individuals, while l-differential privacy ensures that even with some information about other records, the ability to exclude the presence of a specific record from the response to a query is maintained at a high level. For example, suppose it's necessary to analyze a dataset of patients to determine if there is a correlation between age and heart rate frequency. Data anonymization can be used to hide the names and addresses of patients. Then, l-differential privacy can be applied to add noise to the age data, making it more difficult to identify specific patients. This will allow us to conduct the analysis without compromising the confidentiality of patients.

The advantages of combining data anonymization and l-differential privacy include: increased level of confidentiality, which can ensure a higher level of data confidentiality than either of them separately; quantitative assessment of confidentiality providing the ability to quantitatively evaluate the level of data confidentiality; flexibility in using them together with various types of data and analytical methods. As for the disadvantages of the combination: the combination of these two methods can be a challenging task; information loss may lead to loss of information, which can affect the results of data analysis; decreased data utility may result in decreased usefulness of data for analysis.

5. Conclusion

Scientific methods for ensuring privacy when analyzing data in information systems are crucial tools for guaranteeing information privacy and security in the digital world. These methods play a vital role in creating and implementing data protection strategies that ensure not only a high level of privacy but also the ability to process and analyze information effectively for informed decision-making. Understanding and effectively implementing these methods helps safeguard information and maintain user trust in digital technologies. Combining methods can provide a more robust level of data protection. The combination of data anonymization and l-differential privacy offers an effective and flexible approach to ensuring data privacy during analysis. It allows you to strike a balance between protecting personal information and obtaining valuable analytical results.

References

[1] M Ros Martín, Impact evaluation clustering-based k-anonymity for recommendations, 2017.
[2] Kazukuni Kobara, Cyber Physical Security for Industrial Control Systems and IoT, IEICE Transactions on Information and Systems, 2016, Volume E99.D, Issue 4, Pages 787-795, DOI: https://doi.org/10.1587/transinf.2015ICI0001

Секція 4.
Розпізнавання образів, цифрова обробка зображень і сигналів

DOI: https://doi.org/10.30837/IST-2024.P1.060

Оцінка продуктивності методів фотонного мапінгу

Наталія Аксак[1†] та Дмитро Могилевський[1*†]

[1] *Харківський національний університет радіоелектроніки, пр. Науки, 14, м. Харків, Україна*

Анотація

У статті розглянуто різні методи фотонного мапінгу для рендерингу 3D сцен, зокрема RTPM, Hash-PPM та Stoch-PPM. Визначено їхні переваги та недоліки залежно від продуктивності, ефективності використання пам'яті та якості зображення. Оцінка проводилась на основі тестових сцен з використанням HDRI карт для моделювання реалістичних умов освітлення. Результати експериментів показали, що RTPM є найефективнішим для сцен з інтенсивними каустиками та систем з підтримкою апаратного трасування променів. У випадках без такої підтримки або з великою кількістю джерел світла, Stoch-PPM та комбінація RTPM із Stochastic Light Culling можуть бути оптимальними. Використання VCM рекомендовано для сцен зі складним непрямим освітленням.

Ключові слова

фотонний мапінг, рендеринг, каустики, трасування променів, цифрова обробка зображень, оцінка продуктивності

1. Введення до методів фотонного мапінгу

Фотонне мапування — це ефективний метод глобального освітлення, який дозволяє моделювати складні ефекти, такі як каустики та розсіяне світло. Він використовується для точного моделювання взаємодії світла з матеріалами в різних галузях, включаючи біологію та екологію. Різні реалізації фотонного мапування можуть значно відрізнятися за продуктивністю і якістю через вибір алгоритмів, налаштування якості, оптимізацію та використання ресурсів. Покращені реалізації балансують між якістю і швидкістю залежно від завдань користувача.

2. Схожі роботи

У роботі [1] представлено метод глобального освітлення для сцен із залученими носіями на основі двонаправленого трасування променів Монте-Карло з використанням фотонних карт для підвищення ефективності та зменшення шуму. Метод імітує багаторазове об’ємне розсіювання, кольорове розтікання між об’ємами та поверхнями, а також об’ємну каустику. Фотонні карти відокремлені від геометрії сцени, що дозволяє імітувати складні об’єкти без необхідності мозаїки. Метод автоматично адаптується до освітлення і виду, що робить його придатним для візуалізації анімацій. Фотонно-ефективне формування зображень дозволяє використовувати однофотонні датчики для захоплення 3D-зображень з одним фотоном на піксель [2]. Однак на практиці мала кількість фотонів часто змішується з фоновим шумом, що створює проблему для алгоритмів реконструкції. У роботі [3] представлено детектор, який розрізняє до 100 фотонів за допомогою просторово-часового мультиплексування масиву надпровідних нанодротів. Він підходить для застосувань у фотонних квантових обчисленнях та

* Corresponding author.
† These authors contributed equally.
nataliia.axak@nure.ua (Н. Аксак); dmytro.mohylevskyi@nure.ua (Д. Могилевський);
0000-0001-8372-8432 (Н. Аксак); 0009-0003-2889-6208 (Д. Могилевський);

квантовій метрології. У роботі [4] повідомляють про джерело одного фотона з високою наскрізною ефективністю. У дослідженні [5] представлено всебічний огляд поступової еволюції алгоритмів транспортування світла, що сприяло прогресу інструментів моделювання, зосереджуючись на використовуваних процедурах і алгоритмах. У роботі [6] представлено реалізацію прогресивного фотонного картографа для трасування променів (RTPM), що використовує апаратне прискорення (наприклад, GPU) для підвищення продуктивності. RTPM забезпечує реалістичне освітлення, особливо для складних сцен, і підходить для інтерактивних додатків та анімації. Хеш-базований прогресивний фотонний мапінг (Hash-PPM) [7] підвищує швидкість доступу до фотонів і зменшує витрати пам'яті, роблячи його ефективним для великих та складних сцен. Він оптимізує пошук фотонів, що особливо корисно для сцен із численними об'єктами або складною геометрією. Стохастичний паралельний фотонний мапінг (Stoch-PPM) [8] підвищує якість освітлення через паралельні обчислення та стохастичний вибір фотонів, зменшуючи шум і покращуючи стабільність у складних сценах. Паралельна обробка значно прискорює обчислення, що робить метод ефективним для великих систем.

Методи фотонного мапінгу відрізняються продуктивністю залежно від технік і оптимізацій. Вони ефективно моделюють складні сцени, покращуючи реалістичність рендерингу. Оцінка продуктивності важлива для вибору оптимального рішення відповідно до вимог якості зображення, швидкості та ресурсів.

3. Опис тестових сцен

Платформа HDRI Haven [9] надає безкоштовні HDRI (High Dynamic Range Image) карти для освітлення 3D сцен (рис. 1), що допомагає оцінити вплив різних умов освітлення на продуктивність і якість рендерингу фотонного мапінгу. HDRI симулює складні світлові ефекти, як-от каустика, важливі для роботи з прозорими або дзеркальними поверхнями.

Рисунок 2: HDRI карти дозволяють оцінити, як різні умови освітлення впливають на продуктивність і якість рендерингу з використанням фотонного мапінгу.

Для оцінки методів фотонного мапінгу використовується набір тестових сцен з різною складністю та умовами освітлення: Interior_1 – сцена з прозорими об'єктами (рис. 1а); Interior_2 – сцена для демонстрації каустик (рис. 1б); Interior_3 – сцена з добре видимими каустиками; Interior_4 – сцена з високою концентрацією каустичних фотонів (рис. 1в); Interior_5 – сцена з розподіленими каустичними фотонами; Interior_6 – сцена з інтер'єром та екстер'єром і багатьма джерелами світла.

Ці сцени дозволяють об'єктивно оцінити продуктивність алгоритмів, враховуючи якість глобального освітлення, реалістичність каустик і тіней, продуктивність на складних геометріях та оптимізацію швидкості рендерингу.

4. Експерименти

Всі тести проводились з роздільною здатністю 1920x1080 на відеокарті NVIDIA RTX 2080 SUPER. Для кожної сцени ми визначили оптимальну кількість фотонів та початковий радіус пошуку для кожного методу, також оцінили продуктивність за такими критеріями:

1. Час рендерингу на ітерацію:

$$T = \frac{N_{rays} \times T_{\text{ray}} \times N_{\text{photons}} \times \text{T}_{\text{photon}}}{P_{performance}}, \quad (1)$$

де N_{rays} – кількість променів, T_{ray} – середній час трасування одного променя (мс), N_{photons} – кількість фотонів, T_{photon} – середній час обробки одного фотона (мс), $P_{performance}$ – продуктивність системи (залежить від апаратної частини, як-от кількість ядер або швидкість GPU). Час рендерингу T вимірювався для кожного методу на всіх тестових сценах.

2. Якість зображення оцінюється за допомогою метрики SSIM, яка порівнює еталонне та відрендерене зображення, враховуючи яскравість, контрастність і структуру. Значення SSIM = 1 означає ідеальну схожість, а значення, близьке до 0, вказує на значну різницю.

3. Швидкість збіжності C_r (2) визначає, як швидко алгоритм досягає заданого рівня якості, вимірюючись часом від початку рендерингу T_{start} до досягнення порогового значення якості T_{end}. Малий C_r означає швидку збіжність, великий – повільну.

$$C_r = T_{end} - T_{start}. \quad (2)$$

4. Ефективність використання пам'яті M_e визначає частку збережених фотонів для рендерингу відносно згенерованих. Вона обчислюється як відношення збережених фотонів N_{stored} до загальної кількості випущених N_{total}

$$M_e = {}^{N_{stored}}/_{N_{total}}. \quad (3)$$

Високий M_e (близький до 1) свідчить про ефективне збереження більшості фотонів, що покращує обчислення освітлення. Низький M_e (близький до 0) вказує на втрату більшості фотонів, що свідчить про неефективність методу.

RTPM показав найкращий час рендерингу на ітерацію серед усіх методів, будучи вдвічі швидшим за Hash-PPM та на 1.5 рази швидшим за Stoch-PPM. Наприклад, у сцені Interior_1 середній час ітерації був: RTPM — 14 мс, Hash-PPM — 27 мс, Stoch-PPM — 19 мс. Метрика SSIM показала, що RTPM та Hash-PPM мають схожу якість зображення за однакову кількість ітерацій, але завдяки швидкості RTPM досягає кращої якості за однаковий час. Після 10 секунд рендерингу сцени Interior_3, SSIM для RTPM становив 0.94, для Hash-PPM — 0.87, для Stoch-PPM — 0.89. RTPM швидше досягає еталонної якості, наприклад, у сцені Interior_5, де він досяг SSIM 0.9 за 2 секунди, тоді як Hash-PPM потребував 4 секунди, а Stoch-PPM — 3 секунди.

Графік збіжності SSIM для сцени Interior_3 показав, що RTPM досягає майже еталонної якості швидше за конкурентів: після 30 секунд рендерингу SSIM для RTPM становив 0.97, для Hash-PPM — 0.96, для Stoch-PPM — 0.94. RTPM потребує більше пам'яті для структури прискорення трасування променів, але завдяки Photon Culling загальне використання пам'яті часто було меншим, особливо у великих сценах. Наприклад, у сцені Interior_6 RTPM зберігав 153 000 фотонів, тоді як Hash-PPM і Stoch-PPM — понад 1 200 000, що зменшило використання пам'яті та прискорило обчислення. Для розширення аналізу розглянуто також методи Vertex Connection and Merging (VCM) та Stochastic Light Culling (SLC).

5. Висновки

Результати проведених експериментів показали, що: VCM демонструє кращу якість зображення на ранніх етапах рендерингу для сцен з складним непрямим освітленням (наприклад, Interior_2); RTPM залишається більш ефективним для сцен з інтенсивними каустиками; час рендерингу VCM виявився в середньому на 20-30% повільнішим за RTPM; VCM потребує більше пам'яті через необхідність зберігання додаткових структур даних для з'єднання шляхів.

Для систем з апаратним трасуванням променів RTPM є найкращим вибором для сцен зі складними каустиками та численними джерелами світла. Без такої підтримки Stoch-PPM забезпечує баланс між швидкістю та якістю, особливо для середньо складних сцен. Комбінація RTPM із Stochastic Light Culling ефективна для сцен з багатьма джерелами світла. Для дуже складних сцен з непрямим освітленням можна використовувати VCM, якщо час рендерингу не є критичним. Для оптимальної роботи RTPM варто починати з великого радіусу пошуку і поступово його зменшувати.

Список літератури

[1] H. W. Jensen, & P. H. Christensen, Efficient simulation of light transport in scenes with participating media using photon maps. In Seminal Graphics Papers: Pushing the Boundaries, 2023, Volume 2, pp. 301-310.

[2] J. Peng, Z. Xiong, X. Huang, Z. P. Li, D. Liu, & F. Xu, Photon-efficient 3d imaging with a non-local neural network. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16, pp. 225-241.

[3] R. Cheng, Y. Zhou, S. Wang, M. Shen, T. Taher, & H. X. Tang, A 100-pixel photon-number-resolving detector unveiling photon statistics. Nature Photonics, 17(1), 2023, pp.112-119.

[4] N. Tomm, A. Javadi, N. O. Antoniadis, D. Najer, M. C. Löbl, A. R. Korsch, & R. J. Warburton, A bright and fast source of coherent single photons. Nature Nanotechnology, 16(4), 2021, pp. 399-403.

[5] M. Ayoub, A review on light transport algorithms and simulation tools to model daylighting inside buildings. Solar Energy, 198, 2020, pp. 623-642.

[6] R. Kern, F. Brüll & T. Grosch, Accelerating Photon Mapping for Hardware-Based Ray Tracing. Journal of Computer Graphics Techniques Vol, 12(1), 2023.

[7] M. Mara, D. Luebke, M. Mcguire, Toward practical real-time photon mapping: Efficient GPU density estimation. In Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, Association for Computing Machinery, New York, I3D ’13, 2013, pp. 71–78. URL: https://doi.org/10.1145/2448196.2448207. 2, 3, 15, 16.

[8] T. Hachisuka, H. W. Jensen, Parallel progressive photon mapping on GPUs. In ACM Siggraph Asia 2010 Sketches, Association for Computing Machinery, New York, SA ’10, 54:1. 2010. URL: https://doi.org/10.1145/1899950.1900004. 2, 3, 14, 15, 16.

[9] HDRI Haven. 2024. URL: https://hdri-haven.com/category/sunrise-sunset/artificial-light.

DOI: https://doi.org/10.30837/IST-2024.P1.064

Метод калібрування чутливості радіосканерів мобільного зв’язку

Олена Васильєва[1†] та Валерій Огар[12†]

[1] *Національний науковий центр “Інститут метрології”, вул. Мироносицька, 42, 61002, Харків, Україна*
[2] *Харківський національний університет радіоелектроніки, пр. Науки, 14, 61166, Харків, Україна*

Анотація
Якість цифрового (мобільного) зв’язку потребує чіткого визначення площі покриття сигналів саме цього виду і їх амплітудний розподіл. Для цього використовують радіосканери, які застосовуються для вимірювань покриття радіомереж операторів мобільного зв'язку технологій GSM, CDMA2000, WIMAX, LTE TDD, LTE FDD і інших. Радіосканери складаються з вимірювальних антен, блоку приймача і персонального комп’ютера з встановленою програмою аналізу і індикації сигналів Визначити реальну чутливість радіосканера і надати його калібрувальні характеристики щодо рівня вхідних сигналів можливо лише при наявності сигналів відповідного виду модуляції. Калібрування такого комлексованого приладу потребує калібрування його функціональних частин і приладу в цілому. Враховуючи високу чутливість приймача радіосканера і роботу його демодулятора з високостабільними по частоті (еталонними) сигналами, для калібрування можливе лише з використанням еталонного комплексу апаратури. На базі національних еталонів України в ННЦ «ІНСТИТУТ МЕТРОЛОГІЇ» розроблено метод, що позволяє в повному обсязі провести калібрування радіосканерів мобільного зв’язку. Метод використовує найсучаснішу апаратуру (генератори, аналізатори спектра та ін.) зі складу зі складу цих еталонів.

Ключові слова
Калібрування, еталони, модуляція, мобільний зв’язок, радіосканери

1. Вступ

Якість цифрового (мобільного) зв’язку потребує чіткого визначення площі покриття сигналів саме цього виду і їх амплітудний розподіл. Для цього використовують радіосканери. Відмінною рисою таких аналізаторів сигналів є робота одночасно в режимі приймача і демодулятора радіочастотних сигналів. Тому оцінити реальну чутливість радіосканера і надати його калібрувальні характеристики щодо рівня вхідних сигналів можливо лише при наявності еталонних радіочастотних сигналів відповідного виду модуляції. Для формування таких сигналів можуть бути використані лише високостабільні малошумні генератори з можливістю синхронізації сигналів еталонною частотою, а для оцінки рівня сигналу необхідні аналізатори спектру (еталонні радіоприймачі) з відповідною опцією демодуляції. Така унікальна апаратура входить до складу Національного еталона девіації частоти частотно-модульованих коливань НДЕТУ ЕМ-04-2021, що знаходиться в ННЦ «ІНСТИТУТ МЕТРОЛОГІЇ»[1,2] .

Радіосканери мобільного зв'язку застосовуються для вимірювань покриття радіомереж операторів мобільного зв'язку технологій GSM, CDMA2000, WIMAX, LTE TDD, LTE FDD і інших. Радіосканери складаються з вимірювальних антен, блоку приймача і персонального комп’ютера з встановленою програмою аналізу і індикації. Так,

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
† These authors contributed equally.
koropetc@ukr.net (О. Васильєва), valeriy.ogar@nure.ua (В. Огар)
0009-0007-9586-1485 (О. Васильєва), 0000-0001-9489-2107 (В. Огар)

наприклад, радіосканер сигналів R&S TSME6 має діапазон частот від 350 МГц до 6 ГГц , драйв-тестів мереж мобільного зв'язку, входить в склад портативних комплексів для бенчмаркінгу та оптимізації параметрів роботи мереж. Сканер підтримує роботу з сигналами з робочим діапазоном частот від 350 МГц до 6 ГГц, що в комплексі з активованими опціями підтримки технологій мобільного зв'язку забезпечує можливість аналізу сигналів мереж більшості актуальних стандартів. TSMA має вбудований GPS-приймач для синхронізації та фіксації місцезнаходження. Використання внутрішнього обчислювального модуля та накопичувача для збереження даних дозволяє використовувати сканер в автономному режимі після попередньої конфігурації або з керуванням через мобільне підключення (WLAN, Bluetooth, Ethernet,etc). Накопичення даних та обробка внутрішнім обчислювальним модулем дозволяє реалізувати на сканері функцію оцінки положення базових станцій мобільних мереж. Радіосканер підтримує технології

- LTE (Long Term Evolution)
- WCDMA / HSPA / HSPA+
- WiMAX
- TD-SCDMA
- GSM / EGPRS / EDGE Evolution
- GSM-R
- CDMA2000® 1xEV-DO
- CDMA2000® 1xRTT
- cdmaOne
- TETRA

Калібрування такого комлексованого приладу потребує калібрування функціональних частин і приладу в цілому.

2. Основна частина

В Національному науковому центрі "Інститут метрології" розроблено метод калібрування чутливості радіосканерів покриття мобільного зв'язку.

Для калібрування вимірювальних антен, що входять у комплект радіосканера, використовується Національний (державний первинний) еталон одиниці напруженості електромагнітного поля у діапазоні частот від 0,01 МГц до 43 ГГц (НДЕТУ ЕМ-05-2021) [3].

Для перевірки чутливості приймача використовується векторний генератор довільної форми типу N5172B фірми Agilent Technology (Keysight) [4]. Однак він має нестабільність частоти вихідного сигналу при використанні власного внутрішнього генератора опорної частоти $2*10^{-6}$, а для перевірки сигнала для стандарта LTE відхилення носійної частоти від номінального значення повинне бути не більш 500 Гц на частотах до 3 ГГц. Рішенням цієї проблеми є використання в якості джерела опорної частоти 10 МГц цезієвого стандарту частоти типу Microsemi 5071, що входить у склад Національного еталона України часу і частоти НДЕТУ ТF-01-2021 і має похибку встановлення частоти $1,2*10^{-13}$ та нестабільність частоти $5*10^{-14}$[5,6]. Переключення опорного сигнала генератора відбувається автоматично при рівні зовнішнього сигнала з частотою 10 МГц від мінус 3 дБмВт до +0 дБмВт.

Точність вимірювання потужності сигнала при визначенні чутливості забезпечується використанням еталона одиниці потужності НВЧ в діапазоні від 0,03 ГГц до 18 ГГц ДЕТУ 09-06-05 [7] Для контроля форми і спектра сформованого цифрового сигнала використовується аналізатор спектра R&S FSV з опцією вимірювання параметрів цифрових модуляцій К70.

Під час проведення калібрування повинні бути виконані нормальні умови довкілля.

Важливим етапом калібрування є встановлення програмного забезпечення для створення сигналів в генераторі. Методика створення сигналів в генераторі N5172B складається з чотирьох етапів.

На першому етапі виконується скачування з веб-сайту фірми Keysight Technology програмного забезпечення Signal Studio. Після скачування програмне забезпечення встановлюється на комп'ютер і запускається програма. В меню мастера настроювання системи обирають бажану конфігурацію з набору готових сигналів, де вказується вид модуляції і смуга частот. Для ідентифікації сигнала при прийомі радіосканером можна змінити деякі параметри, наприклад можна змінити групу ідентифікаторів комірок фізичного рівня, сектор ідентифікатора комірки фізичного рівня, ідентифікатор комірки. В програмі буде показана діаграма сигнала в координатах час (X) – частота (Y), яка показує наявність службових сигналів. Далі генерується необхідна хвильова форма і експортується файл сигнала, який зберігають на USB –носії.

На другому етапі проводиться запис файла у енергозалежну пам’ять генератора з USB-накопичувача в USB порт генератора, проводиться прив'язка сигналів до лицензійних слотів.

На третьому етапі для перевірки отриманого сигнала підключають вихід генератора N5172B до входу аналізатора спектра, яким оцінюють спектр і сузір'я сигнала.

На четвертому етапі проводиться закріплення (блокування, ліцензування) - запис файла з енергозалежної пам'яті генератора в постійну (енергонезалежну) пам'ять.

Працездатність радіосканера перевіряють виявленням сигналу на одній з частот робочого діапазону при налаштуванні генератора на необхідний стандарт модуляції. Встановлюють необхідну амплітуду і частоту носія на генераторі і виявляють сигнал радіосканером, який демодулює і відображає параметри сигнала на дисплеї комп'ютера при включеній програмі індикації (наприклад R&S Romes). При відповідності параметрів сигнала заданим, вважають, що радіосканер пройшов випробування.

Для визначення рівня чутливості радіо сканера зменшують рівень сигнала генератора до границі чутливості для успішного приймання. Проводять 10 спостережень і визначають середнє арифметичне значення чутливості. Цей рівень сигнала генератора приймаємо за чутливість радиосканера.

Аналогічно вимірюють чутливість на частотах 2140, 1840, 940 МГц, на яких працюють базові станції мобільного зв'язку (або частотах замовника).

Стандартну відносну невизначеність калібрування типу А обчислюють за формулою середньо-квадратичного відхилення результатів вимірювання. Відносна невизначеність калібрування за типом В u_b приймається як невизначеність встановлення вихідного рівня генератора у відповідності до його сертифіката калібрування. Сумарна стандартна відносна невизначеність u_c при калібруванні визначається геометричною сумою невизначеностей типа А і типа В. Розширена відносна невизначеність U для рівня довіри 0,95 при коефіцієнті охоплення 2 при калібруванні обчислюється як подвоєна стандартна невизначеність. Отримана невизначеність вимірювання чутливості радіосканерів не перевищує 1 дБм, що повністю відповідає необхідним технічним характеристикам апаратури, що калібрується.

Метод застосовано на існуючий апаратурі калібрувального комплексу радіовимірювальної апаратури [8].

3. Висновки

1. Якість цифрового (мобільного) зв’язку потребує чіткого визначення площі покриття сигналів саме цього виду і їх амплітудний розподіл. Для цього використовують радіосканери. Визначити реальну чутливість радіосканера і

надати його калібрувальні характеристики щодо рівня вхідних сигналів можливо лише при наявності сигналів відповідного виду модуляції.

2. Калібрування такого комлексованого приладу потребує калібрування його функціональних частин і приладу в цілому. Враховуючи високу чутливість приймача радіосканера і роботу його демодулятора з високостабільними по частоті (еталонними) сигналами, калібрування можливе лише з використанням еталонного комплексу апаратури.
3. На базі національних еталонів України в ННЦ «ІНСТИТУТ МЕТРОЛОГІЇ» розроблено метод, що позволяє в повному обсязі провести калібрування радіосканерів мобільного зв'язку.
4. Отримані при калібруванні значення невизначеності цілком відповідають необхідним технічним характеристикам радіосканерів мобільного зв'язку.

Перелік посилань

[1] Неєжмаков П.І., Павленко Ю.Ф., Огар В.І., Васильєва О.М., Кирієнко С.Р. Новий державний еталон одиниці девіації частоти частотно-модульованих коливань. Український метрологічний журнал. 2020. № 2. С. 3–11. doi: 10.24027/2306-7039.2.2020.208641

[2] П. Неєжмаков, О. Васильєва, Ю. Павленко, В. Огар, Цифровому приладобудуванню – нову метрологію. УМЖ 2023 №3. С. 3-8. doi:10.24027/2306-7039.3.2023.291854.

[3] НДЕТУ ЕМ-05-2021 Національний (державний первинний) еталон одиниці напруженості електромагнітного поля у діапазоні частот від 0,01 МГц до 43 ГГц

[4] Keysight Technologies. Keysight Signal Generator Selection Guide. URL: https://www.keysight.com/us/en/product/N5172B/exg-x-series-rf-vector-signal-generator-9-khz-6-ghz.html

[5] НДЕТУ ТF-01-2021 Національний (державний первинний) еталон одиниць часу та частоти URL:http://www.metrology.kharkov.ua/.

[6] Неєжмаков П.І., О.М. Васильєва, Ю.Ф. Павленко, Особливості квантових еталонів і особливий статус секунди в SI-2019 Український метрологічний журнал. 2023. №4. С. 3–8. doi:10.24027/2306-7039.4.2023.298584

[7] ДЕТУ 09-06-05 Державний первинний еталон одиниці потужності електромагнітних коливань в коаксіальних трактах у діапазоні частот від 0,03 до 18 ГГц URL:http://www.metrology.kharkov.ua/index.php?id=268&L=448

[8] О.М. Васильєва, Ю.Ф. Павленко, В.І. Огар. Калібрувальний комплекс радіовимірювальної апаратури. Український метрологічний журнал. 2024. № 2. С. 10–17. doi: 0.24027/2306-7039.2.2024.307137.

DOI: https://doi.org/10.30837/IST-2024.P1.068

Ознакова модель інтелектуальної ідентифікації радіолокаційних об’єктів

Володимир Жирнов[1] та Світлана Солонська[2]

[1] *Харківський національний університет радіоелектроніки, пр. Науки, 14, м. Харків, 61166, Україна*
[2] *НТУ "Харківський політехнічний інститут", вул. Кирпичова 2, Харків 61002, Україна*

Анотація
Пропонується ознакова модель інтелектуальної ідентифікації радіолокаційних об'єктів, що використовує можливості експерта РЛС, який на основі даних про динаміку змін координат, форму, яскравість і багатоглядову передісторію зображень відміток, здатний ефективно ідентифікувати радіолокаційні об’єкти. Розроблено алгоритм автоматичної ідентифікації рухомих об'єктів на основі семантичних ознак. Показано, як цей підхід використовується для автоматичного виявлення й розпізнавання відміток повітряних і надводних об'єктів.

Ключові слова
інтелектуальна ідентифікація, радіолокаційні об'єкти, розпізнавання, ознакова модель

1. Вступ

У сучасній техніці обробки сигналів та інформації недостатньо ефективно використовуються можливості експерта РЛС, який на основі даних про динаміку змін зображень відміток здатний ефективно ідентифікувати радіолокаційні об’єкти. Основною перевагою такого підходу є варіативний комплексний аналіз просторово-часової картини, що відображається на екрані індикатора. Застосування пропонованого методу під час формування і опису сигнального образу дозволяє визначати семантичні складові його поведінки та на їх основі розробляти системи автоматичної ідентифікації радіолокаційних об'єктів.

2. Метод інтелектуальної ідентифікації радіолокаційних об’єктів на основі ознакової моделі відміток

Розроблено метод ідентифікації радіолокаційних об'єктів на основі аналізу простору ознак. Перевага даного методу пов'язана з використанням додаткової інформації, яка отримана під час створення і аналізу інтелектуального зображення радіолокаційної відмітки, що включає образ об'єкта і семантичну складову. Пропонована модель включає процедури формування й аналізу зображень радіолокаційних відміток, визначення характеристик ознакового простору та автоматичне прийняття рішення щодо виявлення і розпізнавання радіолокаційних об’єктів.

Суть метода полягає у формуванні вектору прийняття рішення щодо виявлення й розпізнавання зображень радіолокаційних відміток за допомогою логічної обробки простору семантичних ознак $W(I_g,\ I_s, I_f)$, який задано на множині семантичних ознак зображень відміток $\{\ I_1,\ I_2, \ldots, I_k\}$ на основі геометричної $\{\ I_{g1},\ I_{g2}, \ldots, I_{gk}\}$, семантичної

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
nauka123@ukr.net (В. Жирнов); solonskaya@ukr.net (С. Солонська)
0000-0002-2397-3126 (В. Жирнов); 0000-0002-8841-7825 (С. Солонська)

$\{ I_{s1},\ I_{s2},\ldots,I_{sk}\}$ складових інтелектуальних зображень і флуктуаційної ознаки $\{ I_{f1},\ I_{f2},\ldots,I_{fk}\}$.

Визначимо через *X* множину об'єктів радіолокації. Ознака - це результат виміру деякої характеристики об'єкта, тобто відображення $I: X - D_I$, де D_I - множина допустимих значень ознаки. Якщо задано ознаки $I_1,\ I_2,\ldots,I_n$, то вектор $x = (I_1(x),\ I_2(x),\ldots,I_n(x))$ називається ознаковим описом об'єкта.

На основі системи ознак: Z_{pij} - наявності сигналу в a_{ij} інформаційній комірці ($i,\ j$ – номери елементів зони огляду РЛС), Z_{dij}, Z_{aij} - переходу сигналу з поточної комірки a_{ij} до суміжної комірки за дальністю або азимутом, сформовані складові інтелектуальної моделі зображень сигнальних відміток для точкових рухомих та малорухомих літальних апаратів: літак, вертоліт, БПЛА, це:

- ознаковий опис відмітки об'єкта радіолокації

$x_1 = \{I_1(Z_{pij}, Z_{dij}, Z_{aij}),\ldots, I_n(Z_{pij}, Z_{dij}, Z_{aij})\}$, створений на основі відношень між подіями в інформаційних комірках в ході формування віртуального просторово-семантичного зображення радіолокаційних відміток;

- множина ознак зображень радіолокаційних відміток

$X = (I_1(x_1),\ \ldots, I_n(x_1)),\ldots,(I_1(x_k),\ \ldots, I_n(x_k))$ для k об'єктів на основі аналізу залежностей первинних семантичних ознак;

- формування матриці ознакових описів відміток об'єктів радіолокації, що створена на множині ознак зображень відміток $\{ I_s(X), I_g(X), I_e(X), I_f(X)\}$ на основі семантичної $\{ I_{s1}(x_1),\ I_{s2}(x_2),\ldots,I_{sk}(x_k)\}$, геометричної $\{ I_{g1}(x_1),\ I_{g2}(x_2),\ldots,I_{gk}(x_k)\}$, енергетичної $\{ I_{e1}(x_1),\ I_{e2}(x_2),\ldots,I_{ek}(x_k)\}$ складових інтелектуальних зображень і флуктуаційної ознаки зображень відміток $\{ I_{f1}(x_1),\ I_{f2}(x_2),\ldots,I_{fk}(x_k)\}$.

Матриця розміру $k \times n$. Стовпці цієї матриці відповідають ознакам: семантична $\{ I_{s1}(x_1),\ I_{s2}(x_2),\ldots,I_{sk}(x_k)\}$, геометрична $\{ I_{g1}(x_1),\ I_{g2}(x_2),\ldots,I_{gk}(x_k)\}$, енергетична $\{ I_{e1}(x_1),\ I_{e2}(x_2),\ldots,I_{ek}(x_k)\}$ складові інтелектуальних зображень і флуктуаційна ознака зображень відміток $\{ I_{f1}(x_1),\ I_{f2}(x_2),\ldots,I_{fk}(x_k)\}$, а кожен рядок є ознаковою характеристикою зображення відмітки одного об'єкта радіолокації. На рисунку 1 наведено зразок матриці.

$I_{s1}(x_1)$	$I_{g1}(x_1)$	$I_{e1}(x_1)$	$I_{f1}(x_1)$
$I_{s2}(x_2)$	$I_{g2}(x_2)$	$I_{e2}(x_2)$	$I_{f2}(x_2)$
…	…	…	…
$I_{sk}(x_k)$	$I_{gk}(x_k)$	$I_{ek}(x_k)$	$I_{fk}(x_k)$

Рисунок 1. Матриця ознакових характеристик інтелектуальних зображень відміток *k* об'єктів радіолокації.

3. Предикатні рівняння ідентифікації

Наведені моделі ідентифікації S_j інтелектуальних зображень для об'єктів радіолокації описуються так:

1. Точковий нерухомий об'єкт радіолокації місцевий предмет: тип зображення S_1, семантична ознака процесу формування зображення відмітки $I_{s1}(x_1) = 1$ при діапазонах зміни ознаки розміру відмітки за азимутом $C1 \le D_{s1}^1\ (x_1) \le C2$, розміру відмітки за дальністю $C3 \le D_{s1}^2\ (x_1) \le C4$; енергетична ознака $I_{e1}(x_1) = 1$ інтелектуальної моделі зображення пачки сигналів при сумі добутків амплітуд сигналів на предикати інформаційних комірок пачки у напрямку, що визначається вектором (l_n) $D_{e1}^l(x_1) =$

$\sum_{l_1}^{l_n} q_{i,j+l_n} Z_{ai,j+l_n} > E1$. Тут $C1$ та $C2$ – мінімальна і максимальна тривалість пачок відміток за азимутом, $C3$ та $C4$ – мінімальна і максимальна тривалість пачок відміток за дальністю, $E1$ – мінімальне значення енергетичної ознаки.

Предикатне рівняння ідентифікації має вид

$$S_1 = I_{s1}(x_1) \wedge I_{e1}(x_1). \quad (1)$$

2. Точковий малорозмірний, малошвидкісний рухомий об'єкт радіолокації спортивний літак, вертоліт, БПЛА: тип зображення S_2, семантична ознака процесу формування зображення відмітки $I_{s2}(x_2)$ = 1 при діапазонах зміни ознаки розміру відмітки за азимутом $C5 \le D_{s2}^1\ (x_2) \le C6$ і ознаки розміру відмітки за дальністю $C7 \le D_{s2}^2\ (x_2) \le C8$; енергетична ознака $I_{e2}(x_2) = 1$ інтелектуальної моделі зображення пачки сигналів при сумі добутків амплітуд сигналів на предикати інформаційних комірок пачки у напрямку, що визначається вектором (l_n) $D_{e2}^l(x_2) = \sum_{l_1}^{l_n} q_{i,j+l_n} Z_{ai,j+l_n} > E2$; флуктуаційна ознака зображень відміток $I_{f2}(x_2)$ = 1, коли підпростір значень спектрально-семантичної ознаки відмітки для оцінки швидкості руху об'єкта визначається ознакою $D_{f2}^1(x_2)$ амплітудних флуктуацій пачки (для малих допплерівських зміщень частоти), з урахуванням амплітудних даних $q_{i,j}$, використовуючи дані про форму, визначаємо ознаку флуктуації як допустиму різницю амплітуд відповідних комірок $D_{f2}^1(x_2) = 1$ при $\Delta g = qmin_{max}$ та ознаку одиночної групи одиниць $D_{f2}^2(x_2) = 1$ при F1=2 (для одиночної групи зімкнутих одиниць у множині результат підсумовування завжди дорівнює двом). Тут C5 та C6 – мінімальна і максимальна тривалість пачок відміток за азимутом, C7 та C8 – мінімальна і максимальна тривалість пачок відміток за дальністю, E2 – мінімальне значення енергетичної ознаки, F1– арифметична сума предикатів у предикатній функції *F(y)* .

Предикатне рівняння ідентифікації має вид

$$S_2 = I_{s2}(x_2) \wedge I_{e2}(x_2) \wedge I_{f2}(x_2). \quad (2)$$

3. Точковий рухомий об'єкт радіолокації літак: тип зображення S_3, семантична ознака процесу формування зображення відмітки $I_{s3}(x_3)$ = 1 при діапазонах зміни ознаки розміру відмітки за азимутом $C9 \le D_{s3}^1\ (x_3) \le C10$ і ознаки розміру відмітки за дальністю $C11 \le D_{s3}^2\ (x_3) \le C12$; енергетична ознака $I_{e3}(x_3) = 1$ інтелектуальної моделі зображення пачки сигналів (відмітки) при сумі творів добутків амплітуд сигналів на предикати інформаційних комірок пачки у напрямку, що визначається вектором (l_n), $D_{e3}^l(x_3) = \sum_{l_1}^{l_n} q_{i,j+l_n} Z_{ai,j+l_n} > E3$; семантична ознака флуктуації зображень відміток $I_{f3}(x_3)$ = 1, коли підпростір значень спектрально-семантичної ознаки відмітки з метою оцінки швидкості руху об'єкта визначається як: ознака одиночної групи одиниць $D_{f3}^2(x_3) = 1$ при F1=2; ознака двох або більше груп одиниць при F1=4 або F1=6 (для двох або більше груп зімкнутих одиниць у множині результат підсумовування завжди дорівнює 4 або 6); ознака кількості нулів між групами одиниць $D_{f3}^4(x_3) = 1$ при $L_1 \le 3$ класифікує прийнятий сигнал як літак, а при $L_i > 3$ - як заваду; ознака кількості нулів від початку координат до першої групи одиниць $D_{f3}^5(x_3) = 1$ при $L_0 > 5$ класифікує прийнятий сигнал як літак. Тут C9 та C10 – мінімальна і максимальна тривалість пачок відміток за азимутом, C11 та C12 – мінімальна і максимальна тривалість пачок відміток за дальністю, $E3$ – мінімальне значення енергетичної ознаки, $F1$– арифметична сума предикатів у функції *F(y)*.

Предикатне рівняння ідентифікації має вид

$$S_3 = I_{s3}(x_3) \wedge I_{e3}(x_3) \wedge I_{f3}(x_3). \quad (3)$$

4. Практичне значення одержаних результатів

Проведено модельні експерименти з перевірки ефективності виявлення й ідентифікації літальних апаратів. Для порівняльної оцінки ефективності технології використовувалися семантичні та енергетичні ознаки інтелектуальної моделі сигнальної відмітки.

З аналізу результатів випливає, що при малих значеннях відношення сигнал/шум ймовірність виявлення й розпізнавання, за умови використання запропонованої технології обробки, значно вище, ніж за умови використання енергетичної (класичної) обробки. При відношенні сигнал/шум від 5 до 20 спільна можливість ймовірності виявлення й розпізнавання збільшується на величину від 0,328 до 0.497.

5. Висновки

Розроблена ознакова модель ідентифікації радіолокаційних об'єктів, таких як літак, вертоліт, БПЛА. Розроблено технологію формування й аналізу зображень радіолокаційних відміток, створено простір семантичних ознак для опису відміток об'єктів радіолокації, а також перелік операцій обробки.

References

[1] Skolnik M. I. (eds) (2021) Radar Handbook, McGraw-Hill, New York

[2] V. Zhyrnov, S. Solonska, Symbolic model of radar images when detecting aircraft, Telecommunications and Radio Engineering, Vol. 81-2, 2022, pp. 25-35.

[3] В.В. Жирнов, С.В. Солонська, Інтелектуальна модель зображень відміток радіолокаційних об’єктів для оглядових РЛС, Радіотехніка, Вип. 212, 2023, с. 148-154.

[4] В.В. Жирнов, С.В. Солонська Методи логічної обробки зображень відміток радіолокаційних об’єктів на основі семантичних ознак, Радіотехніка, Вип. 216, 2024, с. 120-125.

Utilization of Software Defined Radio (SDR) for Locating Radio Emission Sources

Ivan Bokhan[1†*] та Oleksii Bielousov[1†]

[1]*Kharkiv National University of Radio Electronics, 14 Nauky Avenue, Kharkiv, 61166, Ukraine*

Abstract
SDR enables the implementation of signal reception and processing functions at the software level, allowing for flexible adjustment of settings to suit various radiomonitoring tasks.

Keywords
DoA, RDF, TDOA, SDR, Correlation analysis, Spatial coordinates, python.

1. Relevance of application:

Nowadays the need for detecting and monitoring radio emission sources (RES) has become critical, especially in the context of contemporary military conflicts where unmanned aerial vehicles (UAVs) are used on a regular basis. Direction-finding systems based on Software Defined Radio (SDR) provide the capability to develop mobile and efficient solutions for monitoring unmanned aerial vehicles and other sources of radio signals.

2. Key techniques to locate radio emission sources using SDR

Trilateration and time difference of arrival (TDOA) method [2]. TDOA enables the determination of a source's location by using multiple receivers positioned at different locations. By measuring the time difference of signal arrival at each receiver, the system calculates the emitter's location by solving a system of equations.

Measurement of the direction of arrival (DoA) of a signal [1]. Algorithms such as MUSIC, ESPRIT, and MVDR/Capon enable the determination of the direction of radio emission sources, facilitating localization. This is achieved by measuring the phase differences between signals received by multiple antennas arranged in a defined geometric configuration.

Correlation Interferometry [1]. This method uses phase information from signals received by antennas to determine the direction of the signal source. The combination of high-resolution antennas and correlation algorithms ensures precise direction finding of the source. This is particularly useful for tracking drones and other high-speed moving objects.

3. Examples of the use of SDR for monitoring and localization

System for detection and protection against drones [1]. SDR-based systems, such as KrakenSDR or HackRF One, are actively used for the detection and neutralization of drones. Utilizing correlation and TDOA methods, these systems can enable to locate a drone and even identify its operator. This approach provides extensive functionality and flexibility in configuration to ensure the security of facilities.

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
*Corresponding author.
†These authors contributed equally.
ivan.bokhan@nure.ua(I. Bokhan); oleksii.bielousov@nure.ua(O.Bielousov)
0009-0002-9172-1710 (I. Bokhan); 0009-0000-0733-4044 (O.Bielousov)

Military applications for direction finding of signal sources. In modern military conflicts, SDR systems provide the capability to detect, locate, and track radio emission sources (RES), such as radio transmitters or enemy command systems. Systems with TDOA and DoA capabilities can be used for the rapid determination of signal source coordinates, which is especially critical for the protection of strategic positions and mobile military facilities.

Civil use in emergency response systems. SDR technologies can be used to locate emergency signals, such as those from individuals in distress in forested or mountainous areas where mobile communication is limited. The use of DoA and TDOA-based technologies enables precise detection of the distress signal's location, significantly speeding up response times.

4. Advantages and challenges of using SDR

Flexibility and adaptability. One of the key advantages of SDR is the ability to modify functionality through software. This allows the system to be quickly adapted to new frequencies, signal processing algorithms, and user requirements, which is especially relevant for countering unauthorized emission sources.

High accuracy and processing speed. The use of TDOA and DoA-based methods ensures high accuracy and speed, enabling SDR to operate in real-time, even under challenging conditions.

Challenges. SDR systems are sensitive to interference and require precise antenna configuration to ensure accuracy. Additionally, there are legal restrictions on the use of radio jamming technologies in many countries, which can limit the functionality of certain applications.

5. The Python programming language for data processing in SDR [3]

The use of Python for signal processing in SDR offers several advantages, making this tool highly popular among researchers and engineers.

Ease of learning and use. Python has a user-friendly and clear syntax, allowing beginners to learn quickly. This reduces the number of errors when writing code and enables users to focus on solving tasks rather than on the details of the language's syntax.

Extensive library support. Python offers a rich set of libraries for signal and data processing (NumPy, SciPy, Matplotlib, PySerial, GNU Radio). These libraries enable the rapid implementation of algorithms for signal analysis and processing.

Scalability and flexibility. Python easily integrates with other programming languages, such as C or C++, allowing the use of faster interfaces for specific components (e.g., real-time data processing). At the same time, Python retains flexibility due to its extensive capabilities for prototyping and testing various algorithms without requiring a complex compilation process.

Open-source software. Python and most libraries for SDR are open-source software. This reduces licensing costs and provides access to a vast array of resources, including documentation, communities, and forums for knowledge sharing.

Handling large volumes of data. Python performs well with large volumes of data, which is crucial for processing signals with high sampling rates. Libraries like NumPy and Pandas enable efficient handling of large datasets and complex mathematical operations required for radio signal analysis.

Large community. Python has an active community of developers working with SDR technologies. This allows for quick access to support, finding ready-made solutions, or educational materials for implementing SDR projects.

Capabilities for machine learning and signal analysis. Python is one of the primary languages for developing machine learning models. This opens up vast opportunities for applying machine learning algorithms to signal classification, noise filtering, spectrum analysis, and more. Libraries such as scikit-learn, TensorFlow, and PyTorch enable the creation and training of models that can be used to enhance SDR systems.

6. Conclusions

The effectiveness of SDR for modern localization tasks. SDR systems have become an indispensable tool for monitoring, detecting, and direction finding of emission sources due to their high flexibility, processing speed, and accessibility. They are capable of providing reliable frequency spectrum monitoring in the face of modern threats.

Development prospects. Further research in signal processing algorithms, such as improving the accuracy of DoA and TDOA methods, as well as enhancing hardware to reduce the impact of interference and increase resolution, opens up new opportunities for the application of SDR in various areas of security, defense, and the civilian sector.

Programming language. Python is an excellent choice for working with SDR due to its simplicity, rich libraries, ability to integrate with other programming languages, and extensive capabilities for developing and testing signal processing algorithms.

References

[1] Florin-Lucian CHIPER Ph.D. THESIS SUMMARY "Using the Radio Source Localization Methods in Drone Detection, Tracking and Neutralization Systems.» «Doctoral School of Electronics, Telecommunications and Information Technology» Decision No 2 of 06.09.2023.

[2] Stefan Scholl, TDOA Transmitter Localization with RTL-SDRs, Software Defined Radio Academy, Friedrichshafen, Germany, 07/2017 URL: https://panoradio-sdr.de/tdoa-transmitter-localization-with-rtl-sdrs.

[3] Dr. Mark Lichtman, PySDR: A Guide to SDR and DSP using Python, 05/2023 URL: https://pysdr.org/

DOI: https://doi.org/10.30837/IST-2024.P1.075

Розробка математичної моделі GPS спуфінгу в умовах роботи CRPA

Олександр Білик[1*†] та Олександр Мартинчук[1†]

[1] *Харківський національний університет радіоелектроніки, проспект Науки, 14, Харків, Україна*

Анотація
В дослідженні проведено аналіз та створення математичних моделей для оцінки ефективності GPS спуфінгових атак та засобів захисту сигналів навігації з адаптивних використанням антенних решіток (CRPA). Розроблено програму в середовищі моделювання Matlab та проаналізовано на основі моделі впливу спуфінгових атак на правильність прийому сигналів супутникової навігації.

Ключові слова
GPS, CRPA, spoofing, навігація, моделювання сигналів

1. Вступ

CRPA (Controlled Reception Pattern Antenna) — це технологія антен, яка використовує масив антенних елементів для контролю діаграми спрямованості прийому сигналу. Вона дозволяє керувати напрямком прийому сигналу та приглушувати завади, покращуючи якість прийому супутникових сигналів в GPS або GNSS системах. До основні переваг CRPA слід віднести:

1. Придушення завад: CRPA може активно приглушувати сигнали від джерел перешкод, що дозволяє надійно приймати сигнал навіть у складних умовах.
2. Покращена точність позиціонування: завдяки здатності адаптувати приймальну діаграму, CRPA може підвищити точність визначення місцезнаходження.
3. Захист від спуфінгу: така антена може бути більш стійкою до спроб фальсифікації сигналу.

GPS спуфінг — це форма впливу на прийом корисного сигналу, під час якого транслюються підроблені сигнали, які імітують справжні супутникові сигнали, з метою змусити приймач прийняти неправильні дані про своє місцезнаходження або час і в подальшому дезорієнтувати об’єкт у просторі.

2. Основні підходи спуфінгу в GPS/GNSS

1. Імітація супутникових сигналів: Спуфер (той, хто проводить атаку) генерує сигнали, що виглядають як справжні GPS/GNSS сигнали, але містять неправильну інформацію про позицію або час. Приймач, приймаючи ці підроблені сигнали, визначає своє місцезнаходження чи час неправильно.

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
* Corresponding author.
† These authors contributed equally.
oleksandr.bilyk@nure.ua (О.Білик); oleksandr.martynchuk@nure.ua (О.О.Мартинчук)
0009-0007-4846-1585 (О. Білик); 0000-0001-9729-4401 (О.О.Мартинчук)

2. Зміщення позиції: Однією з цілей спуфінгу є зміна координат, які отримує жертва. Наприклад, корабель або автомобіль можуть почати рух у неправильному напрямку через те, що їхній приймач GNSS показує помилкове місцезнаходження.
3. Зміна часу: Системи, які покладаються на точний час від супутників, можуть отримати неправильні дані про час, що може спричинити серйозні наслідки для банківських транзакцій, телекомунікацій або енергетичних систем, які синхронізуються через GNSS.

Створення процесу спуфінга:

- Підроблений сигнал: Зловмисник генерує сигнал, схожий на сигнал справжнього супутника. Цей сигнал зазвичай сильніший за оригінальний супутниковий сигнал, тому приймач помилково приймає його за правильний.
- Захоплення приймача: Після того, як приймач починає приймати підроблений сигнал, спуфер поступово зміщує дані про місцезнаходження чи час, змушуючи приймач показувати некоректну інформацію.
- Перешкоди в роботі: Це може вплинути на транспортні засоби (автомобілі, судна, дрони), які залежать від точного визначення місцезнаходження, і може призвести до серйозних аварій або порушень..

3. Розробка математичної моделі GPS спуфінгу в умовах роботи з CRPA

Розробка математичної моделі GPS спуфінгу в умовах роботи з CRPA (Controlled Reception Pattern Antenna) передбачає врахування таких аспектів:

1. Симуляція сигналу супутника: корисний сигнал із GPS/GNSS.
2. Симуляція спуфінгового сигналу: навмисно сформований завадами сигнал із характерними затримками, зміщенням частоти та поляризаційними параметрами.
3. CRPA-система: антени, які забезпечують просторово-поляризаційну обробку сигналу для придушення спуфінгу.

Спуфінгові сигнали, які намагаються імітувати корисний сигнал, мають схожу фазу та частоту. CRPA здатна розрізняти такі сигнали, оцінюючи фазові зсуви та амплітуду сигналів з різних елементів антенної решітки:

$$\Delta\varphi = arg\ arg\left(W^H y(t)\right) - arg(y(t)) \quad (1)$$

Модель прийнятого сигналу:

$$r(t) = A_{GPS}\ cos\cos(2\pi f_{GPS}t + \varphi_{GPS}) + \\ + A_{spoof}\ cos\ cos\left(2\pi f_{spoof}t + \varphi_{spoof} + \Delta\varphi\right) \quad (2)$$

Основними методами протидії спуфінгу є:

- Аналіз фазових та амплітудних зсувів у системі.
- Використання методів спектрального аналізу для ідентифікації спуфінгових сигналів.

На рисунку 1 представлені результати моделювання за допомогою програмного забезпечення Matlab. На графіках можна побачити нормалізовані значення АЧХ корисного сигналу, спуфінгового сигналу (постановника завади) та результуючий сигнал після обробки з виходу CRPA.

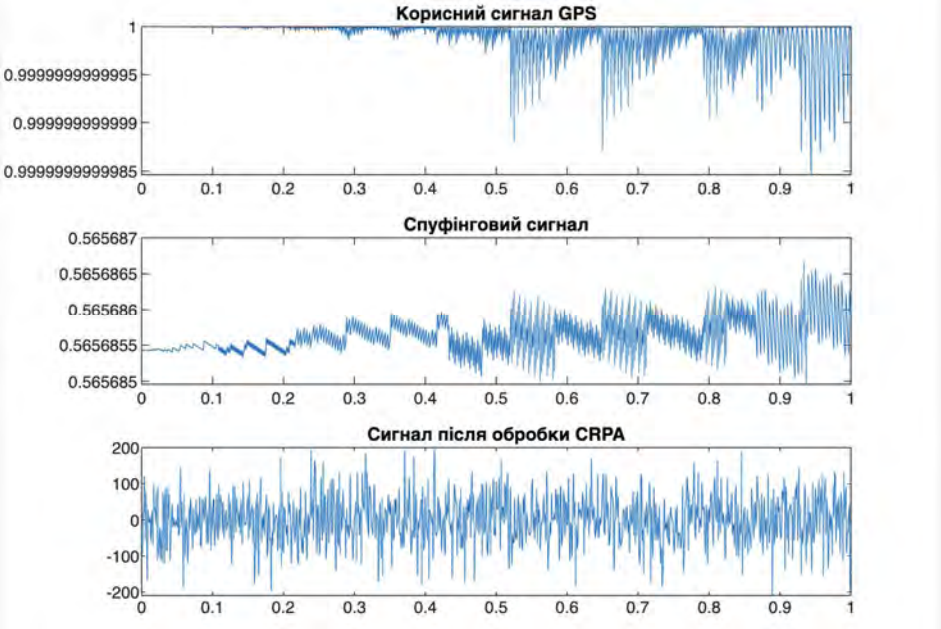


Рисунок 1. Моделювання процесу GPS спуфінгу.

1. Створення сигналів: GPS та спуфінговий сигнал мають різну частоту й фазу, що імітує реальні умови.
2. CRPA: вектор напрямної діаграми враховує просторові параметри.
3. Придушення завад: застосовано інверсію коваріаційної матриці для оптимального формування діаграми направленості.
4. Візуалізація: Аналіз ефективності CRPA в придушенні спуфінгу.

Дана модель може бути розширена додаванням поляризаційного аналізу та більш складними умовами поширення.

4. Висновки

Використання адаптивних антенних решіток та алгоритмів просторово-часової обробкиу навігаційних системах значно підвищує стійкість сигналів на виході антени. Для покращення аналізу можливого впливу на системи навігації дана модель може бути розширена додаванням поляризаційного аналізу та більш складними умовами поширення.

References

[1] A. Altaweel, H. Mukkath and I. Kamel, "GPS Spoofing Attacks in FANETs: A Systematic Literature Review," in IEEE Access, vol. 11, pp. 55233-55280, 2023, doi: 10.1109/ACCESS.2023.3281731.

[2] E. Horton and P. Ranganathan, "Development of a GPS spoofing apparatus to attack a DJI Matrice 100 quadcopter", J. Global Positioning Syst., vol. 16, no. 1, pp. 1-11, Dec. 2018.

[3] T. Humphreys, Statement on the Vulnerability of Civil Unmanned Aerial Vehicles and Other Systems to Civil GPS Spoofing, Austin, TX, USA, pp. 1-16, Jul. 2012.

[4] I. Bisio, C. Garibotto, H. Haleem, F. Lavagetto and A. Sciarrone, "A systematic review of drone based road traffic monitoring system", IEEE Access, vol. 10, pp. 101537-101555, 2022.

Секція 5.
Інформаційні технології в соціумі, освіті, економіці, медицині, управлінні, цивільному захисті, поліграфії, екології та юриспруденції

Neural Network Models for Stroke Rehabilitation

Oleksandr Krupenia[1†], Viktor Kauk[1*†] and Pavlo Kauk[1†]

[1] National University of Radio Electronics, Nauky Avenue 14 61166 Kharkiv. Ukraine

Abstract

This review explores the development of neural network models for stroke rehabilitation, emphasizing their potential to enable personalized, patient-centered interventions that improve outcomes and promote recovery. By integrating machine learning and neurobiological advancements, these models can predict patient outcomes, identify risk factors, and enhance rehabilitation strategies, addressing limitations in traditional standardized care. This paper provides an overview of current research and discusses the opportunities and challenges of implementing these models in clinical practice.

Keywords

neural network models, stroke rehabilitation, personalized recovery, deep learning

1. Introduction

Stroke rehabilitation presents significant challenges for healthcare systems worldwide, with traditional standardized protocols often failing to address individual patient needs effectively. The integration of neural network models into rehabilitation practices has emerged as a promising approach to overcome these limitations.

Recent advances in deep learning architectures have enabled more sophisticated approaches to stroke rehabilitation, processing complex data from multiple sources, including electroencephalography (EEG) signals and movement patterns. These models encompass several key architectures, including Deep Neural Networks (DNN), Improved Deep Neural Networks (IDNN), and Long Short-Term Memory Networks (LSTM), each offering distinct advantages in processing different types of rehabilitation data.

This paper provides a comprehensive review of the current state of the art in neural network applications in stroke rehabilitation. It focuses on their practical implementation, effectiveness, and limitations. It also explores how these technologies improve traditional therapeutic approaches and considers future developments that will contribute to the further development of neurological rehabilitation.

2. Types of Neural Network Models

2.1. Deep Neural Networks (DNN)

Deep Neural Networks employ multiple hidden layers between input and output layers, utilizing a two-step learning process. The network first undergoes unsupervised learning to pre-train network layers, followed by supervised learning that utilizes these pre-trained weights as initial values to optimize performance. This approach effectively mitigates overfitting risks while enabling complex pattern recognition in stroke rehabilitation data, particularly in EEG signal classification for patient assessment and device control [1].

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine

* Corresponding author.

† These authors contributed equally.

oleksandr.krupenia@nure.ua (O. Krupenia); viktor.kauk@nure.ua (V. Kauk); pavlo.kauk@nure.ua (P. Kauk)

0009-0005-5237-7728 (O. Krupenia); 0000-0002-2780-2666 (V. Kauk); 0009-0009-1900-6609 (P. Kauk)

2.2. Improved Deep Neural Networks (IDNN)

IDNNs enhance the traditional DNN framework by incorporating Large Margin Support Vector Machine methodologies. These models optimize edge distribution for improved classification accuracy, which is particularly beneficial when processing high-aliasing characteristics of stroke data. Experimental results demonstrate superior performance in EEG classification, making IDNNs valuable for developing responsive rehabilitation systems [1].

2.3. Convolutional Neural Networks (CNN)

CNNs specialize in processing spatial information from video data, which is particularly useful for analyzing movement patterns during rehabilitation exercises [1]. Through convolutional filters, these networks capture complex visual patterns essential for monitoring patient progress and adapting training protocols accordingly.

2.4. Long Short-Term Memory Networks (LSTM)

LSTMs, a specialized form of Recurrent Neural Network, excel at capturing temporal dependencies in data sequences. While offering superior accuracy in analyzing time-series EEG data, these networks typically require longer processing times than simpler models, presenting a trade-off between performance and computational efficiency [1].

2.5. Graph Isomorphic Networks (GIN)

GINs represent an emerging approach for processing graph-structured data, particularly effective for analyzing EEG-EMG interactions. Recent implementations have achieved a classification accuracy of 88.89%, surpassing conventional methods while providing real-time feedback during rehabilitation sessions [2].

3. Applications in Stroke Rehabilitation

In cognitive rehabilitation, neural network models support recovery through personalized training programs targeting specific functions, with IDNNs demonstrating superior classification accuracy compared to traditional BP Neural Networks, DNNs, CNNs, and LSTM models, particularly in EEG data processing for cognitive assessment [1]. These systems leverage neuroplasticity principles through individualized cognitive exercises and real-time performance assessment.

Robotic rehabilitation systems enhanced by neural networks provide mechanical support and guidance during therapeutic exercises [3, 4]. These systems enable real-time monitoring of movement patterns, adaptive assistance levels based on patient performance, and task-specific training that mimics daily activities through portable home-based devices. The IDNN's performance in controlling these systems has demonstrated high accuracy while maintaining reasonable processing times, comparable to LSTM models [1, 3, 4]. This balance between accuracy and computational efficiency makes it particularly suitable for real-time rehabilitation applications.

The application of machine learning methods in stroke rehabilitation facilitates the classification of stroke residual severity and the monitoring of patient progress during therapy sessions [5]. By analyzing kinematic data collected from in-home therapy, neural networks can assist clinicians in optimizing treatment plans tailored to each patient's specific needs, enhancing the effectiveness of rehabilitation efforts [5, 6]. This data-driven approach not only improves therapeutic outcomes but also empowers stroke survivors by promoting autonomous recovery.

4. Challenges and Limitations of Neural Network Models in Stroke Rehabilitation

The implementation of neural network models in stroke rehabilitation faces several key challenges that impact their effectiveness and widespread adoption.

First, data quality and availability remain significant limitations, with many studies relying on limited sample sizes or unstructured datasets. A recent analysis of 174 models revealed that only 89 had a low risk of bias, while 77 were classified as high-risk due to inadequate handling of missing data and overfitting concerns. Building robust digital infrastructures for systematic data collection and establishing standardized validation protocols could enhance model reliability [7].

Second, computational complexity presents an ongoing challenge as deep learning models offer greater accuracy but require significant computational resources. While the Back Propagation (BP) neural network achieves faster running times, more sophisticated architectures like IDNN must balance accuracy with operational efficiency [1, 8].

Additionally, limited validation poses a substantial methodological concern. Only 3 of 19 primary studies reported external validation, raising concerns about model generalizability. Most studies relied solely on internal validation, potentially compromising the robustness of predictive models in real-world applications [8, 9].

5. Future Directions in Neural Network Models for Stroke Rehabilitation

The advancement of neural network models in stroke rehabilitation focuses on two key areas that promise to enhance patient outcomes and treatment effectiveness:

5.1. Interdisciplinary Collaboration

The future of stroke rehabilitation relies on fostering collaboration between clinical and computational experts, while leveraging machine learning for individualized treatment. This approach enables providers to design personalized rehabilitation plans based on predictive models derived from clinical and neurophysiological data. Regular interactions between researchers and clinicians are essential for effectively transitioning neurorehabilitation devices into practice [3, 10, 11].

5.2. Technological Advancement

The integration of artificial intelligence technologies, such as AIoT and data technology, offers potential for enhancing healthcare automation in rehabilitation. Neural interfaces that deliver real-time feedback show particular promise for improving motor recovery by reinforcing brain-muscle connections. These closed-loop systems, combining neurophysiological sensing with peripheral stimulation, demonstrate significant potential for creating adaptive rehabilitation protocols that respond to patient progress [9, 12].

Conclusion

Neural network models offer promising, personalized approaches to stroke rehabilitation, using architectures like DNNs, IDNNs, CNNs, and LSTMs to improve diagnostic accuracy, cognitive support, and robotic-assisted therapy. While challenges like data quality, computational demands, and validation persist, interdisciplinary collaboration and enhanced data infrastructures are crucial for overcoming these. Future research on neural interfaces and feedback systems could make these models more accessible and effective in clinical settings.

References

[1] Gao, M., & Mao, J. (2021). A novel active rehabilitation model for stroke patients using electroencephalography signals and deep learning technology. Frontiers in Neuroscience, 15. https://doi.org/10.3389/fnins.2021.780147

[2] Li, H., Ji, H., Yu, J., Li, J., Jin, L., Liu, L., Bai, Z., & Ye, C. (2023). A sequential learning model with GNN for EEG-EMG-based stroke rehabilitation BCI. Frontiers in Neuroscience, 17. https://doi.org/10.3389/fnins.2023.1125230

[3] Li, X., He, Y., Wang, D., & Rezaei, M. J. (2024). Stroke rehabilitation: from diagnosis to therapy. Frontiers in Neurology, 15. https://doi.org/10.3389/fneur.2024.1402729

[4] Jeter, R., Greenfield, R., Housley, S. N., & Belykh, I. (2024). Classifying Residual Stroke Severity using Robotics-Assisted Stroke Rehabilitation: A Machine Learning Approach (Preprint). JMIR Biomedical Engineering, 9, e56980. https://doi.org/10.2196/56980

[5] Zu, W., Huang, X., Xu, T., Du, L., Wang, Y., Wang, L., & Nie, W. (2023). Machine learning in predicting outcomes for stroke patients following rehabilitation treatment: A systematic review. PLoS ONE, 18(6), e0287308. https://doi.org/10.1371/journal.pone.0287308

[6] Lin, D. J., Backus, D., Chakraborty, S., Liew, S., Valero-Cuevas, F. J., Patten, C., & Cotton, R. J. (2024). Transforming modeling in neurorehabilitation: clinical insights for personalized rehabilitation. Journal of NeuroEngineering and Rehabilitation, 21(1). https://doi.org/10.1186/s12984-024-01309-w

[7] Campagnini, S., Arienti, C., Patrini, M., Liuzzi, P., Mannini, A., & Carrozza, M. C. (2022). Machine learning methods for functional recovery prediction and prognosis in post-stroke rehabilitation: a systematic review. Journal of NeuroEngineering and Rehabilitation, 19(1). https://doi.org/10.1186/s12984-022-01032-4

[8] Norman, S. L., Wolpaw, J. R., & Reinkensmeyer, D. J. (2022). Targeting neuroplasticity to improve motor recovery after stroke: an artificial neural network model. Brain communications, 4(6), fcac264. https://doi.org/10.1093/braincomms/fcac264

[9] Huang, Y., Yang, B., Wong, T. W., Ng, S. S. M., & Hu, X. (2024). Personalized robots for long-term telerehabilitation after stroke: a perspective on technological readiness and clinical translation. Frontiers in Rehabilitation Sciences, 4. https://doi.org/10.3389/fresc.2023.1329927

[10] Maier, M., Ballester, B. R., & Verschure, P. F. M. J. (2019). Principles of neurorehabilitation after stroke based on motor learning and brain plasticity mechanisms. Frontiers in Systems Neuroscience, 13. https://doi.org/10.3389/fnsys.2019.00074

[11] Sharma, V., Santos, F. P. D., & Verschure, P. F. M. J. (2023). Patient-specific modeling for guided rehabilitation of stroke patients: the BrainX3 use-case. Frontiers in Neurology, 14. https://doi.org/10.3389/fneur.2023.1279875

[12] Vidaurre, C., Irastorza-Landa, N., Sarasola-Sanz, A., Insausti-Delgado, A., Ray, A. M., Bibián, C., Helmhold, F., Mahmoud, W. J., Ortego-Isasa, I., López-Larraz, E., Peiteado, H. L., & Ramos-Murguialday, A. (2023). Challenges of neural interfaces for stroke motor rehabilitation. Frontiers in Human Neuroscience, 17. https://doi.org/10.3389/fnhum.2023.1070404

DOI: https://doi.org/10.30837/IST-2024.P1.083

Особливості визначення точки роси за вуглеводнями природного газу на підземних сховищах газу

Віктор Луценко[1], Юрій Пономарьов[1], Світлана Протас[1] та Вадим Краснокутський[1]

[1] *Інститут транспорту газу" АТ "Укртрансгаз", 16, бульвар Гончарівський, Харків, 61004, Україна*

Анотація

У статті використано матеріали досліджень Американської школи газовимірювальних технологій (ASGMT) та практичні спостереження фахівців «Укртрансгазу», отримані під час введення в експлуатацію відповідного обладнання на підземних сховищах газу. За результатами їх аналізу у звіті зроблено ґрунтовні висновки.

Ключові слова

природний газ; точка роси вуглеводнів; рівняння стану; невизначеність кореляції

Abstract

The article uses research materials of the American School of Gas Measurement Technology (ASGMT) and practical observations of Ukrtransgaz specialists obtained during the commissioning of the relevant equipment at underground gas storage facilities. Based on the results of their analysis, the report draws substantial conclusions.

Keywords

natural gas; hydrocarbon Dew Point; equation of state; uncertainty of the correlation

1. Вступ

У процесі добування, транспортування, зберігання, розподілення природного газу технічні умови контрактів висувають цілком конкретні вимоги до параметрів його якості. Однією з властивостей, що визначає якість природного газу, є температура точки роси за вуглеводнями (HCDP - HydroCarbon DewPoint).

Стандарти [1] та [2] визначають HCDP як температуру за заданого тиску, за якої пари вуглеводнів починають конденсуватися.

Наявна велика кількість способів і методів вимірювання HCDP. Найбільш відомим є пряме вимірювання за допомогою ручних та/або автоматичних приладів із охолоджуваним дзеркалом.

Крім того, є непрямий метод визначання HCDP, із визначенням компонентного складу за допомогою газового хроматографа (ГХ). Вхідною інформацією для аналітичного розрахунку HCDP є компонентний склад природного газу, отриманий за допомогою ГХ, разом із тиском проби газу.

* Corresponding author.
† These authors contributed equally.
lutcenko-va@utg.ua (В. Луценко); ponomarev-yv@utg.ua (Ю. Пономарьов); protas-sa@utg.ua (С. Протас); krasnokutskiy-vs@utg.ua (В. Краснокутський)
0000-0003-0491-3280 (В. Луценко); 0000-0002-5677-3657(Ю. Пономарьов); 0009-0009-6144-6835 (С. Протас); 0009-0009-6823-2661 (В. Краснокутський)

2. Щодо аналітичного визначання HCDP за компонентним складом від газового хроматографа

У доповіді використані матеріали досліджень Американської школи газовимірювальної техніки (ASGMT) та практичні спостереження фахівців АТ «Укртрансгаз».

Типи помилок, які погіршують розрахунки HCDP: помилка розподілу важких вуглеводнів у суміші; помилка, що виникає під час зниження тиску; помилка від похибки газового хроматографа; помилка, що обумовлена використанням конкретного рівняння стану; помилка, одержувана під час відбору репрезентативної проби для ХАЛ.

2.1. Помилка розподілу важких вуглеводнів у суміші

Більшість ГХ, що використовують в умовах експлуатації, вимірюють компонентний склад природного газу до С6+. Тобто, такі ГХ вимірюють N_2, CO_2, C_1-C_5, а всі більш важкі вуглеводні об'єднують в групу С6+.

Таку практику є сенс використовувати для визначення теплотворної здатності. Але для аналітичного визначення HCDP вона неприйнятна, тому що HCDP залежить від концентрації важких вуглеводнів.

Були спроби вводити C_9+GC. Це частково вирішувало проблему, оскільки всі компоненти, важчі за C_9, містяться в одному контейнері. Невизначеність у такому випадку складала до 20 ^{0}C.

2.2. Помилка, що виникає під час зниження тиску

ГХ працюють за тиску, який близький до атмосферного. Однак більшість операцій, що стосуються транспортування газу трубопроводами, зберігання його в підземних сховищах, стиснення тощо, виконують за підвищених надлишкових тисків до 150 бар. Таке зниження тиску може суттєво змінити склад газу, якщо тільки його не виконували в кілька етапів зі значним зовнішнім нагріванням.

2.3. Помилка від похибки газового хроматографа

Як і всі прилади, ГХ має притаманну йому похибку. Похибка визначання важких вуглеводнів зазвичай більша. Під час вимірювання концентрації компонентів природного газу виробники ГХ вказують невизначеність (похибку), що дорівнює 2 %. Однак, типові ГХ в умовах експлуатації, фактично мають похибку 5 % і більше.

2.4. Помилка, що обумовлена використанням конкретного рівняння стану

На сьогодні відома велика кількість рівнянь стану для визначення властивостей реального газу або суміші газів, якою є природний газ [5-8].

Найбільш популярні – модель Пенга-Робінсона (PR), модель Соаве-Редліх-Квонга (SRK), модель GERG-2008. Деякі з цих моделей кращі за умови високих тисків, деякі – за низьких тисків. Можна застосувати дві різні моделі та отримати два абсолютно різних результати для розрахунку HCDP.

2.5. Помилка, одержувана під час відбору репрезентативної проби для ХАЛ

З огляду на всі перераховані вище помилки виникає спокуса отримувати пробу газу, використовуючи більш технологічний ГХ до С12. Хоча ГХ С12 усуває деякі помилки розподілу важких вуглеводнів, він додає ще один набір помилок, а саме похибку відбирання та транспортування проби[6].

Більш того, коли пробу транспортують і потім повторно нагрівають, після використання в лабораторії, може бути внесена додаткова невизначеність.

Похибки обладнання та результати його експлуатування на ПСГ АТ «Укртрансгаз» (експериментального або промислового) наведені в таблиці 1.

Таблиця 1
Похибки обладнання

Найменування	Принцип роботи	Похибка	Виробник	Країна
Торос	Насичення газу вологою за підвищеного тиску зниженням тиску до робочого	±0,5 ^{0}C	ПП "Прибор-центр", м. Київ.	Україна
ФОГ-3Г	Конденсаційний вузол вимірювача виконано за класичною схемою	±1 ^{0}C	«Хімтест Україна +»	Україна
Dewpoint Duo	Використовує метод CEIRS™[8]	±0,5 ^{0}C	ZEGAZ Instruments	USA

Висновки

1. Точність розрахунку температури HCDP на основі газової хроматографії не може бути вищою за точність вимірювання температури HCDP автоматизованим пристроєм з охолоджуваним дзеркалом.
2. Пристрої, принцип дії яких заснований на використанні інфрачервоної спектроскопії, мають велику перспективу використання для аналітичного визначення HCDP (наприклад, технології $CEI\ RS^{TM}$).
3. ПосиланняДля моніторингу працездатності потокового хроматографа щодо визначення HCDP, доцільно використовувати аналітичний алгоритм на основі рівняння стану (рівняння стану, при цьому треба підбирати для конкретних умов і сумішей природного газу).

Посилання

[1] ДСТУ ISO/TR 11150:2016 (ISO/TR 11150:2007, IDT) Природний газ. Точка роси вуглеводнів та вміст вуглеводнів.

[2] ДСТУ EN ISO 14532:2022 (EN ISO 14532:2017, IDT; ISO 14532:2014, IDT) Природний газ. Словник термінів.

[3] ДСТУ EN ISO 23874:2021 (EN ISO 23874:2018, IDT; ISO 23874:2006, IDT) Природний газ. Вимоги до аналізування методом газової хроматографії для обчислювання точки роси за вуглеводнями.

[4] Eric Kelner and Darin L. George, American School of Gas Measurement Technology, Sept. 2008. 1.

[5] Darin L. George, Ph.D., Andy M. Barajas and Russell C. Burkey, Pipeline & Gas Journal, Sep. 2005.

[6] Sohrab Zarrabian, NGSTECH, January 2014 iv Sohrab Zarrabian, Mahmood Moshfeghian, GPA Annual Convention, Apr. 2013.

[7] K. K. Shah, G. Thodos, Ind. Eng. Chem., 1965, 57 (3), pp 30–37.

[8] Sohrab Zarrabian, Measuring hydrocarbon and water dewpoints, ZEGAZ Insnruments,2023.

Автоматизоване робоче місце викладача при освітньому процесі в дистанційній формі

Ігор Сокорчук[1]

[1] *Харківський національний університет радіоелектроніки, проспект Науки, 14, Харків, 61166, Україна*

Анотація

Розглянуто використання автоматизованого робочого місця викладача (АРМ) для підвищення ефективності організації навчального процесу в умовах дистанційної освіти. Розглянуто основні функції та компоненти АРМ, що включають інтеграцію з платформами управління навчанням, системами відеоконференцій, аналітичними інструментами та інструментами для оцінювання. Особливу увагу приділено використанню платформи GitHub для підтримки освітнього процесу, автоматизації перевірки завдань та організації спільної роботи. Окремо розглянуто роль утиліт командного рядка Linux у задачах автоматизації, перевірки студентських робіт, адміністрування та управління навчальними матеріалами. Запропоновано методи інтеграції сучасних інструментів і платформ, які сприяють оптимізації навчального процесу, зменшенню навантаження на викладачів та підвищенню якості освіти.

Keywords

автоматизоване робоче місце викладача (АРМ), дистанційна освіта, платформа GitHub, утиліти командного рядка Linux, автоматизація перевірки, управління навчальними матеріалами, аналітичні інструменти, платформи управління навчанням (LMS)

1. Вступ

Сучасний освітній процес в умовах глобальної цифровізації й активного впровадження дистанційного навчання вимагає нових підходів до організації роботи викладача. Автоматизоване робоче місце викладача (АРМ викладаа) є ключовим інструментом для ефективного управління освітньою діяльністю. Такі засоби дозволяють оптимізувати розробку навчальних матеріалів, взаємодію зі студентами, контроль успішності та автоматизацію адміністративних завдань.

Наведемо основні функції АРМ викладача. АРМ викладача об'єднує технології для вирішення таких завдань:

1. Планування навчальних матеріалів. Розробка й упорядкування лекцій, практичних занять і тестів у цифровому середовищі.
2. Організація дистанційної взаємодії. Інтеграція з платформами відеоконференцій (Zoom, Google Meet) для проведення лекцій і консультацій.
3. Контроль успішності. Використання інструментів моніторингу й аналізу для своєчасного виявлення проблем у навчанні.
4. Автоматизація оцінювання. Автоматична перевірка тестів і завдань для забезпечення об’єктивності оцінювання.
5. Адміністративна підтримка. Ведення електронних журналів, звітності та обмін інформацією зі студентами.

* Corresponding author.
ihor.sokorchuk@nure.ua (І. Сокорчук)
0000-0002-5852-7214 (І. Сокорчук)

2. GitHub у складі АРМ викладача

GitHub [1] є багатофункціональним інструментом, який можна ефективно інтегрувати до АРМ викладача. Цей сервіс надає такі функції, які можуть бути інтегровані до автоматизованого місця викладача та ефективно використовуватись у дистанційному навчанні:

1. Контроль версій і спільна робота. Зберігання навчальних матеріалів із можливістю колективного редагування, створення студентських проєктів.
2. Інтеграція з автоматизацією. Використання GitHub Actions для автоматичної перевірки лабораторних робіт.
3. Зберігання навчальних матеріалів. Структурування й управління доступом до цифрових ресурсів.
4. Зворотний зв'язок. Функції Pull Requests для рецензування робіт студентів.
5. GitHub сприяє підвищенню ефективності роботи викладача, зокрема завдяки автоматизації рутинних процесів і розвитку у студентів навичок роботи з реальними інструментами командної співпраці.

2.1. Переваги GitHub для АРМ викладача

Інтеграція GitHub із АРМ надає низку переваг:

1. Полегшення спільної роботи: GitHub спрощує роботу в командах, що стимулює розвиток комунікативних і технічних навичок у студентів. Вони вчаться працювати над спільним кодом, вирішувати конфлікти злиття та підтримувати якість коду.
2. Ефективність перевірки та зворотного зв'язку: Завдяки автоматизації, викладачі можуть швидко перевіряти завдання та надавати коментарі в режимі реального часу, знижуючи навантаження на ручну перевірку та прискорюючи процес оцінювання.
3. Відкритість і прозорість: Використання GitHub у освітньому процесі стимулює студентів до роботи в умовах відкритого коду, що підвищує відповідальність за якість власної роботи і сприяє об'єктивному оцінюванню.
4. Мотивування студентів: Студенти отримують можливість побачити власний прогрес і переглядати історію змін, що сприяє підвищенню їхнього інтересу до навчального процесу.

2.2. Перспективи використання GitHub у вищій освіті

У майбутньому GitHub може інтегрувати нові інструменти на основі штучного інтелекту (ШІ) та машинного навчання для поліпшення процесів оцінювання та забезпечення адаптивного навчання. Наприклад, використання систем на основі ШІ для автоматичного аналізу коду і виявлення стилістичних помилок може підвищити якість кодування у студентів. Крім того, GitHub може стати основною платформою для створення навчальних спільнот, де студенти зможуть брати участь у проєктах, зберігаючи власні роботи як портфоліо, доступні для потенційних роботодавців.

3. Використання утиліт командного рядка Linux у складі АРМ

Використання операційної системи Linux у поєднанні з утилітами командного рядка (CLI) забезпечує платформу для автоматизації завдань, управління навчальними матеріалами та перевірки студентських робіт.

3.1. Автоматизація перевірки студентських робіт

CLI утиліти Linux [2] дозволяють створити ефективні інструменти для автоматичної перевірки лабораторних і практичних робіт:

bash, diff, cmp – перевірка виконання завдань і порівняння результатів із зразками.

awk, sed – аналіз і обробка текстових даних.

3.2. Автоматизація перевірки оформлення програмного коду

Якість програмного коду є важливим аспектом, особливо для дисциплін, пов'язаних із програмуванням. Утиліти CLI дозволяють викладачеві автоматизувати перевірку стилю та синтаксису [3]: shellcheck, phpcs, shfmt – перевірка скриптів і коду на відповідність стандартам стилю.

3.3. Управління навчальними матеріалами

Утиліти командного рядка Linux спрощують роботу з навчальними матеріалами та файлами [4]:

rsync, tar, gzip – синхронізація, архівування й передача матеріалів.

find, grep – швидкий пошук і фільтрація даних.

3.4. Автоматизація адміністрування

CLI інструменти також можуть використовуватися для управління системами дистанційного навчання [5], наприклад такими як Moodle [6]:

curl, wget – взаємодія з LMS і завантаження даних.

crontab – автоматизація регулярних задач, таких як резервне копіювання.

3.5. Переваги програмних CLI утиліт Linux на АРМ викладача

1. Гнучкість і адаптивність. CLI утиліти легко налаштовуються під конкретні потреби викладача, дозволяючи створювати автоматизовані рішення для широкого спектра задач.
2. Продуктивність. Використання CLI забезпечує високу швидкість виконання операцій, що особливо важливо при роботі з великими обсягами даних.
3. Безкоштовність і відкритість. Більшість CLI утиліт є Open Source, що знижує витрати та дозволяє викладачам адаптувати інструменти до своїх потреб.

4. Засоби штучного інтелекту на АРМ викладача

На автоматизованому робочому місці (АРМ) викладача також можуть використовуватися засоби штучного інтелекту (ШІ) [7]. Перевагами використання ШІ та МН на АРМ викладача є:

1. Індивідуальні рекомендації щодо навчальних матеріалів та завдань сприяють більш глибокому розумінню тем, підвищують мотивацію студентів і покращують ефективність навчання.
2. ШІ може автоматизувати рутинні завдання, виконувати автоматичну перевірку робіт, оцінювати тести та надавати зворотний зв'язок студентам, що дозволяє звільнити викладача для більш творчих і стратегічних завдань.
3. Моделі машинного навчання здатні на ранніх стадіях виявляти ризики академічної неуспішності, аналізуючи поведінку студентів у навчальному середовищі.

4. На основі аналізу великих обсягів даних (Big Data) можна оптимізувати навчальні процеси, визначати найбільш ефективні методи та стратегії навчання для конкретних груп студентів. Таким чином, інтеграція засобів ШІ та МН в АРМ викладача відкриває нові можливості для підвищення якості освіти, робить навчальний процес більш ефективним, гнучким і орієнтованим на потреби студентів.

5. Результати впровадження та висновки

Описані програмні засоби та принципи були використані автором під час розробки автоматизованого робочого місця (АРМ) викладача для проведення занять у дистанційній формі навчання за спеціальністю «Інженерія програмного забезпечення». Розроблена система використовувалася в освітньому процесі протягом двох років і продемонструвала високий рівень ефективності.

Застосування автоматизованого робочого місця дозволило оптимізувати процеси підготовки, проведення занять і перевірки студентських робіт. Зокрема, було досягнуто таких результатів:

1. Автоматизація перевірки лабораторних та курсових робіт за допомогою систем контролю версій і автоматичного тестування (зокрема, GitHub Actions) дозволила забезпечити швидкий зворотний зв’язок для студентів, підвищила якість навчального процесу.
2. Завдяки інтеграції інструментів автоматичної перевірки коду (PHP_CodeSniffer, shfmt, ShellCheck) та платформи Moodle, зменшилася кількість рутинних завдань, пов’язаних із перевіркою робіт, що заощадило час викладача.
3. Використання АРМ забезпечило централізоване зберігання та доступ до навчальних матеріалів, використовуючи хмарні рішення, що підвищило зручність взаємодії між викладачем і студентами, покращило управління навчальними матеріалами.
4. Можливість налаштування середовища відповідно до специфіки дисциплін, його гнучкість і адаптивність, дозволили адаптувати це середовище під різні навчальні дисципліни.

Загалом, впровадження АРМ викладача на основі Open Source та Freeware рішень забезпечило позитивний вплив на якість та організацію освітнього процесу.

Перелік посилань

[1] GitHub – офіційна сторінка сервісу. URL: https://github.com/
[2] GNU Bash, Awk, sed, tar, gzip, grep, wget – офіційна сторінка. URL: https://www.gnu.org/software/
[3] ShellCheck – офіційна сторінка інструмента. URL: https://www.shellcheck.net/
[4] Rsync – офіційна сторінка. URL: https://rsync.samba.org/
[5] Curl – офіційна сторінка. URL: https://curl.se/
[6] Moodle Documentation URL: https://docs.moodle.org
[7] Сокорчук І. П. Автоматизація повсякденних завдань викладачів за допомогою засобів штучного інтелекту: покращення робочого процесу та ефективності / І. П. Сокорчук // Поліграфічні, мультимедійні та web-технології: тези доп. IX Міжнар. наук.-техн. конф., 14-18 травня 2024 р. – Т. 1. – Харків: ТОВ «Друкарня Мадрид», 2024. – С. 330-331.

Звітна документація при освітньому процесі в дистанційній формі

Ігор Сокорчук[1]

[1] *Харківський національний університет радіоелектоніки, проспект Науки, 14, Харків, 61166, Україна*

Анотація

Розглянуто способи автоматизації процесів створення та форматування науково-технічних документів відповідно до сучасних державних стандартів України, зокрема ДСТУ 3008:2015 [1]. Зроблено акцент на використанні мови розмітки Markdown [2], мови опису діаграм Plan-tUML [3] та текстового процесора LibreOffice Writer [4] для ефективного створення структурованих документів, таких як звіти, статті, та інші науково-технічні документи.

Keywords

звітна документація, дистанційна освіта, автоматизація, електронний документообіг, ДСТУ 3008:2015, Markdown, PlantUML

1. Вступ

У сучасних умовах стрімкого розвитку інформаційно-комунікаційних технологій дистанційна форма навчання набуває дедалі більшого поширення. Це зумовлено, як глобалізаційними процесами, так і необхідністю забезпечення доступу до якісної освіти незалежно від місця перебування студентів. Разом з тим, зростає роль ефективної організації звітної документації, що є важливим елементом управління освітнім процесом.

2. Особливості звітної документації при дистанційному навчанні

Звітна документація у системі дистанційної освіти охоплює широкий спектр документів, які можна класифікувати за такими основними групами:

1. Адміністративна звітність — журнали занять, журнали обліку відвідуваності, графіки та розклади освітнього процесу, накази та розпорядження адміністрації.
2. Методична звітність. — плани занять, методичні вказівки, розробки дистанційних курсів та мультимедійних навчальних матеріалів.
3. Оцінювальна звітність. — звіти з лабораторних та інших робіт, курсові та дипломні роботи, результати контрольних заходів.

2.1. Вимоги до звітної документації в дистанційному форматі

Згідно з ДСТУ 3008:2015 [1], в Україні звітна документація повинна мати чітко визначену структуру та відповідати встановленим стандартам оформлення. Звітний документ складається із титульної сторінки, змісту, основної частини, висновків, додатків.

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
ihor.sokorchuk@nure.ua (І. Сокорчук)
0000-0002-5852-7214 (І. Сокорчук)

2.2. Електронний документообіг у дистанційній освіті

При проведенні освітнього процесу у дистанційній формі важливу роль відіграє електронний документообіг, який надає низку переваг:

1. Швидкість і зручність доступу. Електронні документи можуть зберігатися в хмарних сховищах і бути доступними з будь-якого підключеного до мережі пристрою.
2. Автоматизація процесів. Використання електронних платформ дозволяє автоматизувати рутинні операції, такі як створення звітів, перевірка робіт та підготовка звітності.
3. Заощадження ресурсів. Зменшується використання паперу, скорочуються витрати на друк і зберігання документів.
4. Підвищення безпеки. Електронний документообіг дозволяє встановлювати рівні доступу до документів і використовувати цифровий підпис для підтвердження автентичності.

3. Автоматизація створення документації та інструменти для неї

При автоматизованому документообігу у освітньому процесі у дистанційній формі важливою є автоматизація створення, оформлення та аналізу документації, яка дозволяє значно зменшити навантаження на викладачів та адміністративний персонал.

3.1. Markdown як інструмент для створення електронних документів

Для створення електронних документів, зокрема звітів, постає питання щодо вибору формату для таких документів. В освітньому процесі для цієї задачі пропонується використовувати формат Markdown [2]. Цей формат на сьогодні використовується на Internet платформах і сервісах, зокрема у сервісі ШІ ChatGPT та на платформі GitHub.

Основними перевагами формату Markdown є:

1. Легке форматування тексту з допомогою будь-якого текстового редактора, при якому усі елементи, такі як заголовки, списки, таблиці, вставки зображень, створюються за допомогою простих текстових позначок.
2. Можливість конвертації документа у форматі Markdown в інші формати, наприклад у PDF, DOCX, ODT, HTML тощо.
3. Сумісність із системами управління версіями, наприклад із GitHub.

```
Приклад структури звіту в форматі Markdown:
# Звіт про виконання курсового проєкту

## Вступ
Текст вступу...

## Основна частина
### Аналіз предметної області
Текст підрозділу основної частини...

## Використані джерела
1. Джерело 1
2. Джерело 2
```

Позначки: #, ##, ###, використовуються тут для позначення назви документу, заголовків розділів та заголовків підрозділів. Відформатований у такий спосіб тест зберігає структуру документа і є читабельним. Остаточне оформлення документа визнається шаблоном стилів в окремому файлі.

3.2. Діаграми у форматі PlantUML

Для створення діаграм у звітах пропонується використовувати інструмент для генерації UML-діаграм з текстових описів PlantUML [3]. Це дає змогу швидко візуалізувати процеси, структури чи моделі без використання графічного редактора із текстового коду.

Приклад коду для створення діаграми у PlantUML:

```
@startuml
actor Студент
actor Викладач
participant "LMS" as lms

Студент -> lms : Надсилає завдання
lms -> Викладач : Повідомлення про нове завдання
Викладач -> lms : Перевірка завдання
lms -> Студент : Зворотній зв'язок
@enduml
```

Приклад автоматично створеного із наведеного текстового коду рисунка наведений на рисунку 1.

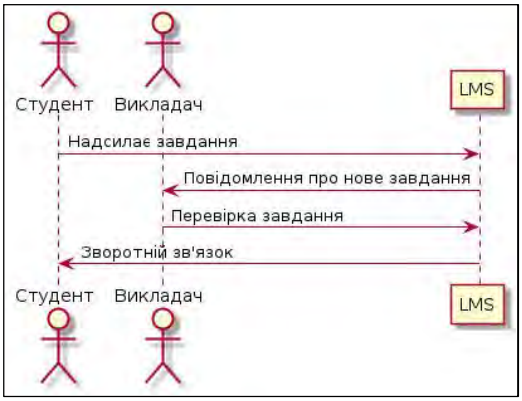


Рисунок 1: Діаграма, створена автоматично із текстового опису у форматі PlantUML

При потребі, в такий рисунок можна легко внести зміни у будь-якому простому текстовому редакторі. Крім того, код рисунка являє собою простий текст, який можна легко автоматично проаналізувати для перевірки та оцінювання.

Такий текстовий опис може знаходитися у тестовому документі у форматі Markdown.

3.3. Форматування звітів у LibreOffice Writer

Для забезпечення відповідності до вимог стандартів, важливо використовувати програмне забезпечення, яке підтримує функції автоматичного форматування. Одним із інструментів для створення документів, які відповідають нормативним вимогам є текстовий процесор LibreOffice Writer [4]. Для автоматизації оформлення звітів відповідно до встановлених правил у ньому використовуються шаблони стилів, які включають: стилі заголовків (Заголовок 1, Заголовок 2 тощо) для створення змісту, стиль основного тексту для забезпечення єдиного формату абзаців, стилі для таблиць, списків і приміток.

Для оформлення звітів був використаний шаблон з такими налаштуваннями: Розмір і тип шрифту. Встановлюються за вимогами (Times New Roman, 12 pt). Поля сторінки. Стандартні налаштування: верхнє і нижнє — 2 см, ліве — 3 см, праве — 1,5 см. Міжрядковий інтервал. Для тексту — 1,5, для заголовків — залежно від рівня (1,5 для заголовків другого рівня). Цей шаблон використовуються у команді pandoc при створенні із тексту в форматі Markdown оформленого відповідно до вимог стандартів документа.

Приклад виклику команди:

```
pandoc -f markdown -t docx —template шаблон.dotx звіт.md -o звіт.docx
```

4. Автоматизація створення та обробки звітної документації

Для створення оформлених відповідно до вимог нормативної документації звітів було використано такі програмні засоби: утиліта Pandoc — для конвертування Markdown документів у інші формати; утиліта PlantUML — для створення діаграм та технічних рисунків; текстовий процесор LibreOffice Writer — для остаточного оформлення звітних документів; шаблони стилів — для забезпечення відповідності звітів вимогам ДСТУ 3008:2015; командний інтерпретатор Bash — для автоматизації процесу обробки.

5. Можливості та результати впровадження

Для впровадження запропонованих способів створення документації у освітніх закладах потрібно:

1. Розробити універсальні шаблони для різних видів звітної документації.
2. Впровадити формати Markdown, PlantUML в документообіг закладу освіти.
3. Інтегрувати автоматизовані засоби створення звітів у системи управління навчанням .
4. Навчити викладачів та студентів використовувати ці засоби.

Описані способи та інструменти були впроваджені автором в освітній процес для підготовки фахівців зі спеціальності «Інженерія програмного забезпечення». Застосування форматів Markdown та PlantUML дозволило підвищити ефективність підготовки навчальних матеріалів, автоматизувати рутинні завдання та покращити взаємодію між викладачами та студентами, сприяло автоматизації освітнього процесу.

6. Висновки

Використання сучасних інструментів для підготовки звітної документації, зокрема Markdown, PlantUML та LibreOffice Writer, дозволяє покращити ефективність та стандартизувати процес оформлення документів. Це сприяє підвищенню якості освітнього процесу в дистанційній формі, забезпечує прозорість і зручність доступу до інформації для всіх учасників навчального процесу. Подальший розвиток автоматизованих систем обліку та аналізу такої документації сприятиме підвищенню якості освіти.

Перелік посилань

[1] ДСТУ 3008:2015. Інформація та документація. Звіти у сфері науки і техніки. — Київ: Держспоживстандарт України, 2015. — 28 с.
[2] Markdown Guide. URL: https://www.markdownguide.org
[3] PlantUML Documentation. URL: https://plantuml.com
[4] LibreOffice Writer Documentation. URL: https://documentation.libreoffice.org

Алгебро-логічні методи синтаксичного аналізу та інтеграції навчальних матеріалів у системах дистанційного навчання

Ігор Сотник[1*†], Ігор Шубін[1†] та Олексій Шапиро[1†]

[1] *Харківський національний університет радіоелектроніки, пр. Науки. 14, Харків, 61166, Україна*

Анотація

В роботі пропонується метод інтеграції різнорідних навчально-методичних матеріалів у єдиний курс навчання. На основі розробленого метода вирішується задача створення інструментальних систем підтримки інтеграції навчально-методичних матеріалів, до яких висуваються жорсткі вимоги, ефективність навчання істотно залежить від форми та якості надання навчальних матеріалів. В основу подібних систем закладено розподілену модель зберігання інформації.

Ключові слова

дистанційне навчання, автоматизовані навчальні системи, комп'ютерні програми навчального призначення, анотування текстів

1. Вступ

Розвиток комп'ютерних технологій і дистанційного навчання зумовив появу задач, пов'язаних з інтеграцією різнорідних навчальних матеріалів у єдиний навчальний курс. Це потребує створення гнучких інструментів, здатних забезпечити адаптивний навчальний процес та персоналізований підхід до кожного студента. У статті запропоновано модель синтаксичного аналізу для завдань анотування і реферування, яка базується на алгебрі кінцевих предикатів і покликана підвищити якість автоматизованого синтаксичного аналізу тексту.

2. Аналіз останніх досягнень

Розглядаються чотири класи автоматизованих навчальних систем (АНС): інформаційні, контролюючі, моделюючі та тренажери. Кожен з цих класів має свої особливості і переваги в залежності від специфіки навчання. Наприклад, моделюючі системи є особливо корисними для виконання лабораторних робіт, де важливо відслідковувати поведінку об’єкта у змінних умовах. Існуючі АНС дозволяють підтримувати навчальний процес, але недостатньо враховують індивідуальні особливості учнів, що є важливим фактором для підвищення ефективності навчання.

3. Постановка задачі і мета роботи

Сучасні тенденції в освіті, зокрема стрімкий розвиток дистанційного навчання, вимагають створення універсальних підходів до організації навчальних матеріалів. Ключова проблема полягає в інтеграції різнорідних навчальних ресурсів (текстових,

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
* Corresponding author.
† These authors contributed equally.
ihor.sotnyk@nure.ua (І. Сотник); igor.shubin@nure.ua (І. Шубін); oleksii.shapyro@nure.ua (О. Шапиро);
0009-0000-5638-6975 (І. Сотник); 0000-0002-1073-023X (І. Шубін); 0009-0005-3539-266X (О. Шапиро)

графічних, аудіовізуальних) у єдиний навчальний курс, що відповідає індивідуальним потребам студентів [1].

Складність полягає у тому, щоб забезпечити:

- Уніфікацію структури курсу, яка дозволяє легко адаптувати навчальний матеріал під різні предметні області та типи навчання.
- Гнучкість навчального процесу, де система автоматично підлаштовується під рівень знань студента та змінює стратегію навчання у реальному часі.
- Безпеку навчального середовища, яке запобігає втручанню студентів у процеси оцінювання та управління навчанням.

Існуючі автоматизовані навчальні системи мають певні обмеження. Вони або не враховують індивідуальних особливостей студентів, або зосереджуються лише на одному аспекті навчання, наприклад, на тестуванні. Тому актуальною є розробка універсальної інструментальної системи, яка інтегрує різноманітні методи навчання, забезпечуючи при цьому високу ефективність та персоналізацію [2].

4. Вирішення задачі

В рамках дослідження розроблено модель навчального курсу, представлену у вигляді орієнтованого графу. Вузлами графу є навчальні матеріали (слайди), а ребрами — можливі переходи між ними. Такий підхід дозволяє забезпечити гнучкість та адаптивність навчального процесу. У кожному вузлі містяться матеріали, які студент може переглянути, перш ніж перейти до наступного етапу навчання. Переходи між вузлами визначаються на основі поточного рівня знань студента, що дозволяє системі автоматично адаптувати навчальний процес.

4.1. Індивідуальна модель студента

Індивідуальна модель студента враховує його знання, психологічний тип та рівень засвоєння матеріалу. Ця модель формується на основі аналізу успіхів студента, що фіксуються у протоколі його дій. Наприклад, модель зберігає дані про виконані завдання, час, витрачений на кожен етап, та результативність тестування. Важливою особливістю цієї моделі є те, що вона дозволяє розробникам навчальних курсів підлаштовувати матеріали та стратегії навчання під індивідуальні особливості кожного студента.

4.2. Використання біноміальної моделі педагогічного тесту

Для контролю знань студента застосовується біноміальна модель педагогічного тесту, яка оцінює виконання завдань за бінарною шкалою: "виконано" або "не виконано". Цей підхід дозволяє точно визначити рівень засвоєння матеріалу та регулювати подальший прогрес студента. Кожен інтерактивний слайд в курсі має два варіанти переходу: один для успішного виконання завдання, інший — для невиконання. Це дозволяє адаптувати навчальний матеріал відповідно до прогресу студента.

4.3. Інтерактивні та інформаційні слайди

Система використовує два типи слайдів:

- Інформаційні слайди — містять навчальні матеріали, включаючи текст, графічні елементи, аудіо- та відеофрагменти. Вони надають інформацію для ознайомлення та запам'ятовування, що є важливою складовою для теоретичної частини курсу.

- Інтерактивні слайди — містять завдання, призначені для оцінки розуміння матеріалу. Вони забезпечують зворотний зв'язок з учнем та контроль за його прогресом. Інтерактивні слайди можуть включати різноманітні типи тестових завдань, що дозволяє не тільки перевірити рівень знань студента, але й допомогти йому засвоїти матеріал за рахунок активного залучення в процес.

4.4. Адаптивне управління навчальним процесом

Навчальний процес керується адаптивно за допомогою аналізу результатів проміжних тестів. Наприклад, якщо студент не виконав завдання, система пропонує додатковий матеріал або інші завдання для кращого розуміння теми. Водночас при успішному проходженні тестування студенту можуть пропонуватися більш складні завдання або можливість швидшого переходу до наступних розділів курсу. Система враховує особливості навчального стилю студента, обираючи для нього оптимальні стратегії освоєння матеріалу.

4.5. Застосування статистичних методів та автоматизація процесів

Запропонована модель також використовує статистичні методи для аналізу та обробки результатів тестування [3]. Це дозволяє автоматично адаптувати модель студента, коригуючи її на основі нових даних. Крім того, розробники курсу можуть вносити корективи у правила переходів між етапами навчання, не редагуючи програмний код, що полегшує оновлення курсу та спрощує підтримку системи.

5. Висновки

Запропонована інструментальна система для інтеграції навчальних матеріалів забезпечує гнучкість і персоналізацію дистанційного навчання. Використання графової структури, моделі студента та індивідуальних стратегій навчання дозволяє підвищити ефективність та результативність навчального процесу. Впровадження такої системи дає можливість швидкого оновлення матеріалів курсу та адаптації навчального процесу під конкретного студента, що є важливим аспектом у розвитку сучасних технологій дистанційного навчання.

Література

[1] Wilson, L., Thompson, D. (2021), “Integrating Annotation Tools in E-Learning Platforms: Enhancing Student Engagement”, Proceedings of the 10th International Conference on E-Learning and Education Technology, Vol. 8, pp. 65–78. DOI: 10.1016/j.elearning.2021.03.005

[2] King, B., Wright, T. (2021), “Enhancing E-Learning Experiences through Interactive Annotation Tools”, Proceedings of the 7th International Conference on Interactive Collaborative Learning, Vol. 5, pp. 12–24. DOI: 10.1016/j.icl.2021.03.002

[3] Smith, J., Johnson, E. (2020), “Automatic Text Summarization: A Comprehensive Review”, Proceedings of the 15th International Conference on Computational Linguistics, Vol. 12, pp. 45–57. DOI: 10.1016/j.cl.2020.01.004

Обґрунтування необіхдності застосування аналізу тепловізійних знімків у функціональній діагностиці систем організму

Костянтин Нечволод[1] та Ірина Кириченко[1]

[1] *Kharkiv National University of Radio Electronics, 14 Nauky Ave., Kharkiv, 61166, Ukraine*

Анотація
Розглянуто основні проблеми при роботі з пацієнтами, визначено необхідність розробки програмного забезпечення для мобільного телефону з метою покращення швидкості та якості прийнятих рішень лікаями.

Ключові слова
тепловізійні знімки, мобільний застосунок, медицина

1. Вступ

Складні умови, в яких опинилася Україна останніми роками, висувають до сучасної науки і практики нові вимоги, що покликані оптимізувати процес надання допомоги людям, травмованим війною, так і іншим категоріям громадян, які мають проблеми із соматичним здоров'ям. Постійні військові дії, обстріли, травмування громадян висувають перед фахівцями різноманітні виклики. У загальному суть цього виклику полягає в необхідності надавати швидко і якісно кваліфіковану і максимально ефективну медичну допомогу великій кількості громадян за короткий проміжок часу. Сьогодні, велика кількість травмованих осіб, потребує своєчасного та якісного лікування, яке не можливо реалізувати наявними в сучасній медицині та сфері охорони здоров'я засобами. Застосування звичних ресурсів у сфері охорони здоров'я (праця лікарів та допоміжного медичного персоналу, застосування класичних методів діагностики, тощо) не дає можливість побороти ті виклики, з якими стикаються медичні працівники та суспільство загалом. Зокрема, наукові дослідження свідчать, що велика кількість негативних наслідків несвоєчасного лікування (наприклад, ампутацій кінцівок внаслідок несвоєчасної чи помилкової діагностики у процесі лікування) є негативним фактором в умовах сьогодення, що накладає відбиток на ставлення громадян до сфери охорони здоров'я та підриває віру самих медичних працівників у свою спроможність надати кваліфіковану допомогу всіх, хто її потребує.

2. Обґрунтування необхідності розробки технології

З метою подолання окреслених вище викликів вважаємо за необхідне розглянути найбільш ефективний та продуктивний шлях – інтеграції методів діагностики соматичних показників пацієнтів із новітніми технологіями у сфері розробки мобільних застосунків. Це виступає новим кроком у рішенні проблеми надання своєчасної допомоги людям, які її потребують. Особливо коли мова йде про терапевтичний супровід осіб із

* Corresponding author.
† These authors contributed equally.
kostiantyn.nechvolod@nure.ua (K. Nechvolod); iryna.kyrychenko@nure.ua (I. Kyrychenko);
0009-0000-5879-6644 (K. Nechvolod); 0000-0002-7686-6439 (I. Kyrychenko);

втратою кінцівок, суттєвим травмуванням кінцівок, суттєвою крововтратою та іншими порушеннями соматичного здоров'я – пораненнями, травмами, опіками, тощо.

Ідея залучення гаджетів та мобільних пристроїв до процесу здійснення функціональної діагностики різноманітних систем організму представлена в ряді сучасних наукових публікацій. Так, наукові пошуки таких авторів як А.Ю.Бойко [1], Є.О.Єрошенко, І.В.Прасол [2], Г.А.Корнієнко [3], І.В.Сімакова [4] присвячені різним аспектам застосування технологій для оптимізації процесу медичної діагностики чи контролю параметрів фізичної активності людини. Разом із тим, наявні наукові розробки в цій сфері, здебільшого, стосуються полегшення процесу обробки зображень, їх класифікації чи зчитування з уже наявних і зроблених іншими засобами досліджень. Разом із тим, застосування мобільних додатків для діагностики параметрів тканин як окремого інструменту залишається поза увагою дослідників.

Інтеграція мобільних додатків у медичну діагностику може значно підвищити ефективність надання допомоги пацієнтам з травмами та іншими соматичними порушеннями. Зокрема, використання мобільних пристроїв для функціональної діагностики тканин організму дозволяє швидко та без додаткових ресурсів оцінити стан уражених ділянок, що є критично важливим у польових умовах або при масових надзвичайних ситуаціях.

Дослідження ефективності використання мобільних додатків у медичній практиці підтверджують їхню доцільність. Зокрема, у роботі [5] розглядаються сучасні прилади та системи для лікування цукрового діабету I типу, включаючи мобільні додатки, які сприяють покращенню контролю за станом пацієнтів. Інше дослідження аналізує використання мобільних додатків для виявлення фахових здібностей абітурієнтів з урахуванням їхнього емоційного стану, що підкреслює широкі можливості застосування таких технологій у різних сферах.

Сьогодні мобільні технології активно змінюють підхід до діагностики, лікування та управління станом здоров'я пацієнтів. Розвиток мобільних додатків у сфері медицини охоплює численні напрямки: від простих програм для моніторингу стану здоров'я до складних рішень для телемедицини.

Мобільні додатки вже стали інструментом для дистанційного консультування, запису на прийом до лікаря та управління особистими медичними даними пацієнтів. Завдяки таким додаткам користувачі можуть отримати доступ до своїх результатів аналізів, історії хвороб і нагадувань про прийом ліків, що робить лікування більш персоналізованим і зручним

У якості вирішення суттєвого ряду проблем із надання медичної допомоги особам, які мають травми і проходять лікування, ми вбачаємо розробку та впровадження мобільного застосунку , метою якого постає різноманітний аналіз системи функціонування тканин організму в зоні ураження, диференціація тканин за ступенем порушеності у відповідності до критерію можливості відновлення чи відмирання, формування імовірнісного прогнозу подальшого лікування певної системи організму людини, тощо. Функціональність застосунку забезпечується використанням комплексу діагностичних параметрів, доступних у гаджеті (зокрема, інфрачервоного випромінювання, температурних реакцій, тощо). Використання цього застосунку дає можливість своєчасно та швидко, без зайвих технологічних навантажень провести первинну діагностику стану тканин організму з метою вирішення стратегії подальшого лікування.

Необхідність застосування новітніх розробок у сфері мобільних застосунків в забезпеченні лікувального процесу (конкретно, застосунку) детермінована рядом чинників сьогодення.

По-перше, в умовах відкритих військових дій травми одночасно отримує велика кількість людей, що потребують своєчасної допомоги. Зважаючи на обмеженість людських ресурсів, медичні працівники не завжди можуть своєчасно провести всі

необхідні технологічні дослідження стану тканин перед вирішенням стратегії подальшого лікування чи ампутації, а часто обмежуються візуальним оглядом, що часто може не відображати дійсні характеристик системи тканин організму і призвести до небажаної ампутації.

По-друге, вагомим чинником запровадження таких технологій виступає обмеженість засобів медичної діагностики (рентгенографії, комп'ютерної томографії, тощо). Апарати, що використовуються для діагностики, є достатньо об'ємними і складно піддаються транспортуванню. По суті, процес мобільної медичної діагностики здійснити дуже складно, і необхідно транспортувати людину із пораненнями чи травмами до медичної лабораторії, що потребує достатньо великих затрат ресурсів та часу. Використання мобільних технологій дозволяє вирішити питання втрати часу та не потребує доставки пацієнта до лікувального закладу для проведення діагностики. При цьому максимально зменшується час, необхідний на проведення дослідження.

По-третє, вагомою проблемною медичної діагностики є низька доступність апаратури. Апаратів, які використовуються для діагностик тканин та систем організму людини, обмежена кількість, зважаючи на їх високу вартість та значну кількість умов їх обслуговування, складність експлуатації. Натомість, використання мобільних застосунків для діагностики систем організму є новим вирішенням, оскільки вони характеризуються вираженою доступністю. Їх використання не передбачає ані необхідності обладнувати спеціальні приміщення у відповідності до вимог охорони праці, ані потребує спеціального транспорту для перевезення. Їхнє використання можливе за умови наявності гаджет, а процес діагностики може бути проведений де завгодно і швидко.

Тож, використання мобільного застосунку для функціональної діагностики систем організму дозволяє подолати велику кількість викликів сьогодення.

Застосунок, покликаний проводити діагностику систем організму, має ряд переваг, порівняно з іншими, класичними засобами діагностики. По-перше, мова йде про його доступність. По суті, використання цієї технології є доступним для усіх осіб, які мають мобільний телефон. Це не передбачає необхідність використання ємних діагностичних процедур у спеціально обладнаних лабораторіях. По-друге, перевагою застосунку є функціональність, тобто, різноманітна спрямованість процесів діагностики та його можливість проводити діагностику. Використовуючи доступні через мобільний пристрій діагностичні параметри, застосунок спроможний здійснити швидку і різноманітну діагностику стану систем організму (зокрема, шкіри, м'язової та кісткової тканин, тощо).

По-третє, перевагою застосунку є простота використання, що досягається зручним та зрозумілим інтерфейсом програми, чіткістю параметрів виводу результатів діагностики. До речі, процедура діагностики та інтерпретації отриманих результатів є доступною навіть для осіб із початковими знаннями медицини. Це, знову ж, свідчить про доступність застосунку [6]. З точки зору лікарів, мобільні додатки полегшують роботу завдяки інтеграції з медичними калькуляторами, довідниками з фармакології та календарями прийомів пацієнтів[7]. Це дозволяє не лише покращити обслуговування пацієнтів, але й оптимізувати внутрішні процеси у медичних закладах

По-четверте, перевагою застосунку виступає здоров'язбережувальна орієнтованість, оскільки процедура діагностики передбачає використання безпечних для організму технологій (інфрачервоного випромінювання), що не несуть шкоди здоров'ю, як наприклад, процедура рентгену чи томографії. Огляд наукових досліджень свідчить, що біосенсори та інші інноваційні пристрої, інтегровані з мобільними технологіями, дають змогу ефективніше відстежувати показники здоров'я, як-от рівень глюкози чи тиск, та миттєво передавати дані лікарю

3. Висновки

Усе це призводить до затримок у діагностиці, можливих медичних помилок, таких як невиправдані ампутації або ускладнення лікування, що ускладнює якісне надання медичної допомоги, підриває довіру до системи охорони здоров'я та негативно впливає на здоров'я пацієнтів. Вирішення цієї проблеми полягає у розробці та впровадженні мобільного додатка для функціональної діагностики тканин організму, що забезпечить доступність, швидкість та ефективність діагностичних процедур без залежності від стаціонарного медичного обладнання. Отже, застосування мобільного застосунку відповідає сучасним практичним викликам до медицини та сфери технологій і має ряд суттєвих переваг, порівняно із типовими засобами функціональної діагностики організму. При чому, застосунок характеризується рядом переваг, до яких ми відносимо доступність, функціональність, простота використання та спрямованість на здоров'язбереження. Тож, розробка та використання такого застосунку є вимогою сучасності і дозволить максимально забезпечити інтеграцію технологій до сфери медичної діагностики для надання швидкої і ефективної допомоги пацієнтам.

Разом із тим, дана публікація не вичерпує всіх аспектів питання, що розглядається. Зокрема, перспективою подальшого дослідження автора виступатиме власне розробка та практична апробація застосунку, оптимізація його функціонального наповнення, припрацювання зручності інтерфейсу та актуального в часі обміну даними.

Література

[1] Бойко А. Ю. Розробка застосунку для вимірювання показників фізичної активності спортсменів : кваліфікаційна магістерська робота : спец. 051 "Економіка" / наук. кер. В. А. Вишневська ; Центральноукраїн. нац. техн. ун-т. Кропивницький : ЦНТУ, 2023. 156 с.

[2] Єрошенко О. А., Прасол І.В. Метод обробки електроміографічних сигналів для побудови системи електростимуляції. II Міжнародна науково-практична конференція «Інформаційні системи та технології в медицині» (ІСМ–2019): зб. наук. пр. Харків: Нац. аерокосм. ун-т ім. М. Є. Жуковського «Харків. авіац. ін-т», 2019. С. 186-188.

[3] Корнієнко Г. А. Аналіз медичних зображень для виявлення патології : магістерська дис. : 122 Комп'ютерні науки. Київ, 2024. 127 с.

[4] Сімакова І. В. Програмно-апаратний комплекс моніторингу та аналізу фізіологічних показників стану людини на базі модулю ESP32 : кваліфікаційна робота на здобуття освітнього ступеня «магістр» : спец. 123 «Комп'ютерна інженерія»; ЧНУ ім. Петра Могили. Миколаїв, 2023. 75 с.

[5] Мобільні додатки і системи для діагностики і лікування цукрового діабету I типу. Вісник Вінницького національного технічного університету, Вінниця, 2023. URL: https://ir.lib.vntu.edu.ua/handle/123456789/21936.

[6] Кириченко І.В., Залазаєв А.М. Порівняння нативної розробки мобільних додатків з уніфікованою розробкою за допомогою REACT NATIVE. XII міжнародна науково-технічна конференція "Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління", 27-28 квітня 2022, Баку – Харків – Жиліна – 2022, Том 2: секція 5, с. 190

[7] O. Cherednichenko, I. Kyrychenko, K. Smelyakov, O. Dolhanenko. Data Security Analysis in EMM Systems. 8th International Conference on Computational Linguistics and Intelligent Systems. Volume II: Modeling, Optimization, and Controlling in Information and Technology Systems Workshop, MOCITSW-CoLInS 2024

Секція 6.
Програмна інженерія

Комп’ютерна симуляція небесної механіки з навчальними цілями

Володимир Бондарєв[1†] та Юлія Черепанова[2*†]

[1] *Харківський національний університет радіоелектроніки, пр.. Науки, 14, Харків, 61166, Україна*

Анотація

В роботі пропонується інтерактивна модель небесної механіки з можливістю як вирішувати, так і створювати навчальні завдання. Модель реалізована програмною системою, якою можна користатися самостійно або з викладачем. Від аналогів роботу відрізняє наявність навчальних завдань, які можна змінювати і поповнювати.

Ключові слова

імітаційна модель, небесна механіка, навчальні завдання

1. Вступ

Вивчення фізики захоплююче заняття, особливо якщо учень сам може ставити фізичні експерименті. Однак далеко не всі розділи підручника можна підкріпити практичною діяльністю. Дуже малі і дуже великі об’єкти можливо тільки уявляти, їх не можна встановлювати, рухати, і навіть, спостерігати. Тут на допомогу викладачу приходить моделювання, а саме комп’ютерні імітаційні моделі, які, завдяки тотальній комп’ютеризації нашого життя, доступні майже кожному.

Учбові моделюючі програми насамперед повинні правдиво відображувати реальні явища, але до того мати простий користувацький інтерфейс, бути інтерактивними і швидко реагувати на дії користувача. Задоволення усіх перерахованих потреб іноді вимагає неабияких зусиль і винахідливості.

Гарним прикладом подібної програми є [1], але модель небесної механіки там носить скорше якісний, а не кількісний характер. Існують вражаючі імітатори, які мають характер планетарію, наприклад, [2,3]. Вони збуджують цікавість, але не навчають фізичним законам, які керують рухом зірок і планет. Існують програмні пакети, які дозволяють чисельно вирішувати рівняння руху і обчислювати орбіти з високою точністю [4], вони можуть бути корисними при створенні комп’ютерних імітацій, але самі такими не є.

В роботі пропонується інтерактивна модель небесної механіки з можливістю як вирішувати, так і створювати навчальні завдання. Модель реалізована програмною системою [5], якою можна користатися самостійно або з викладачем.

2. Модель

Модель імітує поведінку масивних тіл, які рухаються у просторі, підкоряючись закону тяжіння Ньютона [6]. Простір, в якому рухаються тіла, тривимірний. Саме в тривимірному просторі напруга поля гравітації зворотно пропорційна квадрату відстані від точкової

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
* Corresponding author.
† These authors contributed equally.
volodymyr.bondariev@nure.ua (В. Бондарєв); yulia.cherepanova@nure.ua (Ю.Черепанова)
0000-0001-5410-8576 (В. Бондарєв); 0000-0002-9441-6300 (Ю.Черепанова)

маси, у двовимірному просторі залежність напруги і закон тяжіння мали б інший вигляд. Але модельні сценарії побудовані так, що всі вектори положення і вектори швидкості рухомих тіл розташовані в одній площині, тому стан моделі природно відображується на площину екрану.

В кожному модельному сценарії чисельно вирішується задача n тіл. Одночасно з вирішенням тіла і орбіти відображуються на екрані, що створює ілюзію руху тіл.

Модельний час, на відміну від природного, дискретний. При виконанні обчислень одиницею виміру часу вважається один такт, одиницею виміру простору – один піксель, одиниця виміру маси обрана такою, щоб стала тяжіння в законі Ньютона дорівнювала 1. При відображенні моделі одиницям вимірювання можна дати інші назви, наприклад, один такт часу це один день, один піксель це мільйон кілометрів, одиниця маси – маса Землі. Це нічого не змінить у фізичних законах, окрім констант, але ускладнить обчислення для учнів. Тому одиницям вимірювання назв при їх відображенні не надається зовсім.

3. Складові моделі

Планета – основний елемент моделі. Він уособлює не тільки власне планети, а і зірки, астероїди, комети і навіть ракети, які є об'єктами, похідними від планети.

Основні властивості планети це маса, розмір, положення і швидкість. З двовимірності модельних сцен витікає, що положення і швидкість задаються двовимірними векторами. Планета має форму кола, тому розмір планети задається радіусом кола. Для ідентифікації планета слугує її унікальне ім'я. Для відображення планети слугують додаткові властивості – колір і траєкторія. Траєкторія зберігається як колекція точок простору, в яких планета побувала в попередні такти часу.

Планети взаємодіють між собою через силу тяжіння, яка визначається за законом Ньютона. На малих відстанях від центру напруженість поля тяжіння планети стає надто великою, і якщо інші тіла наближаються до центру планети на таку або меншу відстань, обчислювальна схема працює хибно.

Критичну відстань можна визначити з формули для напруженості $E = Gm/r^2$ [7].

$$r_к = \sqrt{Gm/E_к} \quad (1)$$

Запобігти наближенню на критичну відстань може розмір самої планети, тому що, коли тіла стикаються, тобто відстань між центами тіл стає менше за суму їх радіусів, більш масивна планета поглинає планету з меншою масою. Встановлено, що $E_к$ моделі приблизно дорівнює 4, тому радіус планети має бути більшим за $\sqrt{Gm}/2$

Ракети демонструють, як можна пересуватися в космічному просторі і здійснювати міжпланетні подорожі. Ракета є космічним тілом малої маси і розміру. Ракета стартує з обраної планети, отримує миттєвий імпульс під час старту і далі рухається по балістичній траєкторії без можливості її корекції. Відносно ракет діють два припущення: 1) траєкторія ракети починається з центру материнської планети, 2) тяжіння материнської планети не впливає на ракету. Такі припущення суттєво спрощують розрахунки і роблять їх доступними навіть школярам. Як і інші небесні тіла, ракети можуть стикатися з планетами, і це стає закінченням їх життєвого шляху.

Туманності є третьою складовою моделі, і хоча передбачення їх руху не така проста річ, як розрахунки орбіт планет і траєкторій ракет, повчальним є спостереження за їх еволюцією і взаємодією з іншими тілами.

Туманність моделюється сукупністю великої кількості однакових часток малої маси. Внаслідок взаємного тяжіння частки прагнуть злитися в одне і запобігти надто швидкому злиттю можуть тепловий рух і/або відцентрова сила, якщо туманність обертається. Моделювання теплового руху потребує додаткових обчислень, що може зменшити

швидкодію рушія до неприпустимо малої, тому для стабілізації туманності використане лише обертання.

4. Конструктор сцен

Конструктор сцен являє собою веб-сторінку, головним елементом якої є канвас, що відображує космічний простір (рис.1). В тому просторі користувач будує бажану сцену, тобто створює зірки і планети, надає їм бажані параметри, такі як маса, розмір положення, початкова швидкість тощо.

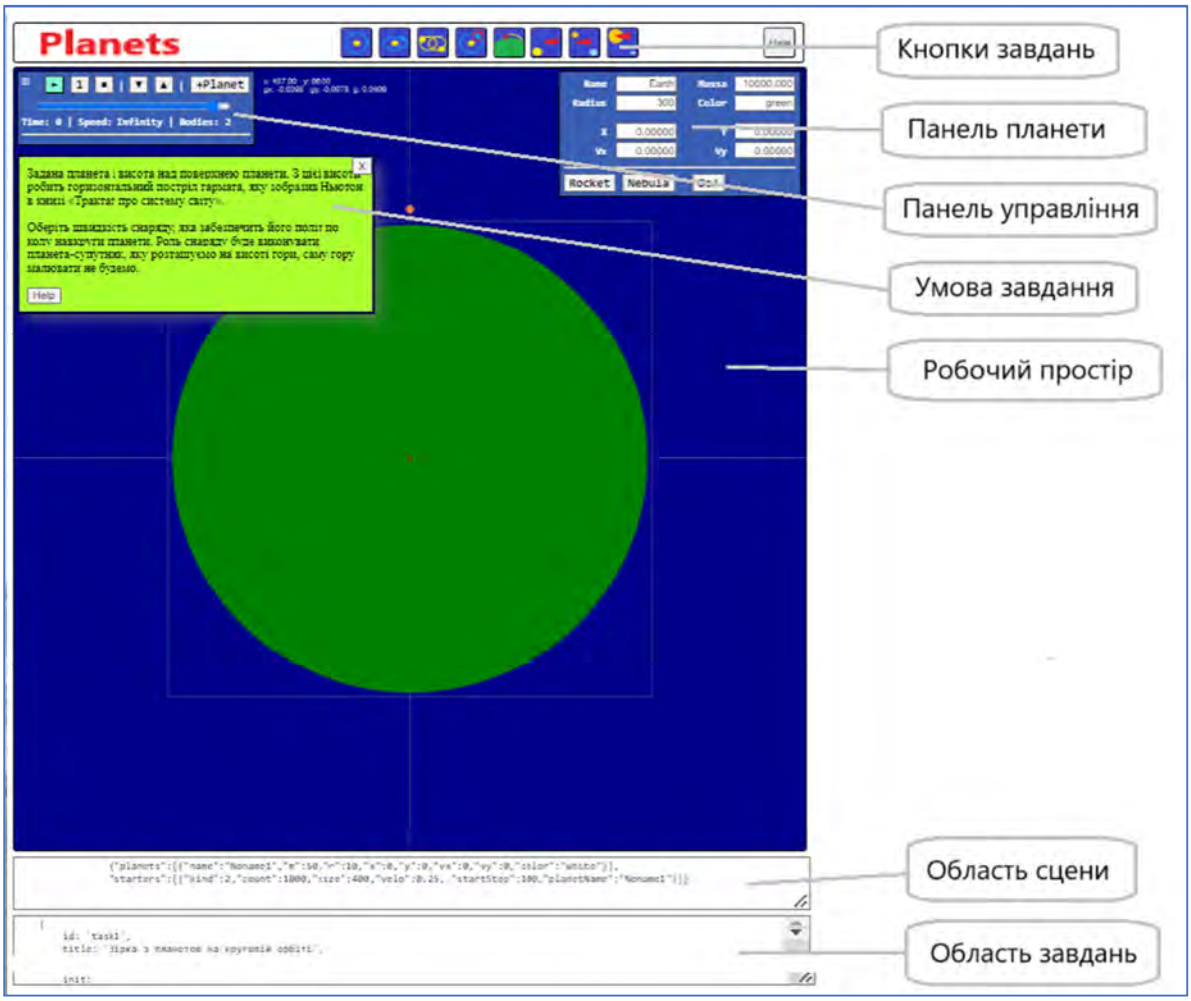


Рисунок 1: Конструктор сцен.

Під час створення сцени вона статична, в ній ніщо не рухається. Коли сцена створена, можна увімкнути плин часу, і всі елементи сцени почнуть рухатися відповідно до законів класичної механіки. В будь-який момент модельний час можна зупинити або продовжити. Можна також просуватись в часі покроково, щоб спостерігати малі зміни в стані моделі.

Сцену можна зберегти у вигляді тексту, щоб потім знову завантажити у простір, коли в тому виникне потреба. Орбіти планет і ракет можна показати або приховати. Масштаб зображення можна змінювати в широких межах. Все перелічене здійснюється за допомогою панелі керування в лівому верхньому куті робочого поля. Панель можна приховати, якщо вона заважає спостереженням.

Будь-який елемент сцени можна зробити обраним. Обраний елемент підсвічується, а в правому верхньому куті робочого поля з’являється панель, на якій можна бачити і змінювати всі параметри обраного елемента.

З обранням елементу з’являється можливість дій, що пов’язані з певною планетою. До них відносяться запуск ракет і перетворення планети на туманність. Такі дії можуть бути відкладені у часі, тобто ракета або туманність виникне не одразу, а через заплановану кількість тактів модельного часу.

5. Учбові завдання

Учбове завдання полягає в тому, що користувач отримує певну сцену і повинен так змінити її, щоб вона задовольняла вимогам, викладеним в завданні. Наприклад, в завданні надається сцена, в якій є масивна зірка і планета на певній відстані від неї. Маса зірки набагато більше за масу планети. Потрібно надати планеті таку початкову швидкість, яка б змусила її обертатися навколо зірки по круговій орбіті.

Всі дані, необхідні для вирішення, такі як координати, швидкості, маси, розміри, користувач отримує з початкової сцени. З тими даними він робить розрахунок швидкості планети, виправляє сцену і бачить результат своїх зусиль, запустивши модельний час. Якщо результат не відповідає вимогам – планета рухається по еліпсу, або впала на зірку, або зовсім відлетіла в відкритий космос, користувач може виправити свої розрахунки і спробувати знову. Якщо досягти мети не вдається, користувач може отримати підказку, як провести обчислення. За бажанням він може отримати остаточне вирішення завдання разом зі сценою, яка точно відповідає вимогам завдання.

Певна низка учбових завдань вже знаходиться в системі. Ці завдання активізуються кнопками, які розташовані над робочим полем.

Опис завдання має текстову форму і повністю відокремлений від програмного коду. Завдяки тому, викладач може створювати власні завдання і додавати їх до тих, що вже є. Питання про те, чи буде підказка і чи буде доступна правильна відповідь, вирішує викладач.

6. Висновки

Імітаційні моделі роблять наочними такі речі, які можна лише уявляти, і тому їх треба застосовувати у навчанні якомога ширше. Хоча вже існують багато імітаційних програм з фізики і зокрема з небесної механіки, запропонована програма [5] поєднує достатньо точне моделювання з учбовими завданнями, які користувачі можуть не тільки вирішувати, а і створювати. Програма має мінімалістичний користувацький інтерфейс, що позбавляє учнів необхідності вивчати що-небудь, окрім фізики.

Література

[1] My Solar System. URL: https://phet.colorado.edu/en/simulations/my-solar-system
[2] Stellarium. URL: https://stellarium.org/
[3] Celestia — real-time 3D visualization of space. URL: https://celestiaproject.space/
[4] REBOUND. URL: https://rebound.readthedocs.io/en/latest/
[5] Planets. URL: https://tss.co.ua/planets/
[6] The Mathematical Principles of Natural Philosophy, Encyclopædia Britannica, London, archived from the original on 2 May 2015, retrieved 13 February 2015.
[7] Feynman, Richard P.; Leighton, Robert B.; Sands, Matthew (2005) [1970]. The Feynman Lectures on Physics: The Definitive and Extended Edition (2nd ed.). Addison Wesley. ISBN 0-8053-9045-6.

Концептуальне моделювання та формалізація ієрархічних знань для систем підтримки прийняття рішень

Заговора Аліна[1] та Шубін Ігор[1]

[1] *Харківський національний університет радіоелектроніки, пр. Науки, 14, м. Харків, 61166, Україна*

Анотація

Стаття присвячена концептуалізації та формалізації знань для систем підтримки прийняття рішень (СПР). Розглянуто ієрархічну організацію знань, що забезпечує структуроване представлення складних предметних областей. Використання категорної теорії дозволяє моделювати відносини між концептами для створення адаптивних і гнучких баз знань. Результати підтверджують ефективність запропонованого підходу для оптимізації роботи СПР у динамічних умовах.

Ключові слова

Системи підтримки прийняття рішень, концептуалізація знань, ієрархічна організація, категорна теорія, формалізація моделей, адаптивні бази знань.

1. Вступ

Концептуалізація та формалізація знань у системах підтримки прийняття рішень (СПР) є важливими компонентами для забезпечення ефективної роботи таких систем. У складних предметних областях концептуальне моделювання дозволяє відобразити основні елементи та взаємозв'язки між ними, забезпечуючи точне представлення знань для підтримки рішень у нестандартних ситуаціях.

Формалізація знань через математичні моделі дає змогу побудувати систему, яка накопичує, адаптує й узгоджує знання в межах ієрархічної структури.

Метою дослідження є розгляд концептуального підходу до побудови моделей знань для СПР та застосування формалізації ієрархічних відносин для підвищення адаптивності таких систем у динамічних предметних областях [1, 2].

2. Концептуалізація та модель знань

Концептуалізація предметної області є важливим етапом у побудові ефективних систем підтримки прийняття рішень (СПР). Вона передбачає створення моделей, що відображають знання у вигляді концептів та їхніх взаємозв'язків, максимально наближених до того, як людина сприймає та організовує інформацію. В основі концептуалізації лежить ідея, що знання про об'єкти та явища реального світу можна представити у формі понять і семантичних відношень між ними [3, 4].

Для цього визначаються базові концепти, які використовуються для моделювання знань про предметну область. Ці концепти організовані у структуру, що дозволяє задавати зв'язки між ними через семантичні відношення. Наприклад, між поняттями можуть існувати відношення типу [5] «частина-ціле», «рівень абстракції», «родо-видова залежність», тощо. Такий підхід дозволяє моделювати предметну область як мережу пов'язаних понять, де кожен елемент має своє місце та значення у загальній картині знань [6, 7].

alina.zahovora@nure.ua (A. Zahovora); igor.shubin@nure.ua (I.Shubin);
0000-0002-1073-023X (I.Shubin);

Важливим аспектом концептуалізації є визначення сигнатури моделі, яка задає імена концептів і відношень між ними. Це дозволяє структурувати знання та однозначно ідентифікувати всі елементи у базі знань. Формалізація семантичних відношень також має важливе значення, оскільки дозволяє представити їх у декларативній формі, що робить модель зручною для використання та аналізу у СПР. Наприклад, кожному відношенню між концептами можна приписати значення істинності, що дозволяє виконувати логічні операції над знаннями.

Побудова концептуальних моделей баз знань також враховує ієрархічну структуру понять, що забезпечує можливість переходу від загального до конкретного рівня опису. Це дозволяє створювати узагальнені та спеціалізовані моделі для різних аспектів предметної області, що може суттєво покращити ефективність роботи СПР. Ієрархічна організація знань, підтримувана концептуальною моделлю, дає змогу СПР надавати ЛПР чітко структуровану інформацію та обирати релевантні рівні абстракції залежно від завдання.

Отже, концептуалізація знань є основою для створення ефективної СПР, де знання структуровані, взаємозалежні та доступні для аналізу й прийняття обґрунтованих рішень.

3. Формалізація ієрархічних відносин у базах знань

Ієрархічна організація знань є ключовим аспектом при побудові систем підтримки прийняття рішень (СПР), оскільки дозволяє структурувати знання таким чином, щоб забезпечити швидкий і зручний доступ до необхідної інформації [8, 9]. Формалізація ієрархічних відносин у базах знань забезпечує можливість представлення складних систем знань як впорядкованих структур, де концепти та відношення між ними відображають взаємозалежності предметної області.

Застосування категорної теорії для моделювання ієрархічних відносин дозволяє описати взаємозв’язки між концептами у формі математичних моделей, які можна формально інтерпретувати та аналізувати. У такій моделі об’єкти предметної області представляються як концепти, а зв’язки між ними — як морфізми, які описують різні типи відношень, такі як ін’єктивність, сюр’єктивність або біморфізм. Ієрархічні відносини, побудовані за таким принципом, дозволяють СПР відображати як загальні узагальнення, так і конкретні деталі.

Одним із основних принципів формалізації ієрархії є побудова відношення «is-a», де кожний концепт може бути пов’язаний з іншими концептами у вигляді ієрархії «рід-вид». Це відношення відображає зв'язок між загальними та спеціалізованими моделями, де узагальнені концепти дозволяють зробити модель більш універсальною, а спеціалізовані — деталізувати знання в конкретній області. В категорній моделі це відношення визначається як сюр’єктивний морфізм, що дозволяє пов’язувати прикладні моделі концептуалізації з їх узагальненими версіями.

Ієрархічна структура також включає відношення між різними моделями, де кожна модель концептуалізації предметної області представляє певний фрагмент знань. Взаємозв’язки між такими моделями можуть включати агрегацію, коли кілька моделей об’єднуються для представлення комплексної системи знань, або узагальнення, коли створюється більш абстрактний опис на основі конкретних моделей. Наприклад, узагальнення дозволяє створювати абстрактні моделі для розв’язання типових завдань, тоді як детальні моделі забезпечують підтримку конкретних рішень у СПР.

Цей підхід до побудови ієрархії у СПР дозволяє системі ефективно управляти складними даними, забезпечуючи адаптивність і гнучкість баз знань. Завдяки чітко структурованій ієрархії знань, система може пропонувати оптимальні рішення для користувача, враховуючи як загальні закономірності, так і специфічні умови завдання.

Формалізація ієрархічних відносин у категорній моделі є важливим елементом, що сприяє розвитку адаптивних СПР, де знання є легко доступними, структурованими та готовими до використання у процесі прийняття рішень.

4. Висновки

Створення ефективних систем підтримки прийняття рішень (СПР) потребує чіткого структурування знань і формалізації ієрархічних відносин. Ієрархічна організація дозволяє системам працювати зі складними даними, зберігаючи баланс між узагальненням та деталізацією. Використання математичних моделей, зокрема категорної теорії, сприяє покращенню структурування баз знань і підвищенню їхньої адаптивності. Такий підхід забезпечує СПР можливістю гнучко реагувати на зміни та ефективно підтримувати прийняття рішень у різноманітних умовах.

Література

[1] Левикін В.М., Неофітна Т.М. Категорне моделювання предметних областей на основі знань у системах прийняття рішень // Науковий вісник Кременчуцького університету економіки, інформаційних технологій та управління. Нові технології. – 2006. – № 4 (14). – С. 21–25

[2] Левикін В.М., Неофітна Т.М. Формалізований опис онтології та концептуальної моделі предметної області // Зб. наукових праць за матеріалами 9-ї Міжнародної конференції «Теорія та техніка передачі, прийому й обробки інформації». – Харків. – 2003. – С. 383–384.

[3] Неофітна Т.М. Використання механізму успадкування в системах штучного інтелекту // Зб. наукових праць за матеріалами 7-ї Міжнародної конференції «Теорія та техніка передачі, прийому й обробки інформації». – Харків. – 2001. – С. 394–395.

[4] Годинський Е.Г. Концепція створення та використання дискретних ситуаційних моделей в управлінні динамічними системами // Інформаційні технології. – 2001. – № 9. – С. 37–43.

[5] UML. Основи, 3-тє видання. – Пер. з англ. – СПб: Символ-Плюс, 2006. – 192 с.

[6] J. Smith, S. Deloach, M. Kokar, K. Baslawski. Category theoretic approaches of representing precise UML Semantics // Presise UML Workshop at ECOOP 2000, Sophia Antipolis, France, June 2000.

[7] Неофітна Т.М. Розробка категорної моделі бази знань інтелектуальної системи прийняття рішень // Проблеми біоніки. – Харків. – 2003. – № 58. – С. 108–115.

[8] Левикін В.М., Неофітна Т.М. Розробка правил поповнення ієрархічної бази знань // Науково-технічний журнал «Біоніка інтелекту». – 2006. – № 1(64). – С. 68–71.

[9] Вагін В.М. та ін. Достовірний і правдоподібний висновок у інтелектуальних системах. – М.: ФІЗМАТЛІТ, 2004. – 704 с.

Адаптивні бази знань для систем підтримки прийняття рішень: підходи та моделі

Лихова Алєся[1] та Шубін Ігор[1]

[1] *Харківський національний університет радіоелектроніки, пр. Науки, 14, м. Харків, 61166, Україна*

Анотація

У статті розглянуто актуальні підходи до проєктування баз знань для систем підтримки прийняття рішень (СППР). Особлива увага зосереджена на онтологічному та проблемно-орієнтованому аналізі предметних областей та структури знань. У роботі пропоновано комплексний підхід до об'єднання онтологічних моделей та прикладних моделей концептуалізації, що дозволяє забезпечити адаптивність та гнучкість системи підтримки прийняття рішень

Ключові слова

Бази знань, системи підтримки прийняття рішень, адаптивність, онтологічний аналіз, моделі концептуалізації, проблемно-орієнтований підхід.

1. Вступ

Системи підтримки прийняття рішень (СПР) набули значної актуальності у сучасних дослідженнях через їхню здатність оптимізувати та підтримувати складні процеси управління. На відміну від традиційних інформаційних систем, СПР не тільки обробляють та зберігають дані, але й забезпечують аналітичну підтримку особам, які приймають рішення (ЛПР), особливо у контексті складних конфліктних ситуацій. СПР дозволяють ЛПР краще зрозуміти контекст, оцінити альтернативи та прийняти обґрунтоване рішення, що є вкрай важливим у швидко змінюваних і невизначених умовах [1].

Актуальність теми підтверджується зростаючим використанням СПР у різноманітних галузях, від медицини до фінансів, де необхідна підтримка в умовах багатофакторного аналізу і високої складності задач. Ця робота зосереджена на аналізі підходів до проектування баз знань для СПР, а також на теоретичних основах структурного представлення знань у складних предметних областях.

Метою дослідження є огляд сучасних методів проектування баз знань, адаптованих до потреб СПР, а також виявлення теоретико-категорних моделей, що забезпечують багатоаспектне представлення ієрархічних систем знань [2].

2. Аналіз джерел та завдання дослідження

Основним завданням систем підтримки прийняття рішень (СПР) є забезпечення допомоги особам, які приймають рішення (ЛПР), у ситуаціях, що потребують швидкої реакції та обґрунтованих рішень. Процес прийняття рішення складається з кількох ключових етапів: постановки мети, визначення альтернатив, оцінки можливих варіантів, вибору найкращого з них та реалізації прийнятого рішення. Кожен із цих етапів вимагає точного представлення знань та оцінки потенційних наслідків. Особливу увагу у СПР приділяють формуванню та впорядкуванню альтернатив, що дозволяє ЛПР обирати

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
aliesia.lykhova@nure.ua (A. Lykhova); igor.shubin@nure.ua (I.Shubin);
0000-0002-1073-023X (I.Shubin);

оптимальний шлях з урахуванням наявних критеріїв і обмежень. Водночас СПР повинна підтримувати процес генерації нових альтернатив, що є особливо важливим для прийняття рішень у нестандартних або швидко змінюваних умовах [3].

Важливим аспектом є вирішення проблем «пошуку» та «виведення» при опрацюванні альтернатив. Проблема пошуку пов'язана з визначенням усіх можливих рішень, тоді як проблема виведення зосереджена на обґрунтуванні та формулюванні висновків на основі доступної інформації. Для того щоб забезпечити ефективну підтримку, СПР повинні враховувати також особливості когнітивних процесів ЛПР.

Таким чином, актуальними завданнями дослідження є вивчення підходів до побудови гнучких і адаптивних СПР, здатних підтримувати структурне представлення знань у процесі прийняття рішень, а також розвиток методів для формування ефективних стратегій вибору та аналізу альтернатив [4].

3. Підхід до проектування баз знань

Розробка баз знань для систем підтримки прийняття рішень (СПР) є одним із найважливіших етапів у побудові ефективної системи управління складними предметними областями. Бази знань забезпечують СПР необхідною інформацією та концептуальними моделями, що дозволяють ЛПР обирати оптимальні варіанти рішень. Сучасні підходи до створення баз знань для СПР зосереджені на максимальній адаптивності та гнучкості структур, що дозволяє системі відповідати потребам користувачів у динамічних умовах.

Особливістю таких баз знань є багаторівневе представлення інформації: бази знань мають підтримувати як загальні концептуальні знання, так і спеціальні прикладні моделі, орієнтовані на розв'язання конкретних функціональних задач. Важливість адаптивності баз знань пояснюється тим, що предметні області, з якими працює СПР, часто зазнають змін, що стосуються зовнішніх умов, нових методів управління або оновлених вимог. Тому ефективне СПР має забезпечувати структуроване й узгоджене представлення знань, яке відповідає поставленим перед системою завданням [5].

Два основні підходи до побудови баз знань включають проблемно-орієнтований і онтологічний аналіз предметної області. Проблемно-орієнтований підхід ґрунтується на вивченні інформаційних потреб користувачів і побудові моделей на їх основі. Він забезпечує точне представлення знань для специфічних завдань, але може обмежувати гнучкість у непередбачених ситуаціях. Онтологічний аналіз, у свою чергу, спрямований на побудову концептуальної моделі, що охоплює найважливіші аспекти предметної області та є універсальною для різних задач. Такий підхід дозволяє створити базу знань, яка може застосовуватися в різних функціональних моделях, зберігаючи при цьому внутрішню узгодженість і актуальність.

При побудові ефективної бази знань обидва підходи доцільно комбінувати. Онтологічний аналіз створює основу для універсальної концептуальної моделі, що не прив'язана до конкретних завдань. На цій базі вже можуть будуватися прикладні моделі, орієнтовані на розв'язання конкретних проблем. У результаті створюється система знань, яка може швидко адаптуватися до нових завдань і відповідати поточним потребам СПР. Завдяки такій структурі можна обирати рівень абстракції відповідно до типу задачі: для загальних задач використовувати абстрактніші моделі, тоді як для конкретних — деталізованіші прикладні моделі.

З метою забезпечення багатоаспектного представлення знань про складні предметні області використовуються прикладні моделі концептуалізації. Кожна з цих моделей описує певний фрагмент предметної області, дозволяючи розглядати об'єкти та процеси в потрібному ракурсі. Це дає можливість оптимізувати процес прийняття рішень, використовуючи лише ті знання, які є суттєвими для поточних завдань.

Побудова баз знань для СПР також передбачає організацію структурних відношень між прикладними моделями концептуалізації, що дозволяє інтегрувати різні аспекти знань у єдину систему. Серед таких відношень виділяють агрегування та узагальнення моделей, що забезпечує системне відображення предметної області. Агрегація дозволяє поєднувати моделі для розгляду комплексних явищ, а узагальнення дає можливість формувати абстрактні моделі на основі конкретних прикладних знань.

Таким чином, ефективне проектування баз знань для СПР потребує комплексного підходу, який передбачає поєднання універсальності онтологічних моделей і специфічності прикладних. Це забезпечує гнучкість системи, її здатність до адаптації та можливість підтримувати ЛПР на різних етапах прийняття рішень [6].

4. Висновки

У сучасних умовах швидких змін і зростання складності управлінських задач розробка адаптивних систем підтримки прийняття рішень (СПР) стає ключовим завданням. Запропонований підхід до проектування баз знань, що поєднує онтологічні та прикладні моделі, забезпечує необхідну гнучкість і ефективність у роботі таких систем.

Онтологічні моделі формують концептуальний фундамент для універсального представлення знань, тоді як прикладні моделі дозволяють налаштувати базу знань для вирішення конкретних задач користувачів. Завдяки цьому бази знань стають не лише адаптивними, але й здатними забезпечувати якісну підтримку процесу прийняття рішень навіть у непередбачуваних ситуаціях.

Таким чином, використання комплексного підходу до проектування баз знань для СПР сприяє ефективнішій підтримці управлінських процесів і створює нові можливості для впровадження інновацій у складних предметних областях. Це відкриває перспективи для подальшого вдосконалення таких систем у напрямку більшої інтеграції знань і технологій.

Література

[1] Левикін В.М., Неофітна Т.М. Категорійне моделювання предметних областей на основі знань у системах прийняття рішень. // Науковий вісник Кременчуцького університету економіки, інформаційних технологій та управління. – 2006. – № 4(14). – С. 21–25.

[2] Кокорева Л.В., Перевозчикова О.Л., Ющенко К.Л. Діалогові системи представлення знань. – К.: Наукова думка, 1992. – 448 с.

[3] Герасимов Б.М., Тарасов В.А., Токарєв І.В. Людино-машинні системи прийняття рішень з елементами штучного інтелекту. – К.: Наукова думка, 1993. – 183 с.

[4] Джексон П. Вступ до експертних систем: Пер. з англ. – М.: Видавничий дім «Вільямс», 2001. – 624 с.

[5] Кіппхан Г. Енциклопедія з друкованих засобів інформації. Технології та способи виробництва: Пер. з нім. — М.: МГУП, 2003. — 1280 с.

[6] Левикин В.М., Неофітна Т.М. Синтаксис і семантика наслідування в концептуальних моделях предметних областей // Вісник НТУ «ХПІ». Нові рішення у сучасних технологіях. — Харків: 2002. — № 7. — С. 6–11.

DOI: https://doi.org/10.30837/IST-2024.P1.112

Рішення для визначення температури точки роси за вологою на основі нерівномірних інтерполяційних таблиць

Віктор Луценко[1†], Сергій Шапко[1*†] та Микола Смаглюк[2†]

[1] *Інститут транспорту газу, АТ "Укртрансгаз", бульвар Гончарівський, 16, Харків, 61004, Україна.*
[2] *АТ "Укргазвидобування", вул. Кудрявська, 26/28, м. Київ, 04053, Україна*

Анотація
У цьому повідомленні надано простий евристичний алгоритм обчислення точки роси за вологою на базі нерівномірних інтерполяційних таблиць [1]. За цим алгоритмом є змога визначати температуру точки роси за вологою та вологовміст природного газу за абсолютного тиску природного газу, що дорівнює 3,92 МПа. Наведено програмну реалізацію цього алгоритму.

Ключові слова
температура точки роси за вологою, вологовміст, алгоритм, програма

Abstract
This paper presents a simple heuristic algorithm for calculating the dew point by moisture content based on uneven interpolation tables [1]. The algorithm is aimed at determining the dew point temperature by moisture and moisture content at an absolute pressure of natural gas equal to 3.92 MPa. The software implementation of this algorithm is presented.

Keywords
dew point temperature by moisture, moisture content, algorithm, program

1. Вступ

У газовій галузі чинні нормативні документи, які унормовують порядок та алгоритми визначання фізико-хімічних показників природного газу. Це повідомлення містить алгоритм, що дозволяє автоматизувати визначання температури точки роси за вологою та вологовміст за абсолютного тиску природного газу 3,92 МПа, на підставі таблиць, що наведені у нормативному документі СОУ 09.1-30019775-186 [1].

Цей нормативний документ чинний для газовидобувних підприємств. Інтерполяційні таблиці наведені в [1]. Точки інтерполяційних таблиць розраховані за алгоритмом, що відповідає ДСТУ EN ISO 18453 [2].

Актуальність повідомлення – автоматизовано визначання точок роси за вологою та вологовмісту в природному газі у польових умовах. Ціль такого визначання точок роси за вологою та вологовмісту – моніторинг працездатності потокових вологомірів, що встановлені на комерційних вузлах обліку природного газу. Характерною особливістю таблиць, наведених у [1], є інтерполяційна нерівномірність, яка ускладнює роботу в автоматичному режимі, тобто таблиці мають не однаковий крок квантування.

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
* Corresponding author.
† These authors contributed equally.
lutcenko-va@utg.ua (V. Lutsenko); shapko-sv@utg.ua (S. Shapko); mykola.smaglyuk@ugv.com.ua (M. Smaglyuk)
0000-0003-0491-3280 (V. Lutsenko); 0009-0005-8755-3066 (S. Shapko); 0009-0002-8233-7389 (M. Smaglyuk)

2. Опис алгоритму

Вхідними даними для алгоритму є абсолютний тиск природного газу ($Р_{абс}$) за якого була виміряна температура точки роси за вологою ($ТТР_в$) та температура точки роси за вологою ($ТТР_в$).

Визначаєм у таблиці стовбець із $Р_{абс}$, за якого виміряна $ТТР_в$ і рядок із значенням, що є найближчим до виміряної $ТТР_в$ та отримуємо значення t_n, t_{n+1}. У цьому рядку отримуємо значення вологовмісту B_{wx} .

Знаходимо в таблиці ще один рядок, найближчий до попереднього, у якому обираєм діапазон значень $ТТР_в$, який містить виміряне значення $ТТР_в$.

Інтерполюємо значення вимріяної $ТТР_в$ за формулою:

$$t_x = t_n + (t_{n+1} - t_n) * \frac{(p_x - p_n)}{(p_{n+1} - p_n)}, \quad (1)$$

Інтерполюємо значення вологовмісту, що відповідає виміряній $ТТР_в$, за формулою:

$$B_{wx} = B_{wnn} + (B_{wn} - B_{wn+1}) * \frac{(p_x - p_n)}{(p_{n+1} - p_n)}, \quad (2)$$

Розраховуємо температуру точки роси за вологою за абсолютного тиску 3,92 МПа за формулою:

$$t_x^{3.92} = t_n^{3.92} + (t_n^{3.92} - t_{n+1}^{3.92}) * \frac{(p_x - p_n)}{(p_{n+1} - p_n)}, \quad (3)$$

Результатом роботи алгоритму є розраховане значення вологовмісту природного газу B_{wx} у мг/м^3 за температури точки роси за вологою ($ТТР_в$), виміряної за абсолютного тиску природного газу ($Р_{абс}$), а також зведена до тиску 3,92 МПа температура точки роси ($t_x^{3.92}$). Невизначеність розрахунку відповідних величин не перевищує значень, унормованих у ДСТУ EN ISO 18453 [2].

3. Програмна реалізація

Алгоритм реалізований на мові програмування – C#.

Відеокадр роботи програми з вхідними даними $Р_{абс}$ = 0,65 МПа. та $ТТРв_{вим}$ = -18,5 ^{0}C наведений, як приклад, на рисунку 1.

Рисунок 1. Результат обчислення програмою.

4. Висновки

Описаний евристичний алгоритм дозволяє розраховувати вологовміст природного газу за температури точки роси за вологою, виміряної за абсолютного тиску природного газу та зведену до тиску 3,92 МПа температуру точки роси за вологою в усьому діапазоні таблиць в [1].

Програмне забезпечення, що реалізує наданий алгоритм, можна використовувати для моніторингу потокових вологомірів у польових умовах.

Література

[1] СОУ 09.1-30019775-186:2019 Система забезпечення якості. Порядок визначення фізико-хімічних показників природного газу, що надходить з об'єктів АТ "Укргазвидобування".

[2] ДСТУ EN ISO 18453:2021 (EN ISO 18453:2005, IDT; ISO 18453:2004, IDT) Природний газ. Кореляція між умістом вологи та точкою роси за водою.

DOI: https://doi.org/10.30837/IST-2024.P1.115

Методи автоматизованої генерації контенту для соціальних мереж

Ольга Ворочек[1] та Ілля Соловей[1†]

[1] *Харківський Національний Університет Радіоелектроніки, Проспект Науки 14 61166 Харків, Україна*

Анотація
Дослідження присвячено генерації контенту у соціальних мережах та автоматизації цього процесу за допомогою технологій Штучного Інтелекту. Розглядаються різні з точки зору створення маніпулятивного контенту підходи, описується структура взаємодії програмної системи із сервісом.

Ключові слова
штучний інтелект, мовні моделі, пропаганда, соціальні мережі

1. Вступ

Сьогодення характеризується високим впливом громадської думки на прийняття державами суспільно-значимих рішень. Інформація давно стала стратегічним ресурсом, але на поточний момент часу від достовірності загальнодоступних даних залежить успіх або невдача економічних, суспільних процесів, реформ у світовому порядку. Досягнення стратегічних цілей можливо лише за умови успішного маніпулювання відповідною інформацією, у тому числі у соціальних мережах.

Фактично, у віртуальному середовищі постійно мають місце інформаційні війни, як один з компонентів гібридної війни. Одним з найперспективніших інструментів генерації контенту для формування суспільної думки є методи та моделі штучного інтелекту.

На поточний момент часу існуючі підходи хоча і вирішують задачі генерації маніпулятивного контенту, але це носить певною мірою хаотичний характер, зосереджуючись на великих обсягах повідомлень та не враховуючи кореляцію згенерованих постів і мети інформаційної атаки.

Таким чином, спираючись на вищесказане, розробка і дослідження методів та моделей програмних систем автоматизованої генерації контенту соціальних мереж, які б адаптувалися до контексту поставленої задачі і повністю відтворювали б поведінку реального користувача, є задачею затребуваною і актуальною.

2. Аналіз останніх досліджень і публікацій

Соціальні мережі відіграють значну роль у нашому повсякденному житті. Згідно даних Datareportal [1] 4.74 млрд людей по всьому світу використовують соціальні мережі, що зіставляє приблизно 59.3% глобальної популяції. Згідно з даними Statista, середній час використання соціальних мереж виріс з 111 хв у 2015 до 143 хв на день у 2024 році [2].

У статті “Соціальні медіа та політична пропаганда” [3] зазначається, що соціальні медіа стали однією з найвпливовіших платформ у майже всіх аспектах суспільства, включаючи

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
olga.vorochek@nure.ua (A. 1); illia.solovei@nure.ua (A. 2)
0000-0002-9054-9894 (A. 1); 0009-0005-5715-2755 (A. 2)

освіту, розваги, охорону здоров'я, економіку та політику. Найбільш популярні платформи в цьому контексті – Facebook, WhatsApp, X, Instagram, TikTok, YouTube, LinkedIn, Telegram та Snapchat.

Згідно з польовим статутом США, пропаганда визначається як будь-яка інформація, ідеї, доктрини або спеціальні методи впливу на думки, емоції, настанови чи поведінку будь-якої спеціальної групи з метою отримання прямих чи непрямих переваг. Це визначення підкреслює маніпулятивний характер пропаганди, спрямований на досягнення певної мети через зміну сприйняття та світогляду аудиторії [4].

Основні етапи пропаганди представлені на рисунку.

Рисунок 1: Графічне зображення основних етапів пропаганди

Ефективне досягнення поставленої мети в пропаганді вимагає не лише впливу та формування світогляду у цільової аудиторії, але й провокування певних дій або реакцій з її боку [5]. Невдала або неправильно побудована пропаганда може досягти стадії, коли інформація засвоюється свідомістю, але не перетворюється у зміну світогляду, необхідну для досягнення заданих дій. Відповідно, очікувана реакція залишається відсутньою.

Пропаганда активно використовується у інформаційній війні проти України протягом багатьох років. Вона може набувати різноманітних форм, таких як фейкових каналів, таргетованої реклами для іноземної аудиторії, Twitter-ботів на основі штучного інтелекту.

Штучний інтелект (ШІ) має застосовуватись не лише для створення, але й для аналізу інформаційного впливу. Наприклад, він може оцінювати ефективність поширення контенту, його охоплення та вплив на аудиторію. Окрім створення пропагандистського контенту, ШІ є ключовим у виявленні ворожих кампаній дезінформації. ШІ може використовуватися для автоматичного аналізу великої кількості даних, що надходять з різних джерел, для виявлення неправдивої інформації та розпізнавання змінених (підроблених) зображень [6].

3. Мета й завдання роботи

Мета даної роботи полягає у дослідженні методів автоматизованої генерації контенту для соціальних мереж, побудова схеми взаємодії програмної системи із компонентом створення нової інформації.

4. Огляд моделей та методів генерації повідомлень

Для генерації повідомлень у сучасному світі широко застосовуються можливості штучного інтелекту, зокрема великі мовні моделі (Large Language Models, LLMs). Ці моделі, що базуються на глибокому навчанні, стали невід'ємною складовою створення та обробки текстового контенту завдяки здатності аналізувати та відтворювати людську мову на високому рівні.

Великі мовні моделі здебільшого побудовані на основі архітектур типу трансформерів, які забезпечують високу ефективність завдяки можливості паралельної обробки даних. Основою їхньої роботи є великі набори текстових даних, на яких моделі проходять навчання. Весь процес навчання мовної моделі поділяється на кілька ключових етапів:

1. Збір та підготовка даних. На першому етапі здійснюється ретельний підбір та обробка текстових даних, що включає як очищення даних від шуму, так і відбір релевантних текстових сегментів для навчання. Масштаб даних на цьому етапі впливає на здатність моделі до генерації та точності відтворення складних лінгвістичних структур.
2. Попереднє навчання. На етапі попереднього навчання модель навчається на загальних текстових наборах, що містять інформацію з різних галузей знань. Це дозволяє моделі оволодіти базовими знаннями про мову, правила граматики, синтаксису та семантики
3. Тонке налаштування (fine-tuning). Після попереднього навчання здійснюється спеціалізоване налаштування моделі, яке адаптує її до конкретних завдань. Тонке налаштування може проводитися за допомогою специфічних наборів даних, що дозволяє моделі демонструвати вищу точність у певних контекстах.
4. Валідація та тестування. Завершальний етап включає валідацію та тестування моделі, де перевіряється її здатність до генерування змістовних і релевантних текстів. Цей процес дозволяє оцінити продуктивність моделі на різних мовних задачах та виявити її слабкі місця для подальшого вдосконалення.

При використанні мовних моделей для генерації повідомлень у соціальних мережах у пропагандистських цілях постає питання цензури, яка відбувається з наступних причин:

1. Етичні міркування. Розробники моделей застосовують методи фільтрації та класифікації контенту, які дозволяють моделі ігнорувати чи виправляти потенційно образливий контент або дискримінаційних висловлювання.
2. Юридичні вимоги. Застосування мовних моделей у глобальному масштабі вимагає врахування законодавства різних країн щодо мови ворожнечі та пропаганди.
3. Відповідальність розробників. Команди розробників несуть відповідальність за контроль над змістом, який можуть генерувати їхні моделі. Це забезпечує безпечне використання продукту та захист користувачів від небажаного контенту.
4. Комерційні причини. Для залучення широкої аудиторії мовні моделі повинні відповідати очікуванням користувачів та дотримуватись культурних норм. Забезпечення прийнятності контенту для широкого кола користувачів сприяє підвищенню довіри та підтримки бренду серед різних сегментів споживачів.

Отже, для виконання поставлених цілей у вигляді генерації повідомлень у соціальних мережах перед нами постає проблема обходу цензурування. Подібний підхід існує і називається джейлбрейкінгом (з англ., втеча з “в’язниці”) – коли мовній моделі нав’язують роль іншого "актора", завдяки чому вона генерує нецензурований контент.

Серед популярних підходів можна виділити Do Everything Now (DAN) [7]. Це тип команди, призначеної для джейлбрейку ChatGPT, тобто для обходу обмежень і доступу до повного функціоналу. У результаті джейлбрейку мовна модель здатна генерувати відповіді, які за звичайних обставин можуть вважатися неприйнятними або неетичними.

Проблема полягає в тому, щоб знайти працюючу інструкцію DAN, оскільки OpenAI регулярно оновлює свою платформу, щоб запобігти неправильному використанню. Через це у останніх версіях мовних моделей використання методу є або проблематичним, або неможливим. Члени команди Adversa знайшли інший підхід, об’єднавши старі принципи джейлбрекінгу завдяки принципу “кролячої нори” [8].

Взаємодію користувача із сервісом можна описати наступним чином:

1. Користувач надсилає запит за допомогою чат-боту.
2. Повідомлення відправляється на сервер, який за допомогою власного Штучного Інтелекту фільтрує повідомлення, та прибирає непотрібні збіги з реальними особами.

3. Відфільтроване повідомлення відправляється великій мовній моделі, та результат повертається на сервер. Кроки повторюються до досягнення необхідного результату.
4. Після обміну повідомленнями між сервером та мовною моделлю, результат повертається у чат-бот.

Рисунок 2: Взаємодія користувача із сервісом генерації повідомлень.

5. Висновки й перспективи подальшого розвитку

Під час дослідження було визначено траєкторію розвитку програмної системи для генерації повідомлень у соціальних мережах. При подальшій розробці важливо винайти свій неповторюваний підхід джейбрейкінгу, так як неприпустимо у розробці покладатися на широкодоступні методи, функціонування яких може бути припинено у будь-який момент.

Штучний Інтелект є перспективним напрямком, але існує черга невирішених питань, які потребують подальшого дослідження, такі як розробка методів підвищення релевантності згенерованого контенту контекстові цільової аудиторії, побудова системи критеріїв якості контенту та методів прогнозування ступеню впливу.

Література

[1] DATAREPORTAL, Global Social Media Statistics, 2024. URL: https://datareportal.com/social-media-users.
[2] Statista, Daily time spent on social networking by internet users worldwide from 2012 to 2024, 2024. URL: https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/.
[3] O. G. Uwa, A. C. Ronke, Social media and political propaganda: A double-edged sword for democratic consolidation in Nigeria, International Journal of Social Science, Management and Economics Research 1 (2023) 1–15. doi:10.61421/IJSSMER.2023.1201.
[4] L. R. Blank, Media Warfare, Propaganda, and the Law of War, Cambridge University Press, Cambridge, 2017.
[5] Є. Глущук, Фейки як інструмент тиску в умовах війни: специфіка застосування та сприйняття, Наукові праці Національної бібліотеки України імені В. І. Вернадського 67 (2023) 96–107. doi:10.15407/np.67.096.
[6] С. Ю. Федоренко, Штучний інтелект та виявлення дезінформації: можливості та виклики, Міжнародна та національна безпека: теоретичні і прикладні аспекти : матеріали VIII Міжнар. наук.- практ. конф. 2 (2024) 391–392. doi:10.31733/15-03-2024/2/391-392.
[7] Workmind, What is DAN Prompt for ChatGPT, 2024. URL: https://workmind.ai/blog/dan-prompt-for-chatgpt/.
[8] Adversa.ai, GPT-4 jailbreak and hacking via rabbithole attack, prompt injection, content moderation bypass and weaponizing ai, 2023. URL: https://adversa.ai/blog/gpt-4-hacking-and-jailbreaking-via-rabbithole-attack-plus-prompt-injection-content-moderation-bypass-weaponizing-ai/.

DOI: https://doi.org/10.30837/IST-2024.P1.119

Exploring the methods for using state machines to optimize and enhance cross-platform mobile applications

Vladyslav Martynov[1,*,†], Ihor Shubin[1,†] and Victoria Skovorodnikova[1,†]

[a] *Kharkiv National University of Radio Electronics, Nauky ave., 14, Kharkiv, 61166, Ukraine*

Abstract

Nowadays, contemporary mobile apps require robust state management, particularly in cross-platform frameworks such as React Native. Finite State Machines (FSMs) provide a structured methodology for managing state transitions, enhancing stability, performance, and testability in complex applications. This article examines the advantages of FSMs and their practical applications in cross-platform development, with a focus on React Native.

Keywords

FSM, React, React-Native, XState

1. Introduction

As modern mobile applications become increasingly embedded in users' daily routines, delivering responsive and efficient functionality is paramount. These applications serve a vast array of purposes—quick access to information, seamless communication, task management, and beyond—placing high demands on developers to continually enhance functionality, performance, and reliability. Cross-platform frameworks like React-Native empower developers to deploy applications across multiple platforms, streamlining development and reducing resource expenditure. Yet, cross-platform solutions introduce distinct challenges, particularly in state management.

Effective state management remains a central challenge in cross-platform development. Complex applications, especially those with multifaceted user interactions, asynchronous tasks, and extensive use of device resources, rely on smooth state transitions to ensure consistent behavior. Without a robust architecture for state management, applications risk crashes, unpredictable behaviors, and a diminished user experience.

Finite State Machines (FSMs) offer a systematic approach to addressing these challenges. By rigorously defining all possible states and permissible transitions, FSMs enforce predictable and stable application behavior, proving especially useful in managing navigation flows, handling asynchronous operations, and structuring complex logic. Within mobile applications, FSMs excel at optimizing scenarios driven by external and internal events, such as user authentication and server requests, where precise state control is essential.

A major advantage of FSMs is their ability to impose a well-defined structure on application control logic, minimizing errors and simplifying state tracking. This structure enhances testability, allowing developers to independently assess each state and transition and

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine

* Corresponding author.

† These authors contributed equally.

vladyslav.martynov@nure.ua (V. Martynov); ihor.shubin@nure.ua (I. Shubin); viktoriia.skovorodnikova@nure.ua (V. Skovorodnikova)

0009-0003-5665-2323 (V. Martynov); 0000-0002-1073-023X (I. Shubin); 0000-0003-0436-4197 (V. Skovorodnikova)

proactively identify potential issues. Furthermore, FSMs contribute to performance optimization by streamlining transitions and reducing unnecessary computations, a critical consideration for cross-platform solutions where execution speed directly impacts user satisfaction.

2. Review of the literature

The concept of state is very familiar. Consider a toaster: initially, the toaster is in the off state; when you push the lever down, a transition is made to the on state and the heating elements are turned on; finally, when the timer expires, a transition is made back to the off state and the heating elements are turned off.

A finite state machine (FSM) consists of a set of states s_j and a set of transitions between pairs of states s_i, s_j . A transition is labeled condition/action: a condition that causes the transition to be taken and an action that is performed when the transition is taken. FSMs can be displayed as a state diagram Fig.1.

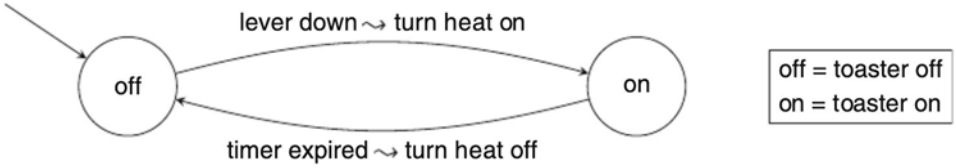


Figure 1: State diagram

A state is denoted by a circle labeled with the name of the state. States are given short names to save space and their full names are given in a box next to the state diagram. The incoming arrow denotes the initial state. A transition is shown as an arrow from the source state to the target state. The arrow is labeled with the condition and the action of the transition. The action is not continuing; for example, the action turn left means set the motors so that the robot turns to the left, but the transition to the next state is taken without waiting for the robot to reach a specific position [1].

React Native is an open source framework for building Android and iOS applications using React and the app platform's native capabilities. With React Native, you use JavaScript to access your platform's APIs as well as to describe the appearance and behavior of your UI using React components: bundles of reusable, nestable code [2].

Unlike hybrid frameworks that run within a web view, React Native compiles to native components, providing applications with a look and feel similar to those built with platform-specific languages like Swift for iOS or Kotlin for Android.

However, managing complex user flows in React Native can be challenging due to platform-specific differences, which impact component behavior, navigation patterns, and resource management. These differences often require additional logic to handle varied platform responses, such as asynchronous tasks, permission handling, and native animations, which can lead to increased complexity and potential inconsistencies in user experience. Addressing these challenges effectively requires a structured approach to state management, particularly when handling intricate navigation and interactive flows across platforms.

Finite State Machines (FSMs) offer an effective solution for such state management challenges by enforcing a strict and predictable sequence of states and transitions. In scenarios involving complex workflows—like multi-step forms, authentication flows, or dynamic content rendering—FSMs provide developers with a clear framework for defining each state and its permissible transitions. This structure mitigates the unpredictability that can arise from asynchronous tasks or varying platform behavior, helping to ensure consistent responses to user actions across iOS and Android.

By using FSMs within React Native applications, developers gain the ability to encapsulate complex user interactions in a way that is both testable and maintainable. Each state transition

is explicitly defined, which reduces the likelihood of unintended behaviors and facilitates debugging. In cross-platform environments where maintaining uniformity is essential, FSMs provide a reliable approach for handling intricate application states, enhancing both the predictability and resilience of mobile applications.

3. Experiments

In this experiment, the basic authentication screen will be implemented with advanced, specialized library called Xstate and then will be compared with default React Native implementation.

XState is a state management and orchestration solution for Javascript and Typescript apps. It uses event-driven programming, state machines, statecharts, and the actor model to handle complex logic in predictable, robust, and visual ways. XState provides a powerful and flexible way to manage application and workflow state by allowing developers to model logic as actors and state machines. It integrates well with React, Vue, Svelte, and other frameworks and can be used in frontend, backend, or wherever Javascript runs [3].

XState enhances the clarity and robustness of state management through its advanced syntax, which allows developers to encapsulate and manage state logic directly within the machine configuration. XState's internal context functions as a dedicated data store, managing state-specific information such as validation flags, input values, and error messages. This encapsulation minimizes reliance on external state management solutions, consolidating all relevant logic and data within a single, cohesive unit.

4. Results

To evaluate the performance of both the React Native default implementation and XState implementations, we used three key metrics: the React Native UI thread, the JS thread, and FPS drop. The UI and JS thread metrics measure the responsiveness of the application by tracking frame rates on each thread, with higher values indicating smoother performance. FPS drop, on the other hand, reflects any decrease in frames per second during interactions, with lower values indicating fewer disruptions in user experience. These metrics provide a comprehensive view of each implementation's ability to handle complex interactions smoothly.

Based on Table 1 metrics, XState demonstrates superior performance across multiple dimensions. In the XState implementation, both the UI and JS threads maintain a steady average of 60 frames per second (FPS). This consistent performance indicates that XState efficiently handles state transitions and user interactions without overloading either thread. The minimal FPS drop (2-3) further suggests that XState is capable of processing complex interaction patterns smoothly, keeping the application responsive and ensuring a seamless user experience even under heavy usage conditions. This level of stability is crucial in maintaining an interactive flow, especially on an authentication screen where real-time validation and feedback are essential.

In contrast, the React Native default implementation shows performance limitations, with the UI thread averaging 55 FPS and the JS thread at 50 FPS. This reduction in frame rates suggests that it introduces more load on the system, potentially due to less optimized transition handling or higher overhead in managing state changes. The FPS drop is also significantly higher, ranging from 10 to 12 FPS during intensive interactions. This larger drop indicates that default implementation experiences performance bottlenecks under complex user actions, resulting in occasional stutters and a less fluid experience. Such delays can negatively impact user perception, especially in scenarios where users expect immediate feedback, such as entering login credentials.

In summary, XState's structured approach and optimized handling of state transitions provide a clear advantage in terms of performance, maintaining high frame rates on both UI and JS threads and minimizing FPS drops.

Table 1
Performance metrics

Metrics	XState	React Native default implementation
UI Thread	60	55
JS Thread	60	50
FPS Drop	2-3	10-12

5. Conclusion

This study demonstrates the effectiveness of using finite state machines to manage complex interactions within mobile applications, particularly in cross-platform frameworks like React Native. By implementing an authentication screen in two variants — React Native default implementation and the XState library—we explored the performance, maintainability, and user experience benefits of each approach.

The XState implementation proved to be superior in terms of performance, with consistently high frame rates and minimal FPS drops, ensuring a smooth and responsive experience. Its advanced abstractions, such as context handling, hierarchical state management, and visual state charts, make it especially well-suited for applications with intricate interaction flows. The ability to visualize state charts not only aids in understanding and debugging complex workflows but also facilitates collaboration among team members and provides a form of living documentation as the application evolves.

React Native default implementation, demonstrated some limitations in handling resource-intensive transitions and maintaining consistent frame rates under heavy user interactions. Although it offers basic state management, the overhead associated with managing states manually makes it less suitable for highly interactive or complex forms, especially when compared to the modular and declarative approach offered by XState.

In conclusion, XState is a powerful tool for managing complex forms and dynamic interactions across various applications. Its structured approach to state management makes it versatile enough for a wide range of use cases, including multi-step forms, onboarding flows, authentication processes, and any scenario requiring predictable state transitions. The library's modular design allows developers to maintain a clean codebase, improve performance, and enhance user experience. As mobile applications continue to grow in complexity, adopting a well-defined state management solution like XState can play a critical role in building robust, scalable, and maintainable applications.

Future work could explore extending this approach to more intricate applications involving real-time updates, multi-user collaboration, or machine learning-driven interactions. Furthermore, the potential for combining XState with other tools in the React Native ecosystem, such as gesture handlers and animations, opens up possibilities for crafting highly interactive and intuitive user experiences.

References

[1] Finite State Machine URL: http://surl.li/xbonql (date of access 20.11.2024)
[2] React-Native URL: http://surl.li/bwzllc (date of access 18.11.2024)
[3] XState URL: https://stately.ai/docs/xstate (date of access 15.11.2024)

DOI: https://doi.org/10.30837/IST-2024.P1.123

Automation of Styling Conversion for Internet Resources

Ganna Sydorenko[1-2*†] and Andrii Biziuk[2†]

[1] *National Technical University«Kharkiv Polytechnic Institute», 2, Kyrpychova str., 61002, Kharkiv, Ukraine*
[2] *Kharkiv National University of Radio Electronics, Nauky Avenue, 14, Kharkiv, 61166, Ukraine*

Abstract
The article examines methods and tools for transforming styling for web resources. In modern web development, ensuring adaptive and correct display of web resources across different devices and browsers is one of the key requirements. One important aspect of this process is automating the conversion of styles to ensure compatibility with various platforms and optimize the user experience. The article discusses the main methods and tools used for automated CSS style transformation, as well as proposes an algorithm that effectively adapts styles for different browsers and devices. The development of such software tools contributes significantly to improving web development efficiency by reducing adaptation time and enhancing the quality of the final products.

Keywords
styling, internet resources, responsive design, automation, web development

1. Introduction

Modern web resources require constant adaptation and optimization to ensure an effective user experience. One of the main aspects of this adaptation is the correct display of styles across various devices, browsers, and operating systems. In this context, styling conversion for internet resources becomes an important stage in the web development process.

Website development often requires substantial time for design and color palette selection. Clients frequently lack a clear vision of the final site's appearance. The primary task of the designer is to meet the client's expectations, taking all details into account to create a quality product [1-2]. Preparing and formatting text materials for electronic publications differs from preparing printed versions. It's essential to consider the specifics of reading on a screen, including font choice, line spacing, and overall page composition.

2. Goal and objectives of the article

Developing software that automates this process can significantly reduce the time spent on manual style editing and make developers' work easier. The purpose of this study is to develop a software tool that automates the process of style configuration for web resources.

To ensure user comfort on a website, it is necessary to utilize all the advantages of the electronic format, such as using color in the design of text blocks and providing options for image enlargement. The presence of hyperlinks and relevant tooltips is also an important component. Working with color is one of the key aspects of the design process. The choice of color scheme should be based on the fundamental principles of color theory, taking into account the symbolic meanings and the impact of colors on user perception. Therefore, the purpose of this study is to create software that helps develop styling designs for internet resources.

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
* Corresponding author.
† These authors contributed equally.
ganna.sydorenko@khpi.edu.ua (G. Sydorenko); andrii.biziuk@nure.ua (A. Biziuk)
0000-0002-0761-2793 (G. Sydorenko); 0000-0001-9830-9206 (A. Biziuk)

The main tasks are:

- Researching modern tools for style adaptation.
- Developing algorithms for automatic style transformation.
- Creating a user interface for configuring the style transformation process.
- Evaluating the effectiveness of the proposed software solution.

3. Selection of Graphic Formats for Web Resources

Overview of Modern Automation Tools. Modern tools such as Shopify, Contao, Wix, Webflow, and others facilitate web resource development by providing functionality for creating HTML code and configuring CSS. However, these tools have certain limitations, especially in terms of artistic design, which may require detailed knowledge of color theory and typography. To improve efficiency, new software solutions can offer greater flexibility in style adaptation, combining aesthetic appeal and technical compliance. For example, additional components in Cogear CMS and Contao provide specific features [3-4], but their functionality may be complex for beginners. Popular formats like GIF, JPEG, and PNG are standards for web graphics due to their properties: support for transparency, compression, animation, and progressive loading. PNG offers higher quality but has larger file sizes.

The main properties of graphic files important for use in the internet environment include the following:

1. Transparency: Allows images to vary in opacity, ranging from fully visible to fully transparent.
2. Compression: Reduces the size of graphic files by applying algorithms that decrease file size by processing pixels as a single group.
3. Interlacing: Enables preloading of images by displaying alternating rows, starting with odd rows, followed by even rows, allowing for a quicker overall visual impression.
4. Animation: Creates the illusion of movement through a series of sequential frames. For example, animated GIF files do not require additional browser plugins and can work on many devices.
5. Progressive Loading: Similar to interlacing, it displays an image in stages, initially loading a low-resolution version, followed by progressively higher-quality versions.

For web graphics, the most widely used formats are GIF and JPEG. Their multifunctional capabilities, such as high compression while maintaining sufficient quality for web pages, have established them as standards in the field of web images. The PNG format, available in two versions (PNG-8 and PNG-24), is also supported by browsers, but its prevalence on web resources is lower compared to GIF and JPEG (see Table 1).

Attitudes toward presenting web content are evolving, with new data formats and development tools emerging to support them. However, older formats (such as GIF and JPEG for graphics) remain quite popular. As shown in Table 1, they are supported by nearly all browsers, and most developers have extensive experience with them. It is important to use the correct format for specific purposes to balance image quality and file size. For example, one image in GIF format may take up more space and produce a lower quality result than in JPEG format, while for another image, the opposite may be true. However, for presenting raster images, PNG has become a better option [5].

New automation tools simplify the creation, management, and configuration of content, especially when it comes to color schemes and fonts, which are typically defined through CSS. Tools like Cogear, ImageCMS, and ReloadCMS offer powerful features for both large and small projects but require specific skills for artistic design according to modern design standards [6].

Table 4
Support for Image Formats in Different Browsers

Browser	JPEG	GIF	PNG	SWG	PDF	BMP
Chrome	+	+	+	part	+	+
Chromiun	+	+	+	part	+	+
Mozilla	+	+	+	part	-	-
Mozilla Firefox	+	+	+	part	+	+
Opera	+	+	+	part	+	+
Safari	+	+	+	part	+	+
Internet Explorer	+	+	+	part	-	+

Color selection and typography are critical to web design. The use of colors, such as complementary or triadic schemes, helps create a harmonious impression. Typography also plays an important role in ensuring readability and aligning with the website's theme. Automated algorithms can adjust fonts and colors according to the content.

4. Automating Design with Software

4.1. Title information

As part of the research, software with an intuitive graphical interface was developed, allowing users to easily customize website designs. With this tool, designers can quickly and efficiently change the style of web resources, creating harmonious visual solutions. The software supports choosing between dark and light themes, using a triadic color scheme based on the principles of the Itten color wheel. The main user page (see Fig. 1) can be divided into a workspace block and a user preview block.

A crucial stage is creating a software solution that automates the style transformation process for online resources. For this, an algorithm will be developed to analyze original CSS files, detect potential compatibility issues, and automatically generate necessary changes.

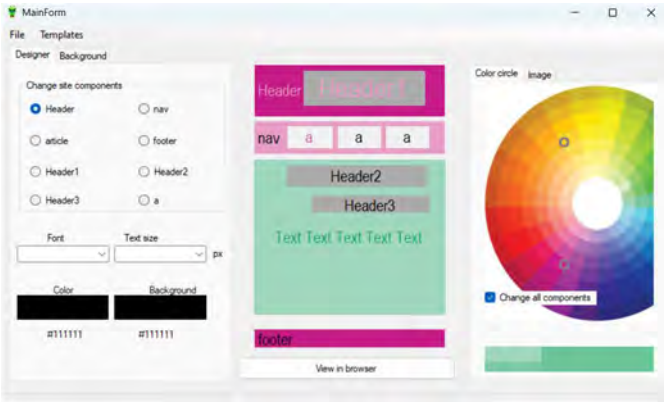
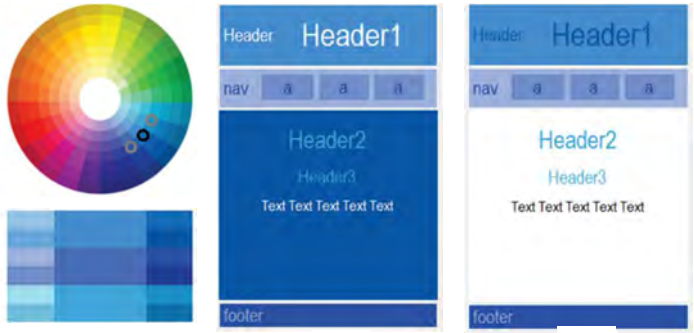


Figure 1: Main user Page.

Software for automating web design simplifies the work of designers by allowing them to quickly create visually consistent solutions. Using artificial intelligence and machine learning, programs can automatically suggest color and typography schemes that match the website's goals. The program supports theme selection (dark/light) based on the triadic color scheme of Itten, providing users with flexible configuration options (Fig. 2).

For the correct functioning of the software tool, the input HTML document should not have explicitly defined styles. It is also necessary for the HTML document to adhere strictly to the HTML 5 language standards for styling.



a. b. c.

Figure 2: Selected styles. a. – Selected colors, b. – Dark scheme, c. – Light scheme

To test the functional capabilities of the software tool, its performance was tested on various HTML pages. It was found that changing accent colors according to Itten's color schemes significantly affects the visual perception of a website, creating the necessary accents and mood. The research showed that the developed color schemes are universal and can be successfully applied to a wide range of websites with diverse content and target audiences.

5. Conclusions

Currently, the visual perception of information on the internet is directly influenced by the color design and typography. With well-chosen colors, it is possible to highlight key elements or hide less important information. Color and typography on web pages can be set using HTML (Hypertext Markup Language), specifically by forming the necessary CSS style sheet. While there are many software tools available for creating such style sheets, they lack the ability for artistic website design based on modern color theory.

Itten's color wheel demonstrates the most successful color combinations. Color schemes combining colors were examined, and a triadic color model was selected. This model provides more options for designing and highlighting key elements on a web page. Automating the style transformation process is an important step in simplifying the work of web developers and improving the efficiency of web resource development. The development of such tools has great potential for further improvement and integration with other web development technologies.

References

[1] Duckett, J. HTML and CSS: Design and Build Websites. Wiley, 2017.
[2] Frain, B. Responsive Web Design with HTML5 and CSS: Develop Future-Proof Responsive Websites Using HTML5 and CSS3. Packt Publishing, 2015.
[3] Jennifer Robbins. Learning Web Design: A Beginner's Guide to HTML, CSS, JavaScript, and Web Graphics. Publisher: O'Reilly Media, 2018.
[4] Alex White. The Fundamentals of Graphic Design – Solon-Press, 2014. – 256 pages.
[5] Steve Krug. Don’t Make Me Think: A Common Sense Approach to Web Usability – ArtHuss, 2022. – 198 pages.
[6] John F. Hughes, Andries van Dam, Morgan McGuire, David F. Sklar, James D. Foley. Computer Graphics: Principles and Practice. Publisher: Addison-Wesley, 2017.

DOI: https://doi.org/10.30837/IST-2024.P1.127

Аналіз алгоритмів кешування в Redis

Матвій Кучапін[1†], Марія Широкопетлєва[1*†] та Олена Шевченко[1†]

[1] *Харківський національний університет радіоелектроніки, проспект Науки, 14, Харків, 61166, Україна*

Abstract

Redis is a widely utilised database management system for various high-performance web applications, analytical tools, and data processing systems. Data caching is one of the principal methods to enhance the performance of such high-loaded systems. This paper analyses the existing approaches to cached data management in Redis, highlighting several common limitations associated with large and complex data structures, such as hashes, sets, lists, etc. It also outlines different data access patterns and the need to apply different caching strategies. The paper proposes the use of adaptive caching algorithms that take into account the data access patterns in other parts of complex structures to implement these strategies.

Ключові слова

Алгоритми кешування, Redis, LRU, LFU, TTL

1. Вступ

У сучасному світі інформаційних технологій, де обсяги даних постійно зростають, з'являються нові виклики щодо ефективної обробки та зберігання інформації. Великі обсяги даних вимагають від розробників програмних систем та баз даних забезпечення швидкого доступу до інформації, мінімізації затримок та оптимальне використання ресурсів.

Одним з ключових методів оптимізації роботи інформаційних систем є кешування, яке дозволяє знизити навантаження на основну базу даних, прискорити обробку запитів та зменшити час відгуку системи. Завдяки цьому даний підхід широко використовується у високонавантажених веб-додатках, системах аналітики та обробки даних, й у багатьох інших галузях, а для забезпечення кешування широко використовується Redis [1].

Метою роботи є дослідження механізмів управління кешованими даними у Redis, що впливають на ефективність використання пам'яті та швидкодію системи.

2. Обґрунтування вибору засобів кешування

Аналізуючи сучасні та популярні in-memory бази даних, що забезпечують швидкий доступ до даних та стабільну роботу систем під значним навантаженням [2], для дослідження було обрано Redis як основний інструмент для кешування. Це зумовлено його високою швидкодією та використанням оперативної пам'яті для зберігання даних, що забезпечує мінімальну затримку при обробці запитів. Redis є ідеальним рішенням для застосунків із критичними вимогами до швидкості обробки інформації [3]. Однак, незважаючи на всі переваги, управління складними структурами даних у Redis пов'язане

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine

* Corresponding author.

† These authors contributed equally.

matvii.kuchapin@nure.ua (М. Кучапін); marija.shirokopetleva@nure.ua (М. Широкопетлєва); olena.shevchenko@nure.ua (О. Шевченко)

0009-0006-1953-2893 (М. Кучапін); 0000-0002-7472-6045 (М. Широкопетлєва); 0000-0003-3177-5530 (О. Шевченко)

з низкою труднощів [4]. Проблеми виникають під час роботи з такими структурами, як хеші, множини і списки, де окремі елементи відрізняються за важливістю та частотою використання. У таких випадках надмірне видалення все ще корисних даних або, навпаки, зберігання старих елементів, які споживають ресурси пам'яті, може негативно позначитися на продуктивності системи. Це призводить до зниження загальної продуктивності через необхідність відтворювати або завантажувати дані повторно. Водночас зберігання застарілої інформації призводить до нераціонального використання пам'яті та знижує ефективність кешування.

Основною проблемою в контексті складних структур даних є відсутність гнучких механізмів управління на рівні окремих елементів усередині цих структур. Це обмежує можливість оптимізації роботи над великими об'єктами, які можуть містити як рідка використовувані, так і часто запитувані елементи.

3. Аналіз існуючих підходів до управління даними в Redis

У сучасних системах управління кешованими даними ключову роль відіграють алгоритми видалення та оновлення даних. Redis, як одна з найпопулярніших систем кешування, реалізує декілька підходів для управління даними: TTL, LRU та LFU. Однак, кожен із цих методів має свої обмеження при роботі зі складними структурами, що впливає на ефективність та продуктивність системи.

TTL є стандартним механізмом керування терміном життя даних у Redis [5]. Він дозволяє призначати певний термін існування кожному ключу, після закінчення якого дані автоматично видаляються з кешу. Це забезпечує контроль над свіжістю даних та автоматичне очищення кешу від застарілих записів. Проте, у випадку великих структур TTL застосовується до всього об'єкта. Це означає, що після завершення терміну життя видаляється вся структура, навіть якщо частина елементів усе ще актуальна. В результаті виникає проблема надмірного видалення корисних даних. Відсутність гнучкого управління TTL на рівні окремих елементів у великих структурах обмежує можливості адаптації системи до різних рівнів важливості та частоти використання даних.

Алгоритм LRU видаляє найдавніше використані елементи, щоб звільнити місце під нові дані [6]. Це дозволяє зберігати у кеші найчастіше використовувані елементи. Однак, при роботі зі складними структурами Redis не завжди коректно визначає, які елементи потрібно видаляти. У випадках, коли деякі елементи у структурі використовуються рідко, але залишаються важливими, LRU може видалити їх, що призводить до втрати критично важливих даних і зниження ефективності кешування. Відсутність розмежування важливості елементів у великих структурах робить цей алгоритм менш ефективним для складних випадків.

LFU орієнтується на частоту використання даних і видаляє найменш часто запитувані елементи [6]. Він більш ефективний для сценаріїв, де частота використання є важливішим критерієм, ніж час останнього доступу. Проте LFU також має свої обмеження при роботі зі складними структурами, тому що він не враховує зв'язки між елементами у структурах, що може призводити до видалення важливих даних, якщо вони запитуються нечасто, але є критичними для програми. Крім того, цей алгоритм потребує додаткових ресурсів для відстеження частоти використання, що ускладнює його реалізацію та знижує продуктивність при високих навантаженнях.

Існуючі підходи управління кешем у Redis мають кілька спільних обмежень, зокрема:

- неефективне використання пам'яті при застосуванні TTL до великих структур даних. Весь об'єкт видаляється після закінчення терміну життя, незалежно від активності окремих елементів, що призводить до перевитрати оперативної пам'яті та необхідності повторного створення кешованих даних;

- відсутність можливості гнучкого управління TTL на рівні окремих елементів у великих структурах, що обмежує адаптивність системи до різних рівнів важливості та частоти використання даних. Це не дозволяє коректно видаляти застарілі елементи з великих структур, які можуть мати різну важливість;
- рідко використовувані елементи у великих структурах продовжують займати пам'ять навіть після того, як їх використання стає неактуальним.

В таблциці 1 наведено результати порівняння алгоритмів LRU та LFU.

Таблиця 5

Порівняльна таблиця алгоритмів LRU та LFU

Параметр	LRU	LFU
Критерій видалення	Час останнього доступу до ключа	Частота доступу до ключа
Переваги	Простота реалізації	Зберігає найпопулярніші ключі
Недоліки	Може видаляти часто використовувані, але нещодавно неактивні ключі	Повільна адаптація до змін у популярності
Підходить для	Динамічних патернів доступу	Стабільних патернів доступу

Дані характеристики вказують на те, що вибір між LFU та LRU залежить від конкретної системи та вимог до кешування. LRU є більш ефективним у системах з динамічними патернами доступу, де актуальність даних змінюється швидко. LFU, навпаки, підходить для систем зі стабільною популярністю даних, де певні ключі постійно користуються попитом. Різні патерни доступу до даних можуть впливати на ефективність кожного з цих алгоритмів, тому вибір алгоритму управління кешем повинен базуватися на аналізі поведінки користувачів та характеру даних у системі.

Загалом, використання стандартних механізмів управління кешем при роботі зі складними структурами даних може призвести до наступних негативних наслідків:

- старі та неактуальні дані можуть залишатися в пам'яті та займати місце, яке могло б бути використане для зберігання актуальної інформації;
- часте видалення та повторне створення великих структур даних збільшує навантаження на систему та призводить до збільшення часу відповіді;
- видалення актуальних даних через недосконалість політик витіснення може призвести до неконсистентності даних та негативно вплинути на користувацький досвід.

При зростанні обсягів даних та кількості користувачів обмеження пам'яті та недостатня гнучкість політик витіснення стають більш помітними, що може обмежити можливості систем.

Одним із перспективних напрямків є розробка динамічних механізмів TTL, які дозволять контролювати час життя даних на рівні окремих елементів складних структур, враховуючи їхню активність та важливість.

Крім того, сучасні дослідження орієнтуються на створення алгоритмів, що поєднують можливості LRU та LFU, але мають покращену гнучкість і здатність адаптуватися до змінних умов роботи системи [7]. Такий підхід дозволяє краще управляти пам'яттю, зберігаючи часто запитувані дані та своєчасно видаляючи рідко використовувані елементи, знижуючи загальну затримку доступу до кешованих даних.

Перспективи розвитку полягають у впровадженні комбінованих алгоритмів, які автоматично підлаштовуються під поточне навантаження та обсяг кешованої інформації

[8]. Подальша інтеграція таких адаптивних алгоритмів у кешуючі системи, зокрема Redis, дозволить ефективніше вирішувати проблеми масштабування кешованих даних, підвищувати продуктивність та забезпечувати високу швидкість доступу до даних у системах із високим навантаженням.

Одним із важливих напрямків у сучасних дослідженнях є адаптивні алгоритми кешування, які здатні пристосовуватися до динамічних патернів доступу до даних.

4. Висновки

Недосконаліть стандартних підходів також проявляється у неможливості тонкого налаштування під конкретні потреби додатків. Стандартні політики витіснення не враховують особливості доступу до даних у різних частинах складних структур. Це може обмежити ефективність кешування та призвести до нераціонального використання ресурсів.

Таким чином, стандартні підходи TTL, LRU та LFU мають суттєві недоліки при роботі зі складними структурами даних у високопродуктивних системах. Для подолання цих проблем необхідно розробити більш гнучкі та адаптивні механізми управління кешем, які враховують специфіку складних структур даних та характер доступу до них, забезпечуючи ефективне використання ресурсів та високу продуктивність системи, що є напрямком подальших досліджень.

Перелік посилань

[1] Redis Docs. 2024. URL: https://redis.io/docs/latest/ (дата звернення 12.10.2024).

[2] Top 27 In-Memory Databases Compared. 2024. URL: https://www.dragonflydb.io/guides/in-memory-databases.

[3] A. Kuzochkina, M. Shirokopetleva and Z. Dudar, Analyzing and Comparison of NoSQL DBMS, 2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), Kharkiv, Ukraine, 2018, pp. 560-564, doi: 10.1109/INFOCOMMST.2018.8632133.

[4] P. Zhang, et al, Redis++: A High Performance In-Memory Database Based on Segmented Memory Management and Two-Level Hash Index, 2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom), Melbourne, VIC, Australia, 2018, pp. 840-847, doi: 10.1109/BDCloud.2018.00125.

[5] J. Liu, X. Wan, Q. Zhu, T. Peng and X. Hu, Research on Adaptive Cache Mechanism Based on TTL, 2022 2nd International Conference on Networking, Communications and Information Technology (NetCIT), Manchester, United Kingdom, 2022, pp. 507-511, doi: 10.1109/NetCIT57419.2022.00125.

[6] Z. Li, D. Liu and H. Bi, CRFP: A Novel Adaptive Replacement Policy Combined the LRU and LFU Policies, 2008 IEEE 8th International Conference on Computer and Information Technology Workshops, Sydney, NSW, Australia, 2008, pp. 72-79, doi: 10.1109/CIT.2008.Workshops.22.

[7] W. Ma, Y. Zhu, S. Wang and Y. Bao, LearnedCache: A Locality-Aware Collaborative Data Caching by Learning Model, 2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), Xiamen, China, 2019, pp. 718-726, doi: 10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00109.

DOI: https://doi.org/10.30837/IST-2024.P1.131

Експериментальне Порівняння Лінійного та Нелінійного SVM-класифікаторів

Катерина Горішня[1*†] та Володимир Кобзєв[1†]

[1] *Харківський національний університет радіоелектроніки, пр. Науки, 14, м. Харків, 61166, Україна*

Анотація
У цій роботі проведено експериментальне порівняння лінійного та нелінійного методів опорних векторів (SVM) для класифікації даних із нелінійними межами. Дослідження виконано на наборі даних make_circles із використанням різних типів ядер. Результати демонструють значну перевагу нелінійних ядер над лінійним у задачах із складними розділювальними межами. Отримані результати підкреслюють важливість правильного вибору ядра для забезпечення максимальної ефективності SVM-класифікації.

Ключові слова
класифікація, метод опорних векторів, лінійне ядро, нелінійне ядро, make_circles, медичні дані

1. Вступ

Обробка наявних результатів медичних досліджень за допомогою методів Big Data дозволяє виявляти закономірності, які пов’язані з перебігом захворювань [1]. Ці методи дають можливість аналізувати величезні обсяги інформації, що утворюються внаслідок сучасних томографічних досліджень. Результатом таких досліджень є набори зображень, які можуть містити ознаки захворювань різного ступеня тяжкості.

Томографічні зображення є основою для аналізу кількості, розмірів, розташування клітин і їх інших характеристик. Класифікація таких зображень за формальними правилами дає медичному персоналу базу для діагностики важких захворювань, визначення стадії розвитку хвороби та підготовки до вибору методів лікування. Актуальність статистичного аналізу такої інформації пояснюється наступними моментами.

1. Онкологічні захворювання є однією з головних причин смертності людей у світі.
2. Дані, пов'язані з онкологічними захворюваннями, мають великий обсяг інформації, використання методів виявлення закономірностей в яких може допомогти у визначенні характерних ознак, що вказують на розвиток раку.
3. Виявлені закономірності або аномалії допомагають окреслити характеристики пацієнта, які визначають ефективний метод лікування.

2. Метод опорних векторів

Для аналізу таких даних важливо вибрати метод, який не лише ефективно класифікує клітини на здорові та хворі, але й забезпечує точну локалізацію ракових клітин на

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
* *Corresponding author.*
† *These authors contributed equally.*
kateryna.horishnia@nure.ua (К. Горішня); volodymyr.kobziev@nure.ua (В. Кобзєв)
0009-0001-9032-4249 (К. Горішня); 0000-0002-8303-1595 (В. Кобзєв)

зображеннях. Метод опорних векторів (SVM) є одним із найефективніших для таких завдань, оскільки дозволяє визначити чіткі межі між класами клітин і виявити приховані закономірності в розташуванні хворих клітин [2, 3].

В основі методу лежить уявлення про гіперплощину, яка розділяє точки (клітини) в просторі відповідно до їх класу. У найпростішому випадку гіперплощина відповідає оболонці органу, навколо якої розташовані клітини. Однак в реальних системах класи рідко бувають лінійно роздільними. У таких випадках застосовуються нелінійні ядра, такі як поліноміальні та радіальні базисні функції, які забезпечують адаптацію до складних форм меж [4].

Локалізація та класифікація клітин на томографічних зображеннях ґрунтується на кількох ключових підходах, які забезпечують точне визначення координат і стану клітин. Першим етапом є визначення центру (середини) групи точок, що представляють клітини на зображенні. Це досягається шляхом обчислення середнього положення точок на основі їх координат. Такий аналіз дозволяє ідентифікувати центральне розташування патології, що є важливим для подальшої діагностики.

Наступним кроком є обчислення радіусу зони класифікації. Радіус визначається як відстань від центру до найбільш віддалених точок хмари, які відповідають клітинам. Це дозволяє встановити межі аналізованої зони, в межах якої відбувається класифікація. Використовуючи SVM, ці межі можуть адаптивно розширюватися залежно від характеристик ядра.

Класифікація точок відбувається у багатовимірному просторі, що дозволяє враховувати їхні просторові координати та інші властивості. Метод SVM створює оптимальну межу між класами, розділяючи точки на здорові та хворі. У випадках, коли ракові клітини знаходяться на межі оболонки органу або виходять за її межі, SVM забезпечує точність класифікації шляхом налаштування ширини розділювальної смуги. Це дозволяє мінімізувати помилки та враховувати специфіку розташування клітин, що є критично важливим для аналізу ранніх стадій онкологічних захворювань.

3. Проведення експерименту

Для проведення експерименту було обрано набір даних make_circles із бібліотеки Scikit-learn [5]. Цей набір є особливо цінним для тестування методів класифікації, оскільки утворює два концентричні кола, які моделюють дані зі складними нелінійними межами. Використання таких даних дозволяє оцінити ефективність різних типів ядер у SVM для класифікації точок у просторових структурах із нетривіальними розділювальними межами.

У рамках експерименту було протестовано лінійне ядро (Linear), поліноміальні ядра другого та третього порядків (Poly-2, Poly-3) та радіальне базисне ядро (RBF). Основними метриками для оцінки ефективності класифікації стали точність (Precision), повнота (Recall) та F1-міра (F1-score). Результати експерименту наведені в таблиці 1.

Таблиця 1

Результати експерименту

Ядро	Precision	Recall	F1-score
Linear	0.5372	0.55	0.5330
Poly-2	0.7138	0.68	0.6790
Poly-3	0.8297	0.78	0.7784
RBF	0.9634	0.96	0.9602

Лінійне ядро (Linear) продемонструвало найгірші результати серед усіх протестованих варіантів. Це пояснюється тим, що лінійне ядро ефективно лише для задач із лінійно

роздільними межами між класами. У випадку набору даних make_circles, де класи утворюють концентричні кола, лінійне ядро не може коректно описати розділювальні межі. Як наслідок, точність, повнота та F1-міра для цього ядра залишаються низькими.

Поліноміальне ядро другого порядку (Poly-2) показало значно кращі результати. Воно здатне враховувати дугоподібні межі між класами, що робить його придатним для задач, де межі мають просту криволінійну форму. Однак Poly-2 має обмежену здатність до адаптації, через що його ефективність знижується при обробці більш складних структур даних.

Поліноміальне ядро третього порядку (Poly-3) забезпечило ще кращу адаптацію до складних меж між класами. Це свідчить про те, що збільшення ступеня полінома дозволяє краще враховувати нерівномірний розподіл даних і складніші форми меж. Poly-3 показало помітне зростання точності, повноти та F1-міри, що підтверджує його переваги над ядром Poly-2.

Найвищі результати продемонструвало радіальне базисне ядро (RBF). Завдяки своїй здатності враховувати складні структури даних, RBF забезпечує максимальну точність, повноту та F1-міру серед усіх протестованих ядер. Його ефективність пояснюється тим, що RBF ядро враховує не лише просторове розташування точок, але й їх відстань від центру класифікаційної зони, дозволяючи створити найбільш оптимальну межу між класами. Це робить RBF ідеальним вибором для задач, де класи утворюють складні нелінійні структури, як-от концентричні кола.

4. Висновки

Результати експерименту підтвердили, що метод SVM із лінійним ядром підходить лише для задач, де класи розділені прямою лінією. У задачах з більш складними геометричними формами об’єктів і меж поміж ними, що представлено у наборі даних make_circles, ефективність класифікації значно зростає при використанні нелінійних ядер, таких як поліноміальні або радіальне базисне ядро. Ядро RBF продемонструвало свою особливу перевагу завдяки здатності адаптуватися до складних форм меж та мінімізувати помилки класифікації.

Таким чином, результати дослідження підкреслюють необхідність використання нелінійних ядер для задач із нелінійними межами, особливо у випадках аналізу медичних даних, де точна класифікація є критично важливою для ранньої діагностики та розробки персоналізованих підходів до лікування.

Список використаних джерел

[1] Choong, H.-J.Y., & Lee, C.H. (2017). Medical big data: promise and challenges. Kidney Research and Clinical Practice, 36(1), 3–11. https://doi.org/10.23876/j.krcp.2017.36.1.3.

[2] Кобзєв В.Г., Яковлєв С. В., Назаров О.С., Назарова Н.В., Горішня К.О. Модифікації методу опорних векторів для класифікації та виявлення аномалій у задачах обробки зображень // Застосування інформаційних технологій у підготовці та діяльності сил охорони правопорядку: Збірник тез доповідей Міжнародної науково-практичної конференції. - Харків: НАНГ, 2024. - с. 149-151.

[3] Kobziev V., Yakovlev S., Gorishnia K. Modifications of the support vector method for classification and detection of anomalies in image processing problems // Інформаційні технології та комп’ютерне моделювання; матеріали статей Міжнародної науково-практичної конференції, м. Івано-Франківськ, 21-24 травня 2024 року. – Електронне видання. Івано-Франківськ: Прикарп. нац. ун-т ім. В. Стефаника, 2024. – с. 221-222.

[4] Anwar, M. (2018). Support vector machines (SVM): Linear and non-linear SVMs. Medium.

[5] Géron, A. (2022). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, Inc.

DOI: https://doi.org/10.30837/IST-2024.P1.134

Моделювання взаємодії популяцій хижаків та травоїдів з використання еволюційного підходу

Олександр Олійник[1] та Олена Олійник[1]

[1] *Харківський національний університет радіоелектроніки, пр. Науки, 14, Харків, 61166, Україна*

Анотація

У статті розглядається розробка моделі взаємодії популяцій хижаків і травоїдів для використання в ігрових або симуляційних системах. В основі роботи лежить аналіз класичної моделі Лотки–Вольтерра, яка описує динаміку взаємодії хижаків і жертв. Однак її обмеження, такі як нескінченність ресурсів та відсутність індивідуальної взаємодії, потребували доопрацювання. Запропонована модель враховує обмежені ресурси, взаємодію з окремими членами популяції та можливість адаптації параметрів для досягнення стабільного стану системи. Вона дозволяє моделювати зміну чисельності популяцій у залежності від зовнішніх умов і параметрів, що робить її універсальною для застосування в ігрових середовищах, природоохоронних системах та інших наукових і практичних завданнях.

Ключові слова

модель хижак–жертва, Лотка–Вольтерра, динаміка популяцій, біом, обмежені ресурси, еволюційний алгоритм, симуляція, взаємодія популяцій, ігрові системи, екосистема.

1. Вступ

Під час моделювання нами ігрового всесвіту виникла задача додати в модель взаємодію тварин. Для того щоб не плодити зайві сутності було вирішено зосередитись на двох популяціях: травоїди та хижаки. По задумці травоїди розмножуються до тих пір доки вистачає ресурсів області ігрового поля та змінюють тип біому при досягнені певних умов. Наприклад, якщо кількість травоїдів в луговому біомі перевищує деяку межу біом становиться пустелею, що зменшує кількість ресурсів для травоїдів. Хижаки стримують ріст популяції травоїдів в заданих межах та взаємодіють з гравцем. Більш детально про взаємодію з ігровим полем та гравцем ми плануємо розповісти в іншій роботі, а тут зосередимось саме на взаємодії популяцій між собою.

Взаємодія двох пов'язаних популяцій є доволі складною задачею. Цю задачу різними способами пробували вирішувати починаючи з початку минулого сторіччя. Найбільш відомим рішенням є модель Лотки–Вольтерра хижак–жертва.

2. Модель Лотки–Вольтерра хижак–жертва

Ця модель описує взаємодію двох видів, хижаків що поїдають травоїдів, та травоїдів що поїдають якийсь сторонній ресурс. Модель спирається на декілька припущень [3]:

- Ресурс потрібний травоїдам невичерпний;
- Потреба в їжі хижаків невичерпна;
- Хижаки їдять тільки цих травоїдів;

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
oleksandr.oliinyk1@nure.ua (A. 1); olena.oliinik@nure.ua (A. 2)
0000-0002-7734-3963 (A. 1); 0000-0002-6769-2844 (A. 2)

- Зовнішнє середовище не змінюється і не надає переваг одному з учасників моделі, а адаптація учасників не істотна
- Розподіл популяцій по віку чи простору не впливає на дінаміку.

В цілому, всю модель можна описати за допомогою пари нелінійних диференціальних рівнянь[3]:

$$\frac{dx}{dt} = ax - \beta xy,$$
$$\frac{dy}{dt} = -\gamma y + \delta xy,$$

де x- щільність здобичі, y – щільність хижака, $\frac{dx}{dt}, \frac{dy}{dt}$- миттєві зростання обох популяцій, t – час, a – швидкість росту популяції травоїдів, β- коєфіцієнт впливу хижаків на загибель травоїдів, γ- рівень смертності хижаків, δ – коефіцієнт впливу травоїдів на швидкість росту популяції хижаків.

Важливим фактором цієї моделі є те, що покоління хижаків та травоїдів повинні збігатись.

На графіку чисельність популяцій за цією моделлю буде змінюватись приблизно так:

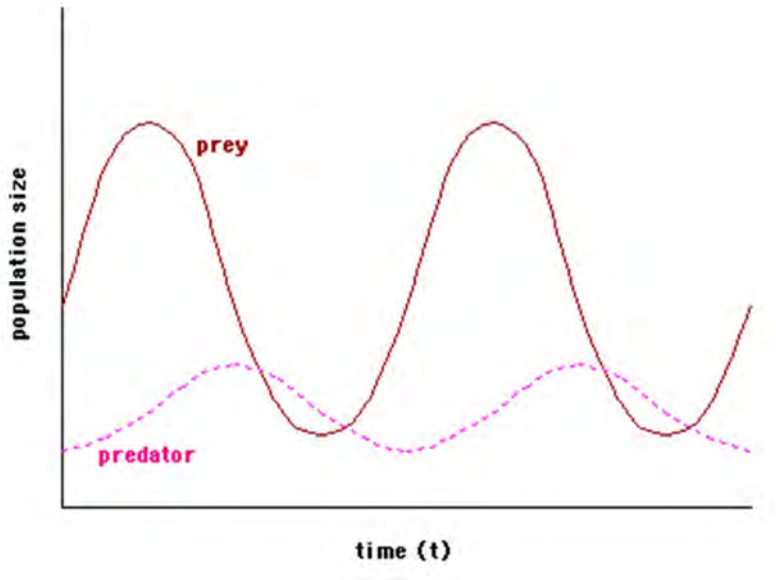


Рисунок 1: Графік зміни чисельності популяцій хижаків та жертв в часі, по моделі Лотки–Вольтерра[3]

Ця модель нам не підходить. По перше по задачі ресурси травоїдів обмежені, відповідно обмежені і розміри популяції хижаків. Ще одна проблема в тому що ця модель не передбачає взаємодії з окремим представником популяції, а нам для інших етапів роботи це потрібно. Але в іншому вона дійсно добре описує взаємодію двох видів і ми можемо спиратись на неї щоб побудувати нашу. Розглянемо дороблену модель.

3. Модель

Наша модель повинна описувати взаємодію двох популяцій приблизно як і попередня, але з можливістю вводу обмежень ресурсів, зовнішньої взаємодії з конкретними членами популяції та обрахунку дій конкретного члена популяції. Також потрібно щоб модель дозволила встановити таке співвідношення між видами, щоб їх чисельність виходила на плато і не призводила до суттєвих викидів. Почнемо з припущень:

- Розмір ресурсів для травоїдів обмежений біомом (для прикладу будемо орієнтуватись на луговий біом, що може прокормити 1000 членів популяції травоїдів);
- Розподіл по віку та полу не впливає на динаміку;
- Покоління починаються в один час;
- Виконання всіх дій відбувається покроково;
- Для приплоду потрібно мінімум два члена популяції;
- За 1 ітерацію хижак з'їдає не більше певної кількості травоїдів;
- Якщо максимальна ємність біому перевищена, тварини починають гинути від голоду.

Для травоїдів доступні наступні дії:

- Якщо травоїдів менше за ємність біому, то кожні два травоїда кожну ітерацію можуть народити від 1 до 3(дуже низька ймовірність) нових травоїдів
- Якщо поріг перевищено, то травоїди починають гинути від голоду цілими групами до 6 членів;
- Зовнішня дія що збільшує або зменшує кількість членів популяції або ємність біому;

Таким чином природним обмеженням чисельності травоїдів є ємність біому та голод. Для хижаків природним обмеженням виступає кількість травоїдів. Для збереження нормальної чисельності потрібно щоб на кожного хижака було мінімум стільки травоїдів скільки йому потрібно для проходження періоду (кроку). Розглянемо доступні для хижаків дії.

- Спіймати та з'їсти до 2 травоїдів.
- Нікого не спіймати, але вижити.
- Нікого не спіймати і вмерти від голоду.
- Якщо знаходиться живих хижаків, то вони можуть народити від 1 до 4 нових хижаків.
- Якщо чисельність хижаків доходить до порогових значень(у нашому випадку чисельності травоїдів поділеної на два), то хижаки зменшують приріст популяції до 1-2 нових хижаків за ітерацію, а також гинути групами до 9 членів.

Всі обчислення відбуваються покроково для кожного члена популяції спочатку рахується чи знайшов він їжу, чи помер він від її недостачі, після цього для кожної пари живих членів обчислюється поява нащадків. Спочатку виконується обчислення для травоїдів, а потім для хижаків. Таким чином система починає еволюційний процесс схрещення та зміни параметрів. У якості модифікатора мутації виступають зовнішні взаємодії.

Всі коефіцієнти(кількість членів в групах, кількість потрібних жертв в сезоні, кількість жертв за 1 охоту, кількість нащадків на одну пару, з групами вірогідностей кожного варіанту) можуть бути налаштовані окремо. Ті що є в прикладі отримані за допомогою еволюційного алгоритму, що шукав параметри що дозволять організувати плато чисельності без великих коливань.

Звісним чином, на модель окрім ємності біому впливають і початкові розміри популяцій. Розглянемо декілька прикладів.

У нас буде три діаграми:

- Чисельність приблизно однакова;

- Травоїдів більше;
- Хижаків більше.

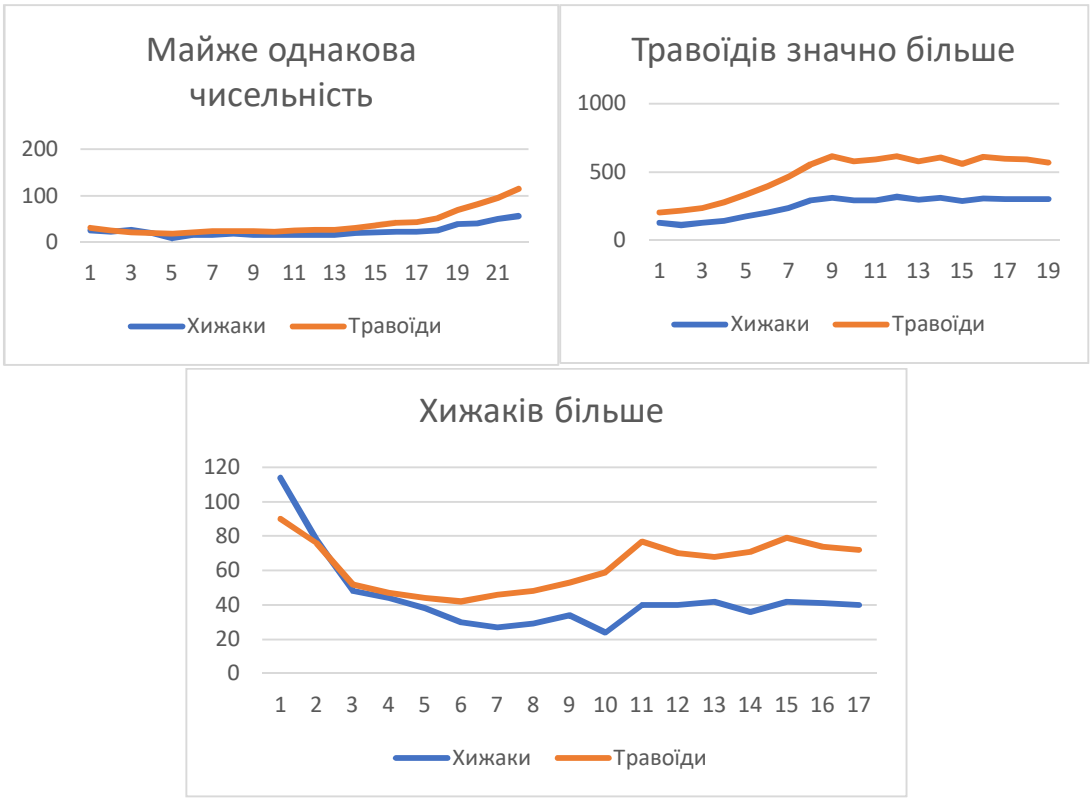


Рисунок 2: Отримані графіки зміни чисельності

4. Висновки

Розроблена модель взаємодії хижаків та жертв що дозволяє окрему взаємодію з кожним членом популяції, враховує розміри кормової бази біома для травоїдів, дозволяє отримати збалансовану на заданому рівні популяцію, та передбачає можливість зовнішніх втручаннь та зміни кормової бази популяції. Модель буде корисна при розробці ігор, де є потреба в організації конкурентних популяцій здатних взаємодіяти без участі гравця, при моделюванні природоохоронних комплексів, що передбачають втручання людини во взаємодію між видами. При певній доробці на реальних прикладах модель може бути корисною при проектуванні протидії інвазійним видам та розробці методів боротьби з «overpopulation».

Література

[1] Lotka, A. J. (1925). Elements of Physical Biology. Williams and Wilkins.
[2] Goel, N. S.; et al. (1971). On the Volterra and Other Nonlinear Models of Interacting Populations. Academic Press. ISBN 0-12-287450-1.
[3] PREDATOR-PREY DYNAMICS: LOTKA-VOLTERRA URL: https://web.archive.org/web/20121215031911/http://www.tiem.utk.edu/~gross/bioed/bealsmodules/predator-prey.html

Секція 7.
Комунікаційні, GRID та хмарні технології. IoT

DOI: https://doi.org/10.30837/IST-2024.P1.139

Порівняння моделей і методів побудови туманних обчислювальних систем для збору та аналізу розподіленої інформації та прийняття рішень

Олександр Сбітнєв[1†] та Людмила Волощук[1†]

[1] *Одеський національний університет ім. І.І.Мечникова, Дворянська 2, Одеса, Україна*

Анотація
У цій роботі проведений аналіз існуючих архітектур туманних комп'ютерних систем, які висвітлюють їхні рівні, функціональні можливості методи побудови та порівняльні переваги.

Ключові слова
Туманні обчислення, Інтернет речей (IoT), архітектури, обробка даних, масштабованість.

1. Вступ

З розвитком Інтернету речей і зростанням потреби в обробці великих обсягів даних у реальному часі, традиційні хмарні обчислення починають демонструвати свої обмеження. Висока затримка, обмежена пропускна здатність і недостатня масштабованість створюють перешкоди для багатьох сучасних додатків, таких як автономні транспортні засоби, розумні міста та промисловий Інтернет речей. У зв'язку з цим виник новий термін— туманні обчислення, вони забезпечують передобробку та аналіз даних ближче до місця їх збору. Туманні обчислення стали важливою парадигмою, що поєднує хмарні сервіси з периферійними пристроями, вирішуючи проблеми, пов'язані з затримкою, обмеженнями пропускної здатності та обробкою даних у реальному часі. [1]

Ця робота присвячена аналізу архітектур туманних обчислень, що розрізняються за своєю структурою, функціональністю та рівнем інтеграції з хмарними і периферійними пристроями. Аналіз архітектур дозволяє визначити їхні переваги та недоліки, а також обрати найбільш підходящий варіант для конкретних IoT-застосувань.

2. Розгляд архітектур

2.1. Три-рівнева Архітектура

Три-рівнева Архітектура є базовою моделлю для туманних обчислень, що розміщує обчислювальні ресурси між хмарою та IoT-пристроями. Вона забезпечує обробку даних, взаємодію та зберігання інформації, отриманої з IoT-пристроїв, зосереджуючись на помірному рівні затримок для сценаріїв із середніми вимогами до швидкості реагування. Рівні включають IoT рівень (збір даних), туманний рівень (первинна обробка), і хмарний рівень (глибока аналітика та зберігання). Така структура дозволяє ефективно обробляти

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
† Ці автори зробили однакові внески.
alexsbitnev99@gmail.com (О. Сбітнєв); lavstumbre@gmail.com (Л. Волощук)
0009-0008-6311-612X (О. Сбітнєв); 0000-0002-2510-0038 (Л. Волощук)

дані з оптимальною пропускною здатністю, але її функціональність обмежена порівняно з більш складними моделями. [2]

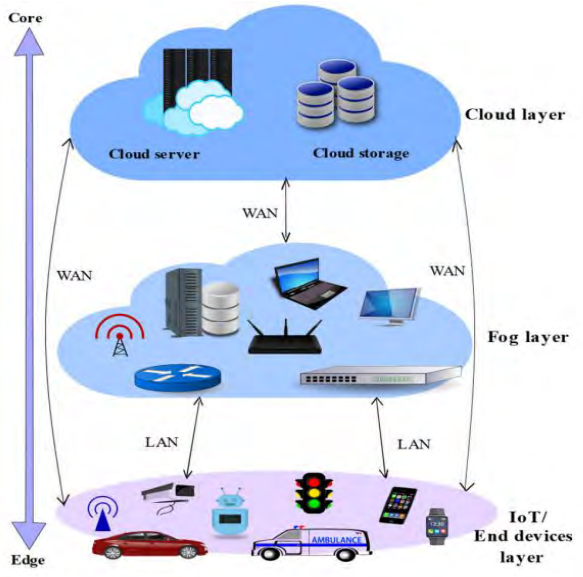


Рисунок 1: Трирівнева архітектура

2.2. N-рівнева Архітектура OpenFog

N-рівнева Архітектура OpenFog пропонує гнучку багаторівневу структуру, де кожен додатковий рівень посилює обчислювальні можливості, безпеку та масштабованість системи. Ноди організовані ієрархічно, причому нижчі рівні зосереджуються на зборі та первинній обробці даних, а вищі — на аналітиці та управлінні ресурсами. Ця архітектура ефективно справляється з обробкою великих обсягів даних і підходить для систем із високими вимогами до обробки та надійності. Такий підхід забезпечує масштабованість та стійкість, але ускладнює управління мережею. [3]

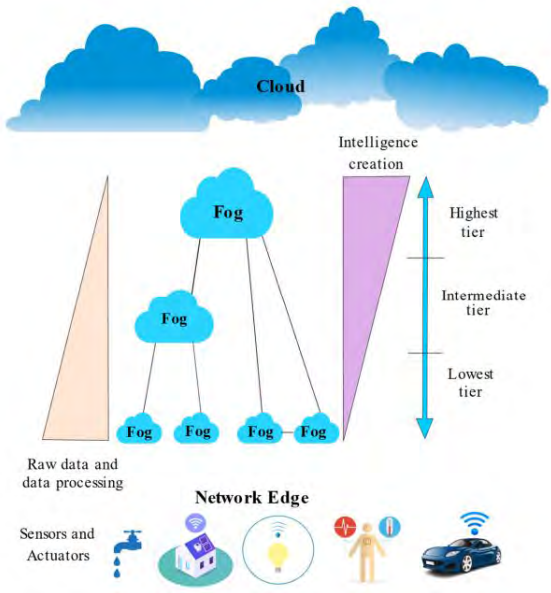


Рисунок 2: N-рівнева Архітектура OpenFog

2.3. Семирівнева Архітектура

Семирівнева Архітектура — комплексна модель для зниження затримок, що розширює хмарні сервіси до IoT-пристроїв. Сім рівнів охоплюють фізичний рівень, туманні пристрої/шлюзи, моніторинг, попередню та постобробку, зберігання, управління ресурсами та безпеку. Кожен рівень виконує чітко визначену роль у процесі обробки даних, моніторингу системи, управління ресурсами та захисту. Така модель підходить для додатків, де затримки є критичними, і забезпечує високий рівень безпеки та управління ресурсами, хоча вимагає більших інфраструктурних витрат. [4]

Рисунок 3: Семирівнева Архітектура

2.4. Багаторівнева архітектура

Найбільш комплексна архітектура складається з дванадцяти рівнів, пропонуючи детальну та складну структуру для туманних обчислювальних систем. Ця архітектура починається з фізичного рівня, який відповідає за збір даних з широкого спектра датчиків та виконавчих механізмів, після чого йде рівень периферійних обчислень, де відбувається первинна обробка даних. Локальна мережа та шлюзові рівні керують комунікацією між периферійними пристроями та широкою мережею, забезпечуючи безперебійний потік даних і підключення. Архітектура також включає рівень зовнішньої мережі, який відповідає за зовнішню комунікацію, та рівень безпеки, що забезпечує надійний захист від потенційних загроз. Крім того, рівень проміжного програмного забезпечення сприяє балансуванню навантаження та управлінню чергою повідомлень, а рівень ETL (Extract, Transform, Load) забезпечує належну обробку та зберігання даних. Рівень великих даних та аналітики використовує алгоритми штучного інтелекту та машинного навчання для просунутої аналітики, а рівні повідомлень та представлення керують оповіщеннями в реальному часі та взаємодією з інтерфейсом користувача. Нарешті, рівень конфігурації забезпечує управління конфігурацією та станом периферійних пристроїв, гарантуючи, що оновлення та зміни ефективно впроваджуються. [5]

3. Порівняльний аналіз

Порівняльний аналіз чотирьох архітектур туманних обчислень базується на таких ключових критеріях, як продуктивність, затримка, гнучкість, безпека та можливості обробки даних.

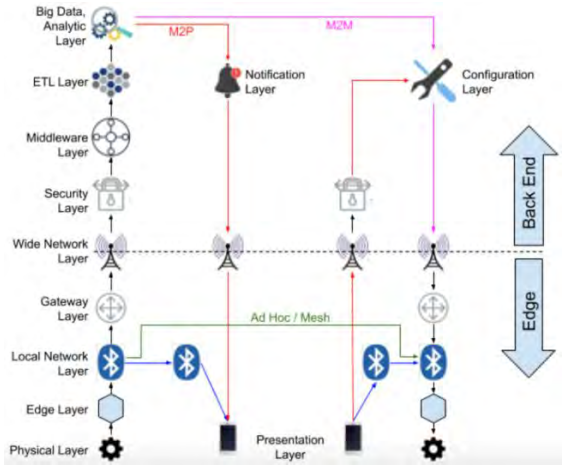


Рисунок 4: Багаторівнева архітектура

Чотири архітектури туманних обчислень відрізняються за складністю, що визначає їхню продуктивність, затримки та можливості обробки даних. Три- та семирівневі моделі є збалансованими рішеннями для швидкої обробки в реальному часі з помірними вимогами до безпеки та ресурсів, що робить їх придатними для простих IoT-додатків. Архітектура N-рівня підходить для сценаріїв із більшим навантаженням, забезпечуючи стійкість і масштабованість, але залишається менш потужною у порівнянні з дванадцятирівневою моделлю. Найбільш складна, дванадцятирівнева архітектура, підтримує розширені функції безпеки та аналітики, дозволяючи працювати з великими обсягами даних, що ідеально підходить для інтеграції різних компонентів у складних середовищах, але вимагає значних ресурсів для налаштування та підтримки. Отже, вибір архітектури залежить не лише від розмірів розгортання, але й від вимог до безпеки, гнучкості та здатності обробляти великі обсяги даних. Складніші архітектури надають більше можливостей для адаптації та інтеграції, тоді як простіші рішення краще підходять для критичних за часом додатків із меншими вимогами до ресурсів.

Література

[1] S. N. Srirama (2023), A Decade of Research in Fog computing: Relevance, Challenges, and Future Directions, Journal of Software: Practice and Experience, 2023;

[2] Hu, P.; Dhelim, S.; Ning, H.; Qiu, T. Survey on Fog Computing: Architecture, Key Technologies, Applications and Open Issues. J. Netw. Comput. Appl. 2017, 98, 27–42, https://doi.org/10.1016/j.jnca.2017.09.002.

[3] IEEE Standard 1934–2018; IEEE Standard for Adoption of OpenFog Reference Architecture for Fog Computing. 2018 ; pp. 1–176

[4] Angel, N.A.; Ravindran, D.; Vincent, P.M.D.R.; Srinivasan, K.; Hu, Y.-C. Recent Advances in Evolving Computing Paradigms: Cloud, Edge, and Fog Technologies. Sensors 2022, 22, 196

[5] Gayratov, Z. K., Kilichov, J. R., & Toshpulatov, A. (2022). Basic definitions of twelve layer IoT architecture for smart city. In International Scientific and Technical Conference: Digital Technologies: Problems and Solutions of Practical Implementation in the Industry.

Секція 8.
Захист інформації. Інформаційна безпека

DOI: https://doi.org/10.30837/IST-2024.P1.144

Enhancing Intrusion Detection Systems through Advanced Feature Engineering and Machine Learning Techniques

Aditya Patel
pateladij1201@gmail.com

Abstract*—In this study, I explore advanced methodologies for improving intrusion detection systems (IDS) by leveraging machine learning algorithms and sophisticated feature engineering. I initially assess the performance of various classification algorithms, including k-nearest neighbors (KNN) and support vector machines (SVM), on the ADFA-LD 12 dataset. My analysis reveals that SVMs outperform other methods, particularly when using two-sequence feature spaces instead of traditional frequency-based approaches. I conduct a detailed evaluation of SVM performance with linear and sigmoid kernels, revealing that the two-sequence feature space significantly enhances detection accuracy. Recursive feature elimination demonstrates that optimal performance can be achieved with fewer than 240 features, underscoring the importance of effective feature selection. In contrast, one-class SVMs, used for outlier detection, show comparatively poor performance, indicating that traditional outlier detection methods may not be as effective in this context. My findings highlight the efficacy of SVM classifiers in both two-sequence and frequency feature spaces, though the latter does not offer substantial improvement. The study also emphasizes the need for further research into scalable algorithms and improved kernel methods for unsupervised outlier detection. Future work should focus on integrating domain-specific knowledge for feature engineering and exploring modern datasets to validate and enhance these findings.

I. Introduction

In the industry I are currently experiencing what many people call the fourth industrial revolution. The main characteristics of this disruptive process are:

- The ubiquitous presence of connected devices, collectively called Internet of Everything.
- The ability of cyber system to interact with and affect the physical world creating the so called cyber-physical systems (CPS).
- The growth of distributed computing and the capabilities that it allows.
- The evolution of Machine Learning enabling cyber-physical systems with increased autonomy.[1]

One common denominator of the aspects of the fourth industrial revolution is increased connectivity of computing devices. This brings the negative side effect that more facets of my life and society are exposed online making cyber-security even more important. During the last couple of years I have seen many examples of this danger.

[1]A good example of an autonomous cyber-physical system are Tesla's self-driving cars.

Further expanding on the example of cars I mention that on the summer of 2015 two american security researchers demonstrated that a contemporary model Jeep Cherokee could be remotely accessed maliciously over mobile telephony network.[2] As demonstrated in the afforementioned article[2] this has resulted in the academia, the industry and regulators considering policy decisions for regulation and standards adoption in order for car safety to keep up with technological innovation. Another significant event is the attack on a regional electricity distribution company, Ukrainian Kyivoblenergo, on 23 December 2015. The attack compromised the company's computer systems and their supervisory control and data acquisition systems (SCADA). It resulted in power outages for around 225,000 customers for a period of hours. The diligence of the company employees and the quick transition on manual mode allowed the company to restore its service with little delay. The attack is thought to have been carried out by a state actor and could have inflicted more damage had the perpetrators more time before making their pressence known. Again this attack has attracted a lot of attention among security researchers and the industry in order to formulate appropriate policies to protect core infrastructure[3].

Now that I have seen examples of the increasing importance of cyber - security let's look deeper at what it actually is. Cyber security generally consists of two parts. Intrusion Prevention Systems (IPS) and Intrusion Detection Systems (IDS). Put it simply an IPS prevents the attacker from getting in and an IDS detects him once he is in. With the emergence of cloud computing the boundaries between the two are getting blurred[4]. Nowadays modern IDS have a wide range of features including the detection of an occuring attack, automated responses and detection of malicious behavior within the system[4].

II. Foundations of the project

In this project I are drawing knowledge from two related but distinct fields. [2] The first field is outlier and anomaly detection techniques from Machine Learning. The second field is Intrusion Detection techniques within deployed cyber-security solutions. On the next sections I will describe the foundations of this project. They have been studied to a big extend during the Project preparation course.

[2]With info from:
Project Preparation: Outlier Detection in Cybersecurity Application, 2016, submitted by Nikolaos Perrakis.

III. Outlier and Anomaly Detection

The concept of bad data has been quite old in the field of statistics. One of the first systematic approaches to dealing with them, dating back to the 19^{th} century, has been Chauvenet's criterion on whether one data point of an experimental data set was spurious or not. Moving on to modern machine learning techniques on outlier and anomaly detection I see that there are many approaches. They are based on different statistical methods but also on different characteristics on the domain from which the dataset comes. Below are most relevant methods with regards to my work.

The first method has been put forward by Roberts et al.[5]. They use a Gaussian Mixture model to map the normal behavior of the system their dataset describes. This creates a set of normal system states their trained model recognizes. In their method, they minimize the set of heuristically chosen parameters used in such techniques by using an evolving threshold on how much a data point can differ from a learned normal state before it gets classified as belonging to another state. When I am training our model with normal operation data, this results in a newly learned normal state. Afterward, the same threshold is used to identify an anomaly in the test data. They then test their method in a set of electroencephalograms (EEG) obtained from patient data. They prove the robustness of their method by successfully identifying epileptic seizures when they occur. They do not quantify the success of their results by mentioning accuracy, precision or fall out rates. However they include results from particular examples of EEG activity and explain their performance metrics result comparing them with input from domain related knowledge.

A pioneering approach in unsupervised approaches for anomaly detection came from Schölkopf et al.[6]. They propose an algorithm that estimates the region where the probability density of some given data lives. New data can be compared to this density for anomaly detection purposes. They describe their method in detail and describe it's conceptual limitations. Then they demonstrate it's characteristics in an artificial dataset and it's effectiveness in a real dataset. Their algorithm has come to be known as One class Support Vector Machine algorithm and it has been a very influential algorithm in the field of outlier detection.

Hayton et al.[7] have written another influential paper. They studied Support Vector machine-based anomaly detection in Jet engine vibration spectra. Their approach also adds the feature of combining a second dataset in order to train their model more accurately. Moreover, during the discussion of their results, Hayton et al. gave a very insightful discussion on whether it is preferable to address novelty detection as a 2-class classification problem or not. They argue that when the "abnormal" data points are representative of the abnormal class, then training a model as a binary classification problem is optimum. On the other hand, they point out that the novel data points may be artificial and that their nature may be non-stationary. In domains with those characteristics, it is better to treat only the normal behavior data points as representative of their class. This is also the case in the Information Technology domain, which is why deployed Intrusion Detection Systems use the latter approach.

Clifton et al.[8] are a team of researchers that use class Support Vector Machines for anomaly detection. study luminosity measurements of a Typhoon G30 combustor engine and create an SVM model for anomaly detection. They use wavelet analysis on the multi-channel combustion data to create their feature space and subsequently perform supervised learning to train their SVM to recognize anomalous behavior. Clifton et al. also compare the results of the SVM approach to the GMM approach on the same dataset. They found that SVM-based anomaly detection performed better and demonstrated that with instances of multi-channel combustion data where the SVM model identified the novelty at earlier times compared to the GMM model.

Another interesting approach on the subject touches on the issue of the nature of the novel data points being non-stationary. Farran et al.[9] study the KDD 1999 Cup dataset[3]. Their technique uses two steps. The first uses a method called Voted Spheres, which runs over the dataset once and performs non-parametric classification. It results in partitioning the feature space into a series of overlapping spheres. The second step is to take into account the possibility that the test set may not come from the same distribution as the training set. They use two algorithms, important weighting, and kernel mean matching, to account for that difference. It is useful to mention the author's note that kernel mean matching does not scale very well as the dataset increases due to its reliance on quadratic programming, something that may negate the low computation load of the Voted Spheres method.

IV. Intrusion Detection Systems

Bhuyan et al.[10] give the following description: "Intrusion is a set of actions aimed to compromise the security of computer and network components in terms of confidentiality, integrity, and availability". Intrusion Detection Systems are my answer to that threat. The basic assumption I make is that my system will behave differently during an intrusion than during normal operation. And this is why it is appropriate to use novelty detection in IDS applications.

There are many attack actions against a computer system that can be classified as intrusions with the above definition. Each of them has different specific characteristics. Thus it is helpful for us to divide attacks into certain classes. As I did during Project Preparation[4] I will use the classification used in Bhuyan et al.[10] and Ahmed et al.[11]. I classify intrusions as Malware attacks, Denial of Service attacks, Network attacks, Physical attacks, Password Attacks, Information Gathering Attacks, Remote to User attacks, and Root attacks. Another important property of intrusions is that they have varying anomaly characteristics. I, therefore, classify them in

[3]DARPA IDS 1998 - 1999 Datasets: https://www.ll.mit.edu/ideval/data/ - Accessed at 9 Aug 2016

[4]Project Preparation: Outlier Detection in Cybersecurity Application, 2016, submitted by Nikolaos Perrakis.

an additional way. There are *point anomalies* when a data point is considered anomalous when it is distant in comparison to the points that model the normal behavior of the system. This type of intrusion closely fits the machine learning problem of anomaly detection, and it will be the main focus of my work. There are also *contextual anomalies* when a data point may be considered anomalous according to the context it is associated with, as well as its place in the feature space. The context may be additional features engineered in my feature space or an evaluation of circumstances associated with the data point from the IDS (or its operator). Lastly, there are *collective anomalies* for which a collection of data is considered anomalous, but each data point, taken individually, is not. In Table I[5], I describe the eight classes of attacks, give examples of them, and classify the examples with their respective type of anomalies.

An operational enterprise IDS solution consists of main parts. Two main parts are a Network Intrusion Detection System (NIDS) which analyses traffic over a company's network to detect intrusion and a Host-based Intrusion Detection System (HIDS) which is installed on the company's computers (hosts) and monitors them in order to detect intrusion.

There are many techniques used to identify anomalies on an IDS. Bhuyan et al.[10] do a good work on categorizing them, and I will follow their classification scheme in presenting them. It should be noted, however, that these methods are not completely distinct from each other, and a particular implementation may include aspects from more than one. The first class is statistical methods, which also includes Bayesian Networks. A model is trained to "learn" the normal operation of the system. A threshold of statistical distance from normal operation is determined, and data points exceeding that threshold are classified as anomalous. For example, Krueger et al.[12] used Bayesian Networks as part of an IDS system and managed to reduce the false alarm rates compared to threshold-based approaches. The second class is clustering methods. A distance or similarity method has to be defined on the feature space. It is then used to cluster the data with various algorithms, such as k-means clustering. I then measure the distance of new data points with my learned clusters And use it to classify them as anomalous or not. One example of such a method is the work of Bhuyan et al.[13], where they create a reference point clustering method to perform outlier detection that works well on large datasets. It is useful to note here that a clustering method, such as k-means clustering, conceptually is quite similar to a Gaussian Mixture model. Such similarities exist for various algorithms that, under different implementations, can be labeled as different classes in my classification scheme. The third class is classification methods. They include both supervised and unsupervised algorithms. One good example that I will use later is support vector machines (SVM). They can be trained either on pre-labeled data or on non-labeled data, depending on the SVM implementation. Another useful example of such techniques is Support Vector Data Description, a one-class classification method described and further extended by Kang et al.[14]. The fourth class is knowledge-based methods. They take advantage of what has been learned from previous attacks so that they are able to identify them when they re-appear. These methods include ontology, signature, and logic-based approaches. One example of a knowledge-based method is Xu 's[15] work. In it, he introduces a sequential anomaly detection method that takes advantage of a Markov reward process in a reinforced learning approach. The fifth class is soft computing methods. The key point behind them is that if finding the exact solution is not feasible, then I can look for approximate solutions. These methods include Artificial Neural Networks, Rough Sets, Fuzzy Sets, Ant Colony Algorithms, Artificial Immune Systems, and Genetic Algorithms. I mention as an example the work of Amini et al.[16], where they use adaptive resonance theory (ART) and self-organizing maps (SOM) unsupervised neural networks for real-time intrusion detection. The last class is called combination learner methods and basically consists of techniques that combine more than one of the previous classes in their implementation. A good example of such a method is the work of Borji[17]. He uses my classifiers, decision trees, SVM's, ANN, and kNN, which he combines with three different strategies: majority voting, belief measure, and Bayesian averaging.

The variety of methods for Intrusion detection can be attributed to their diverse strengths and weaknesses. Additionally, the intrusions each IT infrastructure faces are different, which means that different IT networks are best suited to different Intrusion Detection methods. For example, clustering and Nearest neighbor algorithms have poor performance on high dimensional data because, in those cases, distance methods cannot accurately differentiate between anomalous and normal data points. As an example, this problem is studied by Aggarwal et al.[18], where they mention that the meaning of proximity in high dimensions is problematic. They study the problem empirically, focusing on L_k norms, and propose fractional k's as a potential solution, though I will not try them here. Other ways to overcome this are feature reduction techniques, such as spectral techniques or principal component analysis, but he has to be careful to maintain the separation between anomalous and normal data points. In general, classification techniques suffer because of the need to pre-label data points. Creating those labels is hard and often artificial. This results in the subsequently trained model becoming outdated quite fast because of the fast-paced evolution of the IT field. Using partially labeled data for semi-supervised clustering techniques can be more efficient than classification methods, provided I have a feature space with a good distance measure. If not, statistical techniques are a better option. As I mentioned, another point to consider is how easy it is to update my application. Statistical, clustering, and classification methods are all hard to train. But it can be done offline, and their testing phase is fast. Last but not least, there are times when the assumption that intrusion events

[5]Table taken from Project Preparation: Outlier Detection in Cybersecurity Application, 2016, Nikolaos Perrakis

Table I
TYPES OF ATTACKS AND ASSOCIATED ANOMALIES.

Attack Type	Description	Examples (Anomaly Type)
Malware	Virus, Worm, Troyan: A program that may replicate and transfer on its own and perform harmful operations on the infected computer.	Stuxnet Worm (point)
Denial of Service	Attacks that make Network resources inaccessible.	Smurf (collective)
Network	Compromising the security of a Network by exploiting Network protocols.	Man-in-the-Middle (point)
Physical	Compromising a system or a network through physical access.	Evil Maid (point)
Password	Trying to find a user's password, usually through multiple login attempts.	Dictionary attack (collective)
Information Gathering	gathering information to try to find vulnerabilities in a network.	Port scan (collective)
Remote to User	Trying to get remote access as a user to a system.	phf (point)
User to Root	Upgrading a user's privilege to superuser.	Rootkit (point)

are rare compared to normal events is not true. In that case, an intrusion detection method may end up with a high false positive rate. This is a big problem because it interrupts the user's normal operation.

As I see Intrusion Detection is a very complex problem. Each IT system has different characteristics and they are better addressed by different techniques. This means that addressing Intrusion Detection to it's entirety goes well far beyond the scope of this project. Therefore I will limit ourselves to a particular case.

V. DATASETS

A critical part of creating an IDS application is the data you use to train it. In the enterprise world, you have access to the company's data. In academia, however, good datasets for IDS are hard to find. One of the early attempts to solve that problem was initiated by DARPA, and it resulted in the DARPA 1999 IDS Dataset[6]. The dataset was a result of the work done by Stolfo et al.[19]. Even though the DARPA 1999 dataset has been a standard in academia for more than a decade, it has been heavily criticized as well. McHugh[20] has criticized the procedure used to generate the dataset and, more specifically, the validation process used for the validation set. Mahoney et al.[21] have found artifacts of the simulation used to create the datasets within the data undermining the credibility of performance results from intrusion detection algorithms. Moreover Ahmed et al[22] argue that the software used to create the dataset is no longer relevant and even at it's time it didn't have a significant market share. Lastly, the documentation of the dataset is questioned, with some researchers suggesting that the number of attacks present in the dataset is inaccurate.

Over time, other datasets started to emerge, but none has managed to become a standard. One reason for this is that it is difficult to create a dataset that maintains the quality standards needed to not be criticized for any of the shortcomings the DARPA datasets have. This is the case for the dataset I will be using as well. Another reason is that the diverse needs of an IDS system mean that different datasets address different aspects of the IDS landscape. The dataset used in this project addresses host-based intrusion detection (HIDS). It is called ADFA LD-12 and it was presented by Creech et al.[1]. I will describe it in detail in the next chapter.

[6]DARPA IDS 1998 - 1999 Datasets: https://www.ll.mit.edu/ideval/data/ - Accessed at 9 Aug 2016

VI. PROJECT PLANNING

An important part of any project is time management. In order to decide how to manage my time, I first had to see the situation I was in and the goals I wanted to achieve. While the MSc project supervisor has done previous work in the field of intrusion detection, he and his research team are not currently working on it. Therefore, part of the work of this project is to understand recent developments and trends in this field as well as lay the groundwork for further research in the field. This means that my work will not be focused on only one approach, but I will try many approaches to enhance my wider understanding of the subject. That doesn't mean I shouldn't have a specific goal, and this was to employ unsupervised learning outlier detection algorithms. There were also weekly meetings with the supervisor to report on the progress done and decide on the best way forward. Some of the meetings were with the wider research team, where I each presented the work I was doing to the other members.

To assist us in better management, I created the Gantt Chart shown in figure VI. It is separated in eight parts. Initially, I explored the dataset, set up my system, and decided on the architecture of my workflow. After that, I worked on replicating the results of previous papers[23], [24]. This work overlapped with the work on frequency-based algorithms and 2-sequence-based algorithms. Once I replicated the work of previous papers, I worked on new ways to analyze the dataset. One way of doing so was to get out of the boundaries established by the computer science community in handling the dataset and using it in ways more conventional to the machine learning community. After that, I proceeded to write the initial report and deliver a draft to the first supervisor. Subsequently, I used his feedback to improve my analysis and my report. Lastly, I prepared a presentation in order to present my work to the second supervisor.

An overall view of the velocity of my work can be seen in figure 2. It includes both my programming work done in Python as well as the writing of this report and presentation that were written in LaTeX. It does not include time spent

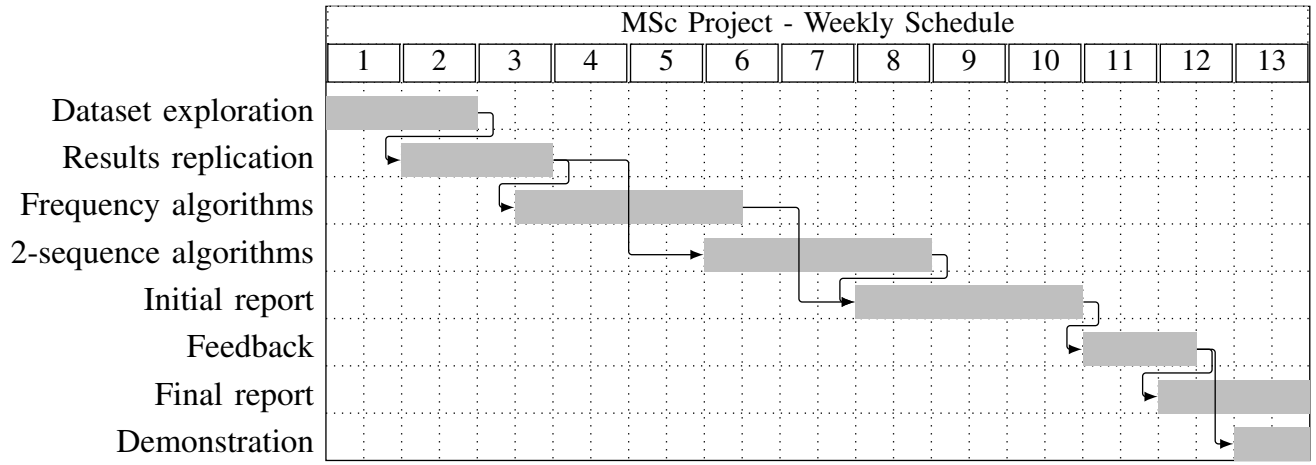


Figure 1. MSc Project Gantt Chart

in work outside of those two, such as computation time and literature review.

VII. Dataset Documentation

The dataset I will use for this project is a modern dataset. It was presented at the 2013 IEEE Wireless Communications and Networking Conference[1]. It uses modern software that simulates real user cases in the IT landscape. The data was produced by Creech et al.[1] by the following described below. Given their goal to simulate a typical real-world case, they used a common architectural configuration used in web servers called LAMP stack. LAMP means Linux, Apache, MySQL, and PHP, after the names of the core software this approach uses. Consequently, Ubuntu 11.04 is used as the server operating system for the creation of the dataset. In order to enable intrusion from the internet, Apache v2.2.17 and PHP v5.3.5 were also installed, and lastly, MySQL v14.14. Apart from that, file transfer protocol (FTP) secure shell (SSH) was enabled in order to simulate the remote administration of the server and the attack surface that comes with it. Their default configuration settings were used. Lastly, a web-based collaborative tool, Tiki Wiki v8.1, was installed and enabled. This version has a documented vulnurenability[7] that allows web exploitation. Those settings are representative of a local server offering basic web services on the internet. Tiki Wiki's known vulnerability represents the ever-present danger of previously unknown vulnerabilities on up-to-date software running on production infrastructure.

A program called *auditd* was used to monitor kernel system call traces. System call traces are API's provided by the kernel of the operating system for userspace applications to access. Ubuntu 11.04 uses the Linux kernel version 2.6.38, which has 325 system calls available[23]. The researchers who created the dataset used 6 different types of attacks to compromise their server. They are presented in table II which was taken from [1]. The dataset is separated into three sets: the training set, the attack set, and the validation set. The training and the validation sets contain sequences of system calls from the normal operation of the system, and the attack sets contain the intrusions performed by the creators of the dataset. Each individual attack method is carried out ten times. Each data point of the dataset consists of a series of system calls. Table III shows the number of data points per subset of ADFA-LD 12.

[7]Packet Storm: All things security, https://packetstormsecurity.com/files/108036/INFOSERVE-ADV2011-07.txt, Accessed 9 August 2016.

Payload/Effect	Vector
Password brute force	ftp by hydra
Password brute force	ssh by hydra
Add new superuser	Client side poison executable
Java-based interpreter	Tiki Wiki Vulnerability exploit
Linux meterpreter payload	Client side poison executable
C100 Webshell	Php remote file inclusion vulnerability

Table II
Attacks used in ADFA-LD 12 dataset.

Subset	Data points
Training	833
Validation	4372
Attack	719

Table III
Subsets of ADFA-LD 12 dataset.

The traditional approach on the cyber security field is to use the training set to train a model, use the attack set to find it's accuracy and use the validation set to find it's false positive rate. While I will use that schema I will not limit ourselves to it.

VIII. Preprocessing and Visualization

The first step in processing my dataset is to download it[8]. Subsequently I unzip the files I see that I create 3 folders for each subset of the dataset. The training and validation set have a long list of text files in their respective folders. Each text file represents a data point and contains a series of integers separated by white spaces that correspond to kernel system calls. The attack set is structured in folders according to the type of attack, and within them are the text files describing the attack data points.

The structure of the dataset after it is unzipped is not so helpful, so I transformed it into a format that will enable us to process it more easily. Some of the standard tools for this purple are pickle and numpy[9]. They are Python libraries that are used to save Python objects. My first preprocessing step is to create a Python dictionary for each of the subsets containing the information about the sequence of system calls for each data point. Then, I use pickle to save that object. Through this process, I create the following three pickle files:

1_training.p , 1_attack.p , 1_validation.p

After that the unzipped dataset files are deleted - they were too inefficient to use. The 3 new files are used to load the dataset for further computations.

my next step is to do a basic exploration of my dataset. The kernel of the host operating system provides 325 system calls, but I am not sure if all are represented and how that representation is distributed. As a first step, I run through my dataset once and create a set[10] where I add all the system calls I encounter. Thus, I end up with the following list of 175 system calls present in my dataset:
[1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 19, 20, 21, 22, 26, 27, 30, 33, 37, 38, 39, 40, 41, 42, 43, 45, 54, 57, 60, 61, 63, 64, 65, 66, 75, 77, 78, 79, 83, 85, 90, 91, 93, 94, 96, 97, 99, 102, 104, 110, 111, 114, 116, 117, 118, 119, 120, 122, 124, 125, 128, 132, 133, 136, 140, 141, 142, 143, 144, 146, 148, 150, 151, 154, 155, 156, 157, 158, 159, 160, 162, 163, 168, 172, 173, 174, 175, 176, 177, 179, 180, 181, 183, 184, 185, 186, 187, 190, 191, 192, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 219, 220, 221, 224, 226, 228, 229, 230, 231, 233, 234, 240, 242, 243, 252, 254, 255, 256, 258, 259, 260, 264, 265, 266, 268, 269, 270, 272, 289, 292, 293, 295, 296, 298, 300, 301, 306, 307, 308, 309, 311, 314, 320, 322, 324, 328, 331, 332, 340]
A curious thing to mention here is the presence of integers higher than 325. One would think that since I only have 325 system calls, only integers up to 325 would be used.[11] This is a very good example that no assumptions should be made when addressing a dataset and that I should always clean and validate the form of the data I expect to have for the next step of my pipeline. Another notable issue here is that I had not identified this problem initially when I was modeling the dataset. However, numerical tests performed after constructing the system called frequency feature space, about which I talked more later, showed that I had not captured all the dataset. This resulted in more rigorous exploration that revealed the unexpected labeling of the system calls.

Moving forward, I count the presence of each system call in my dataset. The result of this computation is presented on figures 3 and 4. As I see, I had to use a logarithmic scale for my y-axis because of how unevenly and spread out the distribution of my system calls is. Moreover, I observe that the norm is that system calls are present in all three subsets, hinting that there is little room for associating particular system calls with malicious behavior.

Another way to improve my basic understanding of my dataset is to look at its principal components. In order to compute the principal components, I merge the whole dataset. As I can see in figures 5 and 6 I have used different colors for the three subsets of ADFA-LD 12. In the figure 5, I only plot the training and attack sets. In figure 6 I also plot the validation set. By observing figure 5 I can see some degree of differentiation between the distributions of the attack and training set. However, once I add the validation set, in figure 6, I see that there is little differentiation between the attack dataset and the training and validation sets combined. The validation set's spatial distribution on the first two principal components is a bit different than the training set's. This means that the training and the validation sets as provided do not carry the same information content.

I can see how much of the variance of the dataset is captured with each principal component of the complete frequency space in table IV. It is worthwhile to mention some specifics about the variance of my datasets. In table IV, I present the variance of the complete dataset in the complete frequency feature space. In that case, one needs 12 dimensions to capture 80% of the variance of the dataset. However, in the work of Xie et al.[23] that I will reproduce in the next chapter, they perform principal component analysis on the training set only. This results in them needing only 9 dimensions in order to capture 80% of the variance of the training set. While I consider that choice sub-optimum because I want my principal components to contain the diversity of the attack dataset in order for a classifier to take advantage of it, I will use their approach in order to be able to compare my results with them. The percentage of the variance of the training set explained in each principal component, in this case, is presented in table V. In both vases I can see that the variance explained by each subsequent principal component decays slowly.

[8]Download url: https://www.unsw.adfa.edu.au/australian-centre-for-cyber-security/cybersecurity/ADFA-IDS-Datasets/

[9]Pickle is the standard library to save python objects. However, when, later, I use numpy objects, it is better to save them with numpy instead of pickle as it is more efficient.

[10]A set, in Python, is an object with the useful property that it does not allow duplicate items.

[11]Nothing in the documentation of the dataset indicated this would (or would not) happen.

Principal Component	Explained Variance
1	0.2042
2	0.1233
3	0.0949
4	0.0777
5	0.0646
6	0.0490
7	0.0386
8	0.0346
9	0.0328
10	0.0303
11	0.0273
12	0.0260

Table IV
FRACTION OF EXPLAINED VARIANCE FROM THE PRINCIPAL COMPONENTS OF THE COMPLETE FREQUENCY SPACE THAT CONTAIN 80% OF THE VARIANCE OF ADFA-LD 12 DATASET.

Principal Component	Explained Variance
1	0.1969
2	0.1548
3	0.1130
4	0.0965
5	0.0709
6	0.0689
7	0.0517
8	0.0380
9	0.0293

Table V
VARIANCE OF THE PRINCIPAL COMPONENTS OF THE TRAINING SET.

IX. FREQUENCY ANALYSIS OF THE ADFA-LD 12 DATASET

I now proceed deeper in my analysis of the ADFA-LD 12 dataset. My first goal is to replicate the results produced by Xie et al.[23]. The paper focuses on frequency-based feature engineering, which is similar to the one I used to explore my dataset. Their analysis is based on two algorithms: k-nearest neighbors and k-means clustering. As I mentioned, a noteworthy part of [23] is that the authors perform principal component analysis not on the whole of the dataset but only on the training set. I will follow their approach while I am trying to replicate their results in the following sections, even though this means my reduced frequency feature space does not capture the dataset as efficiently as it could. This way ensures that I can verify that my replication has been successful. After that, I can move into my own approach.

X. K - NEAREST NEIGHBOURS

I now describe the kNN implementation of Xie et al.[23], which I will replicate.

They begin by performing feature reduction through principal component analysis. The criterion for classifying a data point as normal or not is whether it has more than k data points from the training set within a distance d. The parameters k and d were chosen empirically by the authors. For k they chose 20. Their choice of parameter d depends on the distance metric used. I present their choices[23] in table VI. The change in distance parameter d allows us to calibrate the sensitivity when identifying a data point as abnormal. The bigger the distance, the more likely I am to find k points and call that data point normal. This calibration procedure allows us to construct the ROC curves. After performing my computation, I plot the receiver operating characteristic curves that have also been plotted by Xie et al.[23][12]. I can see them in figures 7 and 8.

Metric	distance (d)	step width
squared euclidean	[0.01, 0.1]	0.01
standardised squared euclidean	[1, 10]	1

Table VI
K-NEAREST NEIGHBOURS PARAMETERS USED IN DIFFERENT ITERATIONS OF THE ALGORITHM IN ORDER TO CREATE THE ROC CURVES FOR KNN ALGORITHM IN THE REDUCED FREQUENCY FEATURE SPACE.

Attack Used	Area under ROC curve
adduser	0.745
hydra ftp	0.593
hydra ssh	0.549
java meterpreter	0.727
meterpreter	0.710
web shell	0.734

Table VII
AREA UNDER THE CURVE FOR SQUARED EUCLIDEAN DISTANCE KNN CLASSIFIER IN THE REDUCED FREQUENCY SPACE.

To assess the performance of my classifiers, I compute the area under the ROC curve. The results are shown in table VII. Calculating the area under the ROC curve for kNN under squared standardized Euclidean distance, I get the results shown in table VIII.

I see that the password cracking attacks (hydra ssh and hydra ftp) are quite harder to identify compared to the other attacks, and my classifier performed poorly on them. For the other attacks, my classifier performed moderately to well overall.

XI. K - MEANS CLUSTERING

I proceed in trying to replicate the results of Xie et al.[23] with regards to the k-means clustering algorithm. I focus on implementing the algorithm with a Euclidean distance metric. The authors used the training data to create 5 clusters. The number of clusters was chosen empirically. A new data point

[12]I should note here that I modified the procedure a bit. Xie et al[23] in their figures do not include the 100% False Positive and True Positive point as well as the 0% True Positive and False Positive point. Thus, they cannot move forward with computing the area under the ROC curve. I added this trivial step in my implementation in order to have a measure that I can use to evaluate and compare results from different algorithms.

Attack Used	Area under ROC curve
adduser	0.696
hydra ftp	0.574
hydra ssh	0.518
java meterpreter	0.689
meterpreter	0.705
web shell	0.697

Table VIII
AREA UNDER THE CURVE FOR SQUARED STANDARDIZED EUCLIDEAN DISTANCE KNN CLASSIFIER IN THE REDUCED FREQUENCY SPACE.

Attack Used	Area under ROC curve
adduser	0.6893
hydra ftp	0.6428
hydra ssh	0.4690
java meterpreter	0.6858
meterpreter	0.7475
web shell	0.7158

Table IX
AREA UNDER THE CURVE FOR K-MEANS CLUSTERING CLASSIFIER USING EUCLIDEAN DISTANCE ON THE REDUCED FREQUENCY FEATURE SPACE.

Attack Used	Regularisation parameter
adduser	10
hydra ftp	0.05
hydra ssh	0.5
java meterpreter	0.1
meterpreter	10
web shell	1

Table X
OPTIMAL REGULARISATION PARAMETER FOR SVM'S WITH LINEAR KERNEL ON ADFA-LD 12 FOR DIFFERENT ATTACK PROFILES IN THE REDUCED FEATURE SPACE.

Attack Used	Area under ROC curve
adduser	0.8303
hydra ftp	0.6978
hydra ssh	0.7801
java meterpreter	0.8628
meterpreter	0.9105
web shell	0.8354

Table XI
AREA UNDER THE CURVE FOR SVM'S WITH LINEAR KERNEL ON ADFA-LD 12 FOR DIFFERENT ATTACK PROFILES ON THE REDUCED FEATURE SPACE.

was classified as normal when it was within a distance d of a cluster center. The distance parameter was chosen by finding the maximum in cluster distance (d_{max}) and getting an evenly spaced sample of 10 numbers from $[0, d_{max}]$. In my case, the maximum distance is 0.7551. This calibration of the distance parameter allows us to control the sensitivity with which I classify a data point as anomalous, enabling the creation of receiver operating characteristic curves.

After performing my computation, I plot the receiver operating characteristic curves. They are presented in figure 9. To assess the performance of my classifier, I compute the area under the ROC curves and present the results in table IX. I can see that in this case, the hydra ssh attack has bad performance, while for the other attacks, the classifier has moderate performance.

The authors of [23] do not comment on the area under the ROC curves for the classifiers they used. Instead, they just assess individual Precision (True Positive) and Fallout (False Positive) rates. With my implementation, I can compute it. By looking through figures 7, 8, 9 and the tables presenting the area under the ROC curve, VI, VII, IX, I can see that the k-means clustering algorithm is a bit more reliable. Additionally k-means clustering is a lot cheaper computationally. The training times for k-means clustering were at least one order of magnitude less than those of k-nearest neighbours on the computing resources I were using for this work.

As a final remark for using k-means clustering for anomaly detection, I should mention that conceptually, it has a lot of similarities to the Gaussian Mixture Model pioneered by Roberts et al.[5].

XII. SUPPORT VECTOR MACHINES

A. Linear SVM on reduced frequency feature space

A natural step in continuing my analysis of the ADFA-LD 12 dataset further than Xie et al.[23] are Support Vector Machines. As a first, I will keep the methodology I used previously, meaning I will work with the first nine principal components of the training set so I can compare my results before moving forward. I will train an SVM model, using a linear Kernel, for each attack separately. More Specifically, I will use the SVC class from the scikit-learn library[28]. Moreover I will use different values of the regularisation parameter C to better map the hypothesis space of available classifiers. The different values of C I used are:

$$[0.05, 0.1, 0.5, 1, 5, 10, 50]$$

I can see the first results of this computation in figure 10 where I plot various performance points on a True Positive versus False positive map. From there, I conclude that for some attacks, namely web shell and meterpreter, the performance of the SVM classifier is not influenced significantly by the choice of regularisation parameter. On the other hand, hydra ftp and hydra ssh attacks are significantly influenced. For each attack, I note the best-performing value, as I can see in table X.

I proceed with computing the optimum linear SVM classifier for each attack and then plot its receiver operating characteristic curve. The results are shown in figure 11. The area under the curve for each classifier is shown in table XI. As usual, the hydra attacks are my worst performers, while the meterpreter attacks are my best performers.

At this stage, I am able to compare my results, from table XI, with those of the k-means clustering algorithm, from table IX. It is quite clear that Support Vector Machines are a more accurate and reliable classifier for every attack. On average, the SVM classifier has 0.15 higher area under the curve.

Moving forward, I want to see the performance of Support Vector Machines on the dataset as a whole instead of training them specifically for each attack. Hence I will treat the problem as a two class classification problem with my classes being normal behavimy and attack behaviour. To get a better idea of the performance of my classifier in this case, I will use cross-validation. To address the issue of class imbalance in my dataset, I will use stratified 8-fold cross-validation[13].

[13]I chose 8-fold cross-validation instead of 10-fold due to the class imbalance against my attack set and the inner divisions within it. The change is minor but helps us sample the attack set better across folds.

Regularisation (C)	Precision	Fall out
0.125	0.62 ± 0.08	0.20 ± 0.03
0.25	0.62 ± 0.08	0.20 ± 0.03
0.5	0.62 ± 0.08	0.19 ± 0.03
1	0.62 ± 0.08	0.19 ± 0.04
2	0.61 ± 0.09	0.19 ± 0.03
4	0.63 ± 0.07	0.20 ± 0.03
8	0.64 ± 0.08	0.21 ± 0.04
16	0.64 ± 0.10	0.23 ± 0.05
32	0.64 ± 0.11	0.24 ± 0.06

Table XII
8-FOLD STRATIFIED CROSS VALIDATION RESULTS FOR VARIOUS REGULARISATION PARAMETER VALUES OF SVM CLASSIFIER WITH LINEAR KERNEL. RESULTS ARE PRESENTED WITH MEAN AND STANDARD DEVIATION FOR TWO METRICS. TRUE POSITIVE RATE (PRECISION) AND FALSE POSITIVE RATE (FALL OUT). REDUCED FREQUENCY FEATURE SPACE WAS USED FOR TRAINING AND VALIDATION.

Regularisation (C)	Precision	Fall out
0.125	0.62 ± 0.10	0.17 ± 0.04
0.25	0.64 ± 0.09	0.17 ± 0.04
0.5	0.66 ± 0.09	0.16 ± 0.04
1	0.68 ± 0.06	0.16 ± 0.04
2	0.76 ± 0.07	0.19 ± 0.05
4	0.81 ± 0.08	0.19 ± 0.04
8	0.91 ± 0.04	0.18 ± 0.03
16	0.91 ± 0.05	0.18 ± 0.03
32	0.92 ± 0.04	0.18 ± 0.03
64	0.93 ± 0.05	0.16 ± 0.04
128	0.92 ± 0.06	0.16 ± 0.04

Table XIII
8-FOLD STRATIFIED CROSS-VALIDATION RESULTS FOR VARIOUS REGULARISATION PARAMETER VALUES OF SVM CLASSIFIER WITH LINEAR KERNEL. RESULTS ARE PRESENTED WITH MEAN AND STANDARD DEVIATION FOR TWO METRICS. TRUE POSITIVE RATE (PRECISION) AND FALSE POSITIVE RATE (FALL OUT). COMPLETE FREQUENCY FEATURE SPACE WAS USED FOR TRAINING AND VALIDATION.

Going even further, I run my simulation for various values of the regularisation parameter. I present my results in table XII. I can see that I get better performance[14] for low regularisation values. The optimum value is $C = 0.5$.

B. Linear SVM on the complete frequency feature space

Moving further away from the approach of Xie et al.[23] I investigate the application of support vector machines in the frequency feature space without reducing it's dimensionality through principal component analysis. I pool all of my dataset together and perform 8 fold stratified cross validation training a linear SVM classifier. I present my results in table XIII.

As I can see by comparing tables XII and XIII using the complete frequency feature space yields significantly better results. The precision rate is increased noticeably, and the fallout rate is slightly decreased. The computation cost was not noticeably increased for the whole frequency feature space when running my scripts.

C. Recursive feature elimination

[14]Increased performance means higher difference between precision and fallout rates, with 1 being perfect classification results.

Moving forward with my analysis of the ADFA-LD 12 dataset, I will conduct recursive feature elimination. I will train an SVM classifier with a linear kernel for a two-pattern classification problem. My two classes will be attack and normal behavior. A property of k-fold cross-validation is that only a small part of the dataset is used as a validation set, meaning my under-represented attack class may not be appropriately represented in the validation set. For this reason, I will use a sampling method to perform cross-validation. I will be sampling my dataset with the StratifiedShuffleSplit method of scikit-learn[28] 6 times, and my training and validation sets will be equal in size. To create a scorer and assess the performance of my classifier, I will use the Precision (True Positive rate) and Fall out (False Positive rate) ratios. But because recursive feature elimination works by optimising one measure, I will use their difference, namely:

$$\text{Scorer} = \text{Precision} - \text{Fallout}$$

When Scorer=1 I have perfect classification and if Scorer is 0 I are classifying everything as attack behaviour. After writing my code and performing the necessary computations, I plot my results in figure 12.

By looking at the plot, I see that when I start removing the less relevant features from the complete feature space, my performance, measured by my scorer metric, remains steady. That plateau remains with minimal variation until I have 50 remaining features with a global maximum of 54 features. Then it has a small decline that keeps increasing and becomes a steep decline after I am left with only 25 features. Therefore I can deduct that the information content required to perform proper identification of attacks in the context of a two pattern classification problem is contained in the 54 most relevant features.

XIII. ONE CLASS SUPPORT VECTOR MACHINES - OUTLIER DETECTION

Before I finish my analysis of ADFA-LD 12 on the frequency feature space, I will study one class, Support Vector Machines. The 1-class SVM algorithm has been introduced by Schölkopf et al.[6] and can work with a variety of kernel functions just like normal SVM classifiers. From my preliminary results, I saw that linear kernels performed badly. After various trials, I found optimum performance on sigmoid kernels with the parameters $\gamma = 0.05$ and $c_0 = 3$.

One more tricky thing about the 1-class SVM classifier is that on the training set, I am asked to specify the upper bound for the fraction of training data that I can tolerate being incorrectly classified. In order to study the upper bound's (ν) effects, I will test its performance by recursively performing cross-validation on a range of values for ν. The results of my first test can be seen in figure 13 and table XIV.

By studying the results of figure 13 and table XIV, I see that there is big volatility for different numbers of ν. Moreover, the standard deviation of my metrics is quite variable, which is something I didn't expect. This might be due to the relatively small number, 10, of re-sampling or because of

ν	Precision	Fall out
0.1	0.55 ± 0.10	0.19 ± 0.05
0.2	0.53 ± 0.04	0.22 ± 0.03
0.3	0.59 ± 0.02	0.30 ± 0.02
0.4	0.82 ± 0.04	0.41 ± 0.01
0.5	0.89 ± 0.01	0.49 ± 0.02
0.6	0.910 ± 0.001	0.61 ± 0.01
0.7	$0.912 \pm 1 \cdot 10^{-16}$	0.73 ± 0.10
0.8	$0.91 \pm 6 \cdot 10^{-4}$	0.81 ± 0.10
0.9	0.93 ± 0.007	0.904 ± 0.009
1.0	1.0 ± 0.0	1.0 ± 0.0

Table XIV
EXPLORING 1-CLASS SVM WITH A SIGMOID KERNEL TO ASSESS ITS PERFORMANCE DEPENDING ON THE UPPER BOUND FOR THE FRACTION OF TRAINING ERRORS. SHUFFLESPLIT METHOD WAS USED FOR CROSS-VALIDATION TO CREATE TWO, EQUAL IN SIZE, NORMAL BEHAVIOR DATASETS FOR TRAINING AND VALIDATION. OPTIMAL PERFORMANCE FOR UPPER BOUND $\nu = 0.4$.

ν	Precision	Fall out
0.35	0.67 ± 0.04	0.36 ± 0.01
0.36	0.68 ± 0.04	0.37 ± 0.01
0.37	0.71 ± 0.06	0.39 ± 0.02
0.38	0.75 ± 0.04	0.40 ± 0.01
0.39	0.77 ± 0.04	0.40 ± 0.01
0.40	0.83 ± 0.04	0.41 ± 0.02
0.41	0.85 ± 0.01	0.42 ± 0.02
0.42	0.85 ± 0.02	0.42 ± 0.02
0.43	0.858 ± 0.001	0.431 ± 0.010
0.44	0.862 ± 0.002	0.449 ± 0.006
0.45	0.864 ± 0.004	0.45 ± 0.01

Table XV
EXPLORING 1-CLASS SVM WITH A SIGMOID KERNEL TO FIND ITS OPTIMUM PERFORMANCE DEPENDING ON THE UPPER BOUND FOR THE FRACTION OF TRAINING ERRORS. THE SHUFFLESPLIT METHOD WAS USED FOR CROSS-VALIDATION TO CREATE TWO EQUAL-IN-SIZE, NORMAL-BEHAVIOR DATASETS FOR TRAINING AND VALIDATION. OPTIMAL PERFORMANCE FOR UPPER BOUND $\nu = 0.41$.

some structure within my dataset that I haven't identified yet. Further investigation is needed to determine this. In order to identify the best-performing value of the upper bound for misclassification, I repeat my procedure on a smaller range of values. The results of my computation are presented in figure 14 and table XV.

From table XV, I can identify $\nu = 0.41$ as the best-performing value for the upper bound on training errors. Moreover, from looking at both tables XIV and XV, I can see that the fallout rate follows the value of ν, which makes sense given that the training data are all from normal behavior and the fallout rate represents misidentification of normal data.

Because the LIBSVM library[30] implementing the One-class SVM algorithm that I use does not predict probabilities for a data point to be classified as outlier but only classifies them as such I cannot create ROC curves. However, by observing the performance of the algorithm depending on the bound for the training error at figures 13 and 14, I see that there is a resemblance to a receiver operating characteristic.

Regularisation (C)	Precision	Fall out
0.125	0.93 ± 0.02	0.068 ± 0.010
0.25	0.93 ± 0.02	0.062 ± 0.009
0.5	0.92 ± 0.02	0.054 ± 0.009
1	0.91 ± 0.02	0.049 ± 0.007
2	0.91 ± 0.02	0.047 ± 0.006
4	0.88 ± 0.03	0.046 ± 0.007
8	0.86 ± 0.03	0.043 ± 0.006
16	0.84 ± 0.04	0.041 ± 0.005
32	0.81 ± 0.03	0.038 ± 0.004

Table XVI
8-FOLD STRATIFIED CROSS-VALIDATION RESULTS FOR VARIOUS REGULARISATION PARAMETER VALUES OF SVM CLASSIFIER WITH LINEAR KERNEL. RESULTS ARE PRESENTED WITH MEAN AND STANDARD DEVIATION FOR TWO METRICS. TRUE POSITIVE RATE (PRECISION) AND FALSE POSITIVE RATE (FALL OUT). TWO-SEQUENCE FEATURE SPACE WAS USED FOR TRAINING AND VALIDATION.

XIV. 2-Sequence Analysis of the ADFA-LD 12 Dataset

After my analysis of the ADFA-LD 12 dataset on the frequency feature space, I proceed to analyze it on the two-sequence feature space.

I begin by giving a specific definition of the two sequence feature space I will use. As I have said, every point of my dataset consists of a series of kernel operating system calls. Instead of counting the frequency of a individual system calls in each point I will count the frequency of combinations of two subsequent system calls in each point. From my previous explorations of my dataset, I know that I have 175 distinct system calls present. Out of the 30625 possible combinations of 2-sequences, only 3792 are present in my dataset. Hence, my two-sequence feature space has 3792 dimensions.

XV. Support Vector Machines

I start my analysis of the two-sequence feature space by addressing my data as a two-pattern classification problem. I will use a linear kernel for my SVM classifier. I will test its performance on various values of the regularisation parameter (C) performing stratified cross-validation. After computing the necessary calculations, I get the results shown in table XVI.

I see that I get my best-performing results when I have $C = 0.25$. By comparing the results from tables XIII and XVI I see that the two-sequence feature space yields significantly better results compared to the full frequency feature space. This is to be expected since the two frequency feature spaces capture more information from the dataset.

Moving forward, I will conduct recursive feature elimination to see how much information is contained in each feature. I will use the same procedure as before. I will train my classifier to distinguish a two-pattern classification problem. I will use the Stratified Shuffle Split method to perform cross-validation. In order to make the computation time reasonable, I will use a reduction step of 24 features (out of a total of 3792) per iteration. After performing my computation, I get the results depicted in figure 15. I see that performance plateaus at 240 features and stays relatively stable afterward, with an obtuse maximum at 2088 features. Below 240 features performance

ν	Precision	Fall out
0.1	0.55 ± 0.03	0.20 ± 0.01
0.2	0.58 ± 0.01	0.245 ± 0.009
0.3	0.72 ± 0.02	0.32 ± 0.02
0.4	0.826 ± 0.004	0.41 ± 0.02
0.5	0.876 ± 0.006	0.50 ± 0.01
0.6	0.9269 ± 0.0008	0.64 ± 0.07
0.7	0.930 ± 0.001	0.76 ± 0.09
0.8	0.936 ± 0.004	0.86 ± 0.07
0.9	0.9560 ± 0.0005	0.909 ± 0.007
1.0	1.0 ± 0.0	1.0 ± 0.0

Table XVII
EXPLORING 1-CLASS SVM WITH A SIGMOID KERNEL TO ASSESS ITS PERFORMANCE DEPENDING ON THE UPPER BOUND FOR THE FRACTION OF TRAINING ERRORS. THE DATASET WAS FEATURE-ENGINEERED ON A TWO-SEQUENCE FEATURE SPACE. THE SHUFFLESPLIT METHOD WAS USED FOR CROSS-VALIDATION TO CREATE TWO EQUAL-IN-SIZE, NORMAL-BEHAVIOR DATASETS FOR TRAINING AND VALIDATION. OPTIMAL PERFORMANCE FOR UPPER BOUND $\nu = 0.4$.

ν	Precision	Fall out
0.35	0.788 ± 0.008	0.366 ± 0.007
0.36	0.805 ± 0.008	0.379 ± 0.012
0.37	0.807 ± 0.006	0.380 ± 0.013
0.38	0.820 ± 0.008	0.398 ± 0.011
0.39	0.822 ± 0.003	0.403 ± 0.012
0.40	0.827 ± 0.003	0.410 ± 0.010
0.41	0.832 ± 0.005	0.428 ± 0.018
0.42	0.832 ± 0.004	0.424 ± 0.013
0.43	0.835 ± 0.004	0.435 ± 0.011
0.44	0.839 ± 0.005	0.448 ± 0.012
0.45	0.849 ± 0.006	0.464 ± 0.017

Table XVIII
EXPLORING 1-CLASS SVM WITH A SIGMOID KERNEL TO FIND ITS OPTIMUM PERFORMANCE DEPENDING ON THE UPPER BOUND FOR THE FRACTION OF TRAINING ERRORS. THE SHUFFLESPLIT METHOD WAS USED FOR CROSS-VALIDATION TO CREATE TWO EQUAL-IN-SIZE, NORMAL-BEHAVIOR DATASETS FOR TRAINING AND VALIDATION. OPTIMAL PERFORMANCE FOR UPPER BOUND $\nu = 0.36$.

drops significantly, hence I can say that the core information content on normal and attack classes is contained in the 240 most important features.

XVI. ONE CLASS SUPPORT VECTOR MACHINES - OUTLIER DETECTION

The last step of my analysis of ADFA-LD 12 on the two-sequence feature space is studying one class, Support Vector Machines. Again, from my preliminary results, I saw that linear kernels performed badly. On the other hand the sigmoid kernel gives us good performance. In order to get results directly comparable to the ones on the frequency feature space, I will use the same parameters that I used before, $\gamma = 0.05$ and $c_0 = 3$.

As I have already said, one more tricky thing about the 1-class SVM classifier is that on the training set, I am asked to specify the upper bound for the fraction of training data that I can tolerate being incorrectly classified. In order to study the upper bound's (ν) effects on the two-sequence feature space, I will test its performance by recursively performing cross-validation on a range of values for ν. The results of my first test can be seen on figure 16 and table XVII.

While figure 16 gives us a good overview of the performance of one class support vector machine the iteration step, 0.1, for ν (maximum error ratio tolerance for training set) is too big to tell us the optimum value. Hence I re-iterate with smaller iteration step, 0.01, for values between $\nu \in (0.35, 0, 45)$ to find the best performing value. I see my results for $\nu = 0.36$.

XVII. VALIDATION AND TESTING

Before I do that, however, it is important to mention a key part of my work that has been lurking in the shadows. This is validation and testing. All of my numerical analysis has been done through computations. It is easy to make a mistake that introduces a bug in the Python code, but it is very hard to identify that by looking at the results. Therefore, I took the approach that every step of a numerical calculation, once implemented, should be tested through the form of debug messages printed by the terminal or my log files in order to verify them. This approach has not only helped us identify bugs in my code immediately, but it has also shaped how my code was written to make it more resilient.

A key success of that testing was the proper identification of the integers representing the kernel system calls. Initially, through manual exploration of the data files, sampled randomly, I had assumed that they were represented by the integers from 1 to 325. Numerical testing, in this case mostly through summation of the relevant probabilities, revealed that something was missing. A more rigorous exploration of the dataset revealed the problem.

XVIII. RESULTS ANALYSIS

My first goal was to replicate the results of Xie et al.[5]. I successfully achieved that and implemented some trivial improvements that allowed us to fully compute the ROC curve and the area below it.

I then moved further by implementing Support Vector Machines classifiers. Comparing the area under the ROC curves for the various cases described in tables VII, VIII, IX, and XI is not very straightforward, so I created figure 18 to condense that information. I can see that k-means clustering and k-nearest neighbors with squared standardized Euclidean distance are the worst performers, while support vector machines perform a lot better.

I then moved to the complete frequency feature space. When formulating the problem as a two pattern classification I saw, from tables XII and XIII, a significant improvement in performance. This means that the feature reduction approach used by Xie et al.[5] was not efficient in maintaining the information contained in the dataset while reducing its dimensionality. To that extent, I tried recursive feature elimination, and I saw that I could eliminate up to 121 out of 175 dimensions of the feature space with no loss of performance. As I can see from figure 12, I could even go below that, and unless I keep less than the 35 most informative features, I am not sacrificing much on the performance of my algorithm.

I then moved to try an unsupervised learning approach using one-class support vector machines, where I saw a significant drop in performance, as evident from my results on tables XIV and XV.

My next step was to move to a feature space that captured more information from the dataset, the two-sequence feature space. To compare how much my performance improved, I created figure 19, which takes information from tables XIII and XVI. As an evaluation metric I use the scored I have defined earlier as the difference between precision and fall out rate.

Last but not least, I compare the performance of one class support vector machine. This algorithm acts as outlier detection, and from what I see from figures 19, 20, its performance is significantly poorer compared to support vector machines classifier. I can also see that there is very little difference between the two feature spaces. This is in contrast with my previous results when intrusion detection was framed as a two-pattern classification problem.

XIX. Further Research

This work was intended to initiate research on the the field of intrusion detection through modern machine learning approaches. As such it leaves a lot of open questions. One straightforward path to go forward is to take note of the system calls identified as more relevant from recursive feature elimination. Someone with expertise in computer science and, more specifically, with the Linux kernel can give us domain input and provide useful information for more efficient feature engineering.

Another avenue for further research is to check on the scalability of the algorithms I am using. The AWID 2015 dataset by kolias et al.[25] is a very good candidate. Its compressed size is 10Gb, making it a very good candidate for assessing algorithms that are to be trained in a distributed setting or high-performance computer systems. It is a modern dataset using tools representative of the current IT landscape, and results on it will have applications on current networks.

Last but not least, there is great room for improvement in using better kernels in unsupervised outlier detection. The fact that I saw very small performance increase by moving to the two-sequence feature space means that I not training the one-class support vector machine algorithm on a feature space that it can take full advantage of. Taking advantage of kernel methods has the potential to greatly improve performance. And it even if it doesn't, it is useful to know the extent of each algorithm's efficiency. This enhances my overall understanding in the Machine Learning field and helps us find areas in need of novel approaches.

References

[1] G. Creech, J. Huy *Generation of a new IDS Test Dataset: Time to Retire the KDD Collection*, 2013 IEEE Wireless Communications and Networking Conference (WCNC)

[2] M. Schellekens *Car hacking: Navigating the regulatory landscape*, Computer Law & Security Review 32 (2016) 307 - 315

[3] C. Sun, C. Liu and J. Xie *Cyber-Physical System Security of a Power Grid: State-of-the-Art*, MDPI - Electronics open access journal.

[4] I. Raghav, S. Chhikara, N. Hasteer *Intrusion Detection and Prevention in Cloud Environment: A Systematic Review*, Journal of Network and Computer Applications, Volume 36, Issue 1, January 2013, Pages 25-41

[5] S. Roberts and L. Tarassenko, *A probabilistic resource allocating network for novelty detection*, Neural Computation, vol. 6, no. 2, pp. 270 - 284, 1994.

[6] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola, R. Williamson, *Estimating the Support of a High-Dimensional Distribution*, Neural Computation, Vol. 13, No. 7, pp. 1443-1471, 2001.

[7] P. Hayton, B. Scholkopf, L. Tarassenko, and P. Anuzis, *Support vector novelty detection applied to jet engine vibration spectra*, in NIPS, pp. 946 - 952, 2000.

[8] L. Clifton, H. Yin, and Y. Zhang, *Support Vector Machine in Novelty Detection for Multi-channel Combustion Data*, Proceedings of the third international conference on Advances in Neural Networks - Volume Part III (2006)

[9] B. Farran, C. Saunders and M. Niranjan, *Machine Learning for Intrusion Detection: Modeling the Distribution Shift*, 2010 IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2010)

[10] M. Bhuyan, D. Bhattacharyya, and J. Kalita (2014) *Network Anomaly Detection: Methods, Systems and Tools*, IEEE Communications Surveys & Tutorials, Vol. 16, No. 1, 2014

[11] M. Ahmed, A. Mahmood, J. Hu *A survey of network anomaly detection techniques*, Journal of Network and Computer Applications. Vol. 60, January 2016, p. 19-31

[12] C. Kruegel, D. Mutz, W. Robertson, and F. Valeur, *Bayesian event classification for intrusion detection*, in Proc. 19th Annual Computer Security Applications Conference, 2003

[13] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, *RODD: An Effective Reference-Based Outlier Detection Technique for Large Datasets*, in Advanced Computing. Springer, 2011, vol. 133, pp.76–84.

[14] I. Kang, M. K. Jeong, and D. Kong, *A differentiated one-class classification method with applications to intrusion detection*, Expert Systems with Applications, vol. 39, no. 4, pp. 3899–3905, March 2012.

[15] X. Xu, *Sequential anomaly detection based on temporal-difference learning: Principles, models and case studies*, Applied Soft Computing, vol. 10, no. 3, pp. 859–867, 2010

[16] M. Amini, R. Jalili, and H. R. Shahriari, *RT-UNNID: A practical solution to real-time network-based intrusion detection using unsupervised neural networks*, Computers & Security, vol. 25, no. 6, pp. 459–468, 2006

[17] A. Borji, *Combining heterogeneous classifiers for network intrusion detection*, in Proc. 12th Asian Computing Science Conference on Advances in Computer Science: Computer and Network Security. Springer, 2007, pp. 254–260.

[18] C. Aggarwal, A. Hinneburg, and D. Keim, *An empirical evaluation of supervised learning in high dimensions*, ICML '08: Proceedings of the 25th International Conference on Machine learning Pages 96-103

[19] S. J. Stolfo, W. Fan, W. Lee, A. Prodromidis, and P. K. Chan, *Cost-Based Modeling for Fraud and Intrusion Detection: Results from the JAM Project*, in Proc. DARPA Information Survivability Conference and Exposition, vol. 2. USA: IEEE CS, 2000, pp. 130–144.

[20] J. McHugh, *Testing intrusion detection systems: A critique of the 1998 and 1999 Darpa intrusion detection system evaluations as performed by Lincoln laboratory*. ACM Transactions on Information Systems and Systems Security, Volum 3, Issue 4, pages 262 - 294, 2000

[21] M. Mahoney and P. Chan, *An Analysis of the 1999 DARPA/Lincoln Laboratory Evaluation Data for Network Anomaly Detection*, Recent Advances in Intrusion Detection, Volume 2820 of the series Lecture Notes in Computer Science pp 220-237

[22] M. Ahmed, A. Mahmood, J. Hu, *A survey of network anomaly detection techniques*, Journal of Network and Computer Applications. Vol. 60, January 2016, p. 19-31

[23] M. Xie, J. Hu, X. Yu, and Elizabeth Chang *Evaluating Host-Based Anomaly Detection Systems: Application of the Frequency-Based Algorithms to ADFA-LD*, 11th International Conference on Fuzzy Systems and Knowledge Discovery, 2014

[24] M. Xie, J. Hu and J. Slay *Evaluating Host-based Anomaly Detection Systems: Application of the One-class SVM Algorithm to ADFA-LD*, Proceedings of the 11th IEEE International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2014), Xiamen, 19-21 August 2014, 978-982.

[25] C. Kolias, G. Kambourakis, A. Stavrou, and S. Gritzalis *Intrusion Detection in 802.11 Networks: Empirical Evaluation of Threats and a*

Public Dataset IEEE Communication Surveys & Tutorials, Vol. 18, No. 1, 2016

[26] G. van Rossum, *Python tutorial, Technical Report CS-R9526*, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995.

[27] D. Ascher, P. F. Dubois, K. Hinsen, J. Hugunin, and T. Oliphant, *Numerical Python* Lawrence Livermore National Laboratory, UCRL-MA-128569, 1999

[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot , E. Duchesnay *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research, volume 12, pages 2825 - 2830, 2011

[29] J. Hunter, *Matplotlib: A 2D graphics environment*, Computing In Science & Engineering, vol. 9, no. 3, pp 90 - 95, 2007

[30] C. Chih-Chung, L. Chih-Jen, *LIBSVM: A library for support vector machines*, ACM Transactions on Intelligent Systems and Technology, Volume 2, Issue 3 (2011)

Секція 9.
Інтелектуальний аналіз даних, DataMining і BigData–технології.

DOI: https://doi.org/10.30837/IST-2024.P1.158

Quantum Entropy in Machine Learning Tasks WS

Victor Syneglazov[1†] and Petro Chynnyk[2*†]

[1] *National Aviation University, Lubomyr Huzar avenue 1, 03058 Kyiv, Ukraine*
[2] *NTUU "Igor Sikorsky Kyiv Polytechnic Institute", 37 Beresteysky prospect, 03056 Kyiv, Ukraine*

Abstract
This paper investigates quantum entropy as a function of error for model training. A comparative analysis of the quality of models trained using quantum entropy with models trained based on standard error functions: cross-entropy and MSE was performed. The results demonstrate an advantage in the quality and training time of the models that used quantum entropy for training, compared to those models that used classical functions.

Keywords
quantum machine learning, quantum entropy, loss function, cross-entropy, mse

1. Introduction

In recent years, quantum computing has become a promising frontier that could change the way we process information. Improve artificial intelligence models, improve information protection, speed up modeling of complex natural processes.

If we consider the use of artificial intelligence in quantum computing, one of the more difficult problems is learning. Today, classical error functions such as cross-entropy and MSE are used to train quantum models of artificial intelligence. We suggest using another function, namely quantum entropy.

Derived from quantum information theory, quantum entropy offers a probabilistic framework that captures uncertainty in ways not accessible through classical measurements. Incorporating such quantum principles into the design of error functions for machine learning models opens up new ways to optimize learning processes and increase efficiency.

The main goal of this work is to perform a comparative analysis of models trained by quantum entropy with models trained by classical functions such as cross-entropy and MSE. In particular, we investigate whether the introduction of quantum entropy can lead to improvements in convergence speed, model accuracy, and overall quality. By investigating the behavior of quantum entropy as a function of error, we aim to find out whether it offers measurable advantages over traditional optimization strategies.

2. Quantum entropy

Quantum entropy is an extension of the concepts of entropy for quantum systems. Its development began within the framework of quantum mechanics and quantum information in order to describe uncertainty and mixed states in quantum systems.

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
* Corresponding author.
† These authors contributed equally.
svm@nau.edu.ua (V. Syneglazov); chynnyk@vivaldi.net (P. Chynnyk)
0000-0002-3297-9060 (V. Syneglazov); 0009-0002-6787-5429 (P. Chynnyk)

The foundations were laid in 1932 by John von Neumann, who first introduced the von Neumann entropy. This entropy generalizes the classical Shannon entropy to the case of quantum states (1).

$$S(\rho) = -Tr\big(\rho log(\rho)\big) \tag{1}$$

where ρ – density matrix.

In order to translate the quantum entropy into an error function, it is necessary to introduce a reference density matrix that will represent the class of the object. This reference density matrix must represent a pure quantum state.

For binary classification, we propose these reference density matrices (2,3)

$$\rho = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{2}$$

$$\rho = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{3}$$

The Scaling and Squaring Algorithms algorithm was used to calculate the logarithm from the matrix.

For learning the quantum machine learning model we suggest the next loss function

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N} -Tr\left(\rho_i log\big(\rho_i(\theta)\big)\right) \tag{4}$$

where N – number of objects, θ – parameters of model.

3. Result of experiments

For training, we used the iris dataset, from which we selected two classes. We selected 70 objects for the training sample, 30 for the test sample.

The quantum model was as follows (Fig. 1)

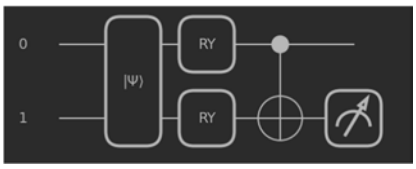


Figure 1: Quantum circuit of quantum machine learning model

The training was carried out using Adam's algorithm.

Result of learning show in Table 1

Table 1
Result of learning

Optimize	Type Loss	Train	Test	Accuarcy Train	Accuaracy Test	F1 Train	F1 Test
MSE	MSE	0,4748	0,50203	1	1	1	1
	Cross Entropy	0,41924	0,43402				
	Quantum Entropy	0,45703	0,4955974				

Optimize	Type Loss	Train	Test	Accuarcy Train	Accuaracy Test	F1 Train	F1 Test
Cross Entropy	MSE	0,54974	0,56639	1	1	1	1
	Cross Entropy	0,46166	0,470315				
	Quantum Entropy	0,60931	0,67563				
Quantum Entropy	MSE	0,45424	0,45344	1	1	1	1
	Cross Entropy	0,40953	0,40893				
	Quantum Entropy	0,42522	0,42784				

Below are graphs of model training and graphs of other error functions based on model training

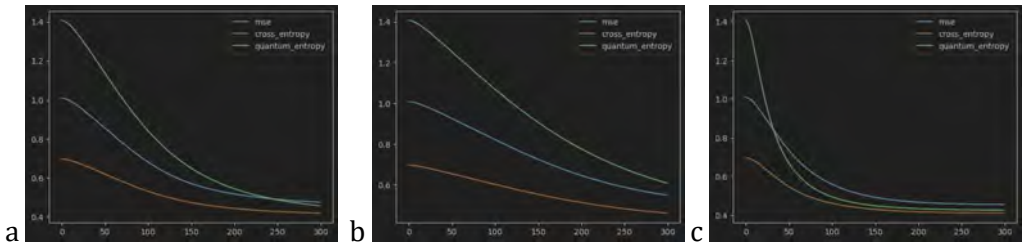


Figure 2: Graphic of learning of model. a - MSE, b – cross-entropy, c – quantum entropy

4. Conclusion

As we can see from the experimental results, all models show excellent prediction results regardless of the type of loss function. As we can see from graph 2, the quantum entropy loss function shows better results.

The error function used for training, quantum entropy, trains the model faster than MSE or cross-entropy. Already at 100 iterations, the error value is smaller than that of MSE. And the value of MSE itself is smaller than when using MSE for training.

Learning with quantum entropy is better because it better takes into account the very nature of quantum computing and exploits the implicit, hidden patterns between qubits.

References

[1] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles, Variational Quantum Algorithms, 2019. URL: https://arxiv.org/pdf/2012.09265.pdf

[2] L. Wright, F. Barratt, J. Dborin, V. Wimalaweera, B. Coyle, and A. G. Green, Deterministic Tensor Network Classifiers, 2022. URL: https://arxiv.org/pdf/2205.09768.pdf

[3] Diederik P. Kingma, Jimmy Lei Ba, Adam: A Method For Stochastic Optimization, 2014. URL: https://arxiv.org/pdf/1412.6980

[4] John von Neumann, Mathematical Foundations of Quantum Mechanics, 1932

DOI: https://doi.org/10.30837/IST-2024.P1.161

Агентна технологія інтелектуального моніторингу

Анатолій Білоніг[1] та Сергій Голуб[1]

[1] *Черкаський державний технологічний університет, бульвар Шевченка 460, 18000 Черкаси, Україна*

Анотація
Система інтелектуального моніторингу являє собою складну структуру, яка складається з переліку взаємозв'язаних елементів, що при взаємодії забезпечують коректну роботу системи прийняття рішень. В роботі наведено технологію – послідовність етапів, методів та засобів, що необхідні для коректної та результативної роботи такої системи, реалізованої у вигляді програмного агенту, описано їх основний функціонал та запропоновано перелік рішень для подальшого покращення ефективності їх роботи.

Ключові слова
програмний агент, інтелектуальний моніторинг, модель, штучний інтелект, розпізнавання образів

1. Вступ

Основна задача програми інтелектуального моніторингу полягає у забезпеченні належним набором знань системи прийняття рішень, що досягається завдяки забезпеченню безперервних спостережень за предметом моніторингу, обробці та перетворенню результатів спостереження [1]. Разом із застосуванням технологій агентного програмування така система моніторингу забезпечує високі показники автономності та ефективності.

Для реалізації поставлених завдань зазвичай будується складна система, що складається з переліку більш вузько-направлених підсистем, при коректній взаємодії яких програмний агент здатний забезпечити задовільний результат. Такі підсистеми можуть як входити до складу самого агенту, так і бути зовнішніми, не зв'язаними з ним, модулями. Для побудови подібних систем існує як необхідний набір ключових елементів, наявність яких є обов'язковою для працездатності агенту, так і перелік додаткових модулів та методів, що можуть, при коректному їх використанні, значно покращити загальну ефективність роботи програми.

2. Основна частина

Якщо розділити роботу системи інтелектуального моніторингу на етапи, то першим таким етапом, незалежно від призначення агенту, буде процес отримання експериментальних даних. Такі дані можуть бути отримані різними способами відповідно до задач конкретної системи, але фундаментально цей процес також можна розділити на дві категорії. У першому випадку до системи безпосередньо під'єднуються додаткові модулі, що дозволяють отримувати дані з навколишнього середовища в реальному часі: мікрофони, камери, сканери, датчики, сенсори, тощо. В іншому випадку експериментальні дані надходять до основної частини системи вже у вигляді файлів того чи іншого формату: фото, відео, аудіо, текст, числові дані. Такий підхід є ефективним коли

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
a.v.bilonih.asp23@chdtu.edu.ua (А. Білоніг); s.holub@chdtu.edu.ua (С. Голуб)
0009-0007-0114-9519 (А. Білоніг); 0000-0002-5523-6120 (С. Голуб)

предметом моніторингу є певне середовище, що характеризується великим набором статистичних параметрів [2].

Не зважаючи на спосіб отримання експериментальних даних, наступним етапом роботи системи інтелектуального моніторингу є перетворення набору незв'язаних та неструктурованих показників на масив вхідних даних (МВД). Попередньо, якщо експериментальні дані надійшли до системи в оригінальному форматі, відбувається процес їх перетворення на масив числових даних, цей процес залежить від формату даних, відповідно якого будуть обрані методи та алгоритми, що будуть застосовані для того чи іншого типу файлів. У випадку звукових повідомлень можуть бути застосовані різні варіації алгоритму перетворення Фур'є, зображення зазвичай перетворюються на масиви RGB представлень для їх пікселів, чи наборів пікселів, що містять в собі певні закономірності. Відео можна розбити на кадри а потім повторити процес, що використовується для перетворення зображень, текст, в свою чергу, перетворюється на набір чисельних параметрів, що відображають кількість або частоту появи тих чи інших букв і символів, разом із визначеними вказівниками на певні параметри, що наперед вказуються перед процесом перетворення.

Основним етапом формування МВД є процес під час якого набори числових даних об'єднуються та структуруються до певного, визначеного наперед формату після чого відбувається їх попередня обробка. Процес попередньої обробки зазвичай складається із декількох циклів перетворення та стиснення даних з метою пошуку найбільш визначних властивостей та ознак, що притаманні тим чи іншим об'єктам які розглядаються системою: процес глобальних перетворень, розкладання даних по мінімальному числу стандартних векторів, кластерний аналіз, тощо [3]. Варто зазначити, що методи та способи проведення попередньої обробки є досить різноманітними і якщо певний набір методів та дій є оптимальним для одного набору даних і покращує результат розпізнавання образів, то це не завжди означає, що цей самий набір методів покаже аналогічний результат і з наступним набором даних. Тому, найкращим рішенням буде створити декілька методів, чи наборів методів, що будуть застосовуватися або лише до конкретного типу даних, або, за допомогою програмного агенту, будуть застосовуватися окремо і тестуватися з метою визначення системою найкращого результату обробки. Також, можливим способом реалізації буде модульний підхід, в якому методи попередньої обробки будуть протестовані наперед з різними типами даних і, в випадку коли до системи моніторингу на вхід надходить новий набір даних, агент буде використовувати (підключати модуль) лише певний набір операцій, що на відміну від тестування загального переліку методів буде відносно зменшувати загальний час обробки.

Після обробки МВД надходять до основної обчислювальної підсистеми програмного агенту – до бази модельних знань (БМЗ). БМЗ – набір навчених та підготовлених до подальшої роботи моделей, що і забезпечують систему інтелектуального моніторингу ефективним інструментом прийняття рішень. Подібні моделі можуть бути побудовані як на основі алгоритмів Методу групового урахування аргументів (МГУА), так і за допомогою використання штучного інтелекту, на основі різноманітних алгоритмів побудови нейронних мереж. Ці підходи можна як і поєднувати, так і використовувати окремо, оскільки, наприклад відомо що, МГУА є найбільш ефективним методом для отримання поліноміального опису стохастичної системи з малої кількості експериментальних даних [4], тоді як нейромережі є більш гнучкими та адаптивними для використання під час вирішення самих різноманітних задач. Узагальнено нейронні мережі можна розділити на декілька типів, що мають схожу основу і загальне призначення, але будуть відрізнятися за внутрішньою будовою та принципом роботи і будуть найбільш ефективні під час виконання конкретних завдань [5]. Ще одним досить ефективним варіантом є створення модульної нейронної мережі, що являє собою

поєднання незалежних мереж в одну систему. Декілька нейронних мереж виконують свої конкретні завдання, тоді як вихід із них поєднується у загальний вихід глобальної мережі.

Незалежно від методів синтезу моделей, вони поєднуються у єдину структуру, де навчені моделі групуються за певними параметрами та їх призначенням, що наперед визначені експертним методом розробниками та спеціалістами під час створення агенту інтелектуального моніторингу. Наприклад, якщо метою конкретного агенту є моніторинг певного типу об'єктів за допомогою транслювання зображень у реальному часі, моделі можуть бути розділені за типами цих об'єктів, за їх кольором, розміром, формою, тощо. Конкретні правила класифікації моделей залежать від призначення системи моніторингу та заданих розробником інструкцій.

Важливо відмітити, що в системі моніторингу використовуються вже навчені моделі, що були наперед треновані для виконання конкретних задач. Це означає, що процес навчання та тестування моделей має відбуватися в окремій системі, або зовнішньому модулі конкретного програмного агенту, проте, в такому випадку підсистеми з навчання та власне моніторингу ніяким чином не повинні взаємодіяти, щоб не допустити внутрішніх змін у уже готових моделях. Оптимальним варіантом буде створити окрему систему для навчання та тестування моделей за певним набором критеріїв та параметрів, що дозволить проводити точний контроль та налаштування для кожної конкретної моделі, незалежно від її загальної структури або конкретних особливостей.

Отже, програмний агент інтелектуального моніторингу за один робочий цикл здатен отримати набір вхідних даних у їх оригінальному вигляді, за потреби перетворити їх в набір числових даних та, після попередньої обробки, сформувати масив вхідних даних, що на наступному кроці надійде до бази модельних знань, В результаті, після певної кількості обчислень та операцій з наборами моделей, агент на виході забезпечить користувача результатом процесу розпізнавання образів відповідно до призначення та будови агенту. Наведений алгоритм дій, є оптимальним мінімумом, що забезпечить високий показник ефективності та точності агенту, проте це не означає, що така система є завершеною і не потребує покращень. Існує велика кількість алгоритмів, функцій та методів для покращення результативності агентної технології інтелектуального моніторингу і модульна структура системи забезпечить можливість модифікації конкретного етапу алгоритму без глобальних змін у загальній структурі системи.

3. Висновки

В роботі було наведено агентну технологію інтелектуального моніторингу, та його узагальнена структура. Було запропоновано розділити процес побудови системи на етапи, що характеризуються їх призначенням та особливостями.

Описано алгоритм дій та засоби отримання експериментальних даних для подальшої обробки в системі моніторингу, вказано на можливість застосування методів для перетворення вихідних даних до числового формату напряму під час процесу формування масиву вхідних даних.

Описано процес та алгоритми, що можливо застосувати під час побудови масиву вхідних даних, запропоновано модульну систему де конкретні алгоритми будуть застосовуватися до конкретних типів даних, та не будуть впливати на роботу підсистеми в цілому.

Наведено визначення та структуру бази модельних знань, описано основні методи та алгоритми синтезу моделей, особливості побудови підсистеми обчислень та розпізнавання образів у програмному агенті інтелектуального аналізу. Запропоновано можливість під'єднання до системи моніторингу окремого модулю, де буде напряму здійснюватися навчання моделей без взаємодії такого модулю із глобальною системою.

Список літератури

[1] Kunytska, Svitlana & Holub, Serhii, Multi-agent Monitoring Information Systems, Mathematical Modeling and Simulation of Systems (2020), 164-171, doi:10.1007/978-3-030-25741-5_17.

[2] Croushore, Dean, Frontiers of Real-Time Data Analysis, Journal of Economic Literature, 49 (2011), 72–100. doi:10.1257/jel.49.1.72.

[3] В. Хомюк, Методи попередньої обробки інформації в процесі розпізнавання образів, 2002. URL: http://ir.lib.vntu.edu.ua/handle/123456789/14988.

[4] Ivakhnenko, Alexey, Polynomial theory of complex systems. IEEE Transactions on Systems, Man and Cybernetics. SMC-1, 4 (1971), 364-378. doi:10.1109/TSMC.1971.4308320.

[5] Shrestha, Ajay & Mahmood, Ausif, Review of Deep Learning Algorithms and Architectures, IEEE Access, PP (2019), 1-29. doi:10.1109/ACCESS.2019.2912200.

DOI: https://doi.org/10.30837/IST-2024.P1.165

Механізми запитів у системах на основі подій: огляд та аналіз методів

Юрій Гунченко[1,†] and Ігор Янкін[1,*,†]

[1] *Одеський національний університет імені І. І. Мечникова, Всеволода Змієнка 2, 65082 Одеса, Україна*

Анотація

У роботі проведено огляд існуючих підходів до обробки запитів у системах на основі подій, зокрема застосування зворотного індексу та проекцій. Висвітлено їхні переваги, обмеження та можливості комбінування методів для підвищення ефективності. Особлива увага приділена важливості інструментів оцінки продуктивності впроваджених рішень та розробки нових методів, які виходять за рамки класичних підходів. Результати дослідження підкреслюють потребу в інноваційних підходах, що забезпечать гнучкість, масштабованість та продуктивність систем у динамічних середовищах.

Ключові слова

системи на основі подій, оцінка ефективності, механізми запитів

1. Вступ

Зростання популярності систем на основі подій та розвиток сучасних технологій обробки даних сприяє формуванню нових підходів до збереження та аналізу інформації. Ефективна робота з механізмами запитів у таких системах є викликом для розробників. Існуючі рішення, такі як патерни «Event Sourcing» і «CQRS», пропонують основи для реалізації подібних механізмів, однак значна частина поточних підходів стикається з обмеженнями, що пов’язані з продуктивністю та складністю отримання актуального стану об'єкта.

Водночас важливим аспектом є необхідність виконання запитів до агрегатів, а не до окремих подій. Як показано в дослідженнях, робота лише з окремими подіями часто є недостатньою для виконання аудиту або отримання повного контексту стану системи [1]. Сучасні системи мають справу з кореляційними зв’язками між подіями, які формують загальний стан агрегатів. Без урахування цих зав’язків відтворення повного стану системи стає складним завданням, що впливає на її ефективність та можливість відновлення після збоїв. Таким чином, необхідно вдосконалювати підходи до обробки даних у таких системах, щоб забезпечити ефективну роботу з агрегатами та їх станами.

Дана робота присвячена аналізу існуючих методів і огляду матеріалів, що стосуються механізмів запитів в системах на основі подій, з метою визначення можливостей для підвищення ефективності запитів у межах обраного напряму дослідження.

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
* Corresponding author.
† These authors contributed equally.
gunchenko@onu.edu.ua (Y. Gunchenko); yankiniigor@gmail.com (I. Yankin)
0000-0003-4423-8267 (Y. Gunchenko); 0009-0007-8262-5504 (I. Yankin)

2. Постановка задачі

Основна мета дослідження полягає в аналізі існуючих підходів до механізмів запитів у системах на основі подій. У межах цього аналізу розглядаються обмеження існуючих рішень, а також можливості їх подолання. Особлива увага приділяється забезпеченню швидкого доступу до актуального стану об'єктів, оптимізації продуктивності запитів і зменшенню затримок у виконанні операцій.

3. Аналіз існуючих рішень

Аналіз існуючих рішень демонструє, що проблема ефективного виконання запитів у системах на основі подій є широко відомою, і для її вирішення запропоновано низку підходів. Кожен із розглянутих підходів має свої переваги та обмеження, які впливають на їх ефективність у різних умовах.

Одним з існуючих рішень є створення зворотного індексу подій [2]. Основна ідея полягає в тому, що події групуються в документи залежно від їх кореляційних зав'язків, що дозволяє виконувати пошук за безпосередньо або опосередковано пов'язаними подіями.

Імплементація зворотних індексів включає інтеграцію подій у реальному часі, що часто створює виклики через потребу постійного оновлення документів у міру додавання нових подій. Так, у деяких з систем, кожна подія додається до відповідного документа, який потребує повної переіндексації, що може стати вузьким місцем у продуктивності всієї системи.

Через це, багато досліджень зосереджені саме на покращенні індексації для конкретних сценаріїв, таких як високопродуктивні бази даних для систем на основі подій, наприклад, «ATLAS» [3]. У цьому випадку використовується реляційна модель з оптимізованим зберіганням та компресією даних для мінімізації обсягів і покращення продуктивності запитів, забезпечуючи швидке додавання подій.

Однак, такі методи мають суттєві обмеження, серед яких високі витрати на підтримку актуальності індексу, труднощі з обробкою великих обсягів даних у реальному часі, що є зазвичай однією з ключових вимог до систем на основі подій, та необхідність вирішення питань узгодженості даних при оновленні. Це вимагає подальших досліджень для розробки більш ефективних і гнучких рішень.

Ще один підхід передбачає використання проекцій для створення окремих представлень даних для обробки специфічних запитів [4]. Проекції дозволяють зберігати дані у формі, яка оптимізована для певних типів запитів, забезпечуючи швидкий і ефективний доступ до них. Підхід базується на асинхронній обробці подій, де дані з основного сховища відтворюються у вигляді окремих представлень, що можуть бути збережені в базах даних залежно від потреб системи.

Хоча цей спосіб і дозволяє ефективно і швидко доступитись до даних, необхідність підтримки актуальності проекцій потребує значних обчислювальних ресурсів. Проекції повинні бути заздалегідь визначені, що обмежує їх гнучкість, особливо у випадках, коли структура або типи запитів змінюються, що є розповсюдженим явищем для систем на основі подій у зв'язку з тим, що під час довготривалого користування такою системою можуть змінюватись бізнес процеси, які ця система відображає, що призводить до необхідності відповідних змін у системі [5]. Крім того, проекції залежать від консистентності подій і можуть втратити актуальність у разі затримок в обробці або помилок у потоці подій. Це створює додаткові виклики для розробників, які повинні забезпечити автоматичне оновлення і відновлення проекцій для збереження точності даних.

Проекції є ефективним рішенням для систем, де структура запитів і обробка даних можуть бути передбачені наперед, однак їх застосування у динамічних середовищах може бути ускладнене через необхідність адаптації до змінних вимог та високі витрати на підтримку.

Ці підходи є класичними рішеннями для вирішення проблеми ефективного виконання запитів у системах на основі подій. Вони широко застосовуються в різних сценаріях завдяки своїй зрозумілості та перевіреності часом. Зворотній індекс дозволяє забезпечити універсальний підхід до пошуку подій, що робить його використання виправданим для систем з високим ступенем невизначеності, де неможливо заздалегідь передбачити набір запитів, що буде виконуватись. Однак, через його високі вимоги до ресурсів, для систем з заздалегідь визначеним набором запитів частіше використовують створення проекцій.

Комбінування зворотного індексу подій та проекцій є перспективним підходом для підвищення ефективності запитів у системах на основі подій. Ця стратегія дозволяє об'єднати переваги обох методів, мінімізуючи їх обмеження. Зворотний індекс забезпечує універсальність і потужні можливості пошуку за кореляційними зв’язками подій, тоді як проекції пропонують швидкий доступ до попередньо визначених представлень даних, оптимізованих для специфічних запитів.

У цій комбінації зворотний індекс використовується для ідентифікації подій, які відповідають складним критеріям пошуку, а проекції – для швидкого отримання результатів, попередньо підготовлених на основі цих подій. Такий підхід дозволяє значно зменшити навантаження на систему, оскільки основний обсяг обчислень виконується під час підготовки проекцій, а не під час виконання запиту.

Одним із викликів у комбінуванні цих методів є забезпечення узгодженості між даними в зворотному індексі та проекціях, особливо у системах із високою частотою оновлень. Затримки в оновленні проекцій або неузгодженість між подіями та їхнім представленням у проекціях можуть призвести до некоректних результатів запитів. Щоб подолати ці проблеми, можна застосовувати асинхронні механізми оновлення або оптимізувати процеси індексації та побудови проекцій, що забезпечить мінімальні затримки і високу актуальність даних.

Комбінування індексації та проекцій доцільно використовувати в системах, де є широкий спектр типів запитів: від складних пошуків за подіями до швидкого доступу до агрегованих даних. Такий підхід дозволяє збалансувати продуктивність системи, гнучкість у запитах і ефективність обробки великих обсягів даних.

Незважаючи на те, що зворотний індекс та проекції є перевіреними та робочими підходами, їхні обмеження стимулюють пошук інноваційних підходів, які могли б забезпечити швидкий доступ до даних, зменшити накладні витрати та підвищити ефективність роботи з запитами в системах на основі подій.

Крім підвищення ефективності запитів, деякі дослідження орієнтуються на створення зрозумілого способу оцінити ефективність впроваджених рішень. Така оцінка є критично важливою, оскільки дозволяє не лише визначити переваги та недоліки певних підходів, але й обґрунтувати вибір технологій та оптимізувати архітектуру систем. Так, у роботі «Developing a performance evaluation benchmark for event sourcing databases» [6] було запропоновано інструменти, що дозволяють оцінити ефективність певних баз даних під час роботи з подіями. А у дослідженні «On the impact of event-driven architecture on performance: An exploratory study» [7] проведено комплексне порівняння систем на основі подій з монолітними системами. Такі дослідження підкреслюють важливість оцінки не лише з точки зору продуктивності, а й з урахуванням специфічних характеристик систем на основі подій, таких як їх масштабованість, адаптивність та здатність обробляти великі обсяги даних у реальному часі. Наявність чітких методів оцінки дозволяє ефективно

інтегрувати нові рішення, мінімізувати ризики і забезпечити високий рівень продуктивності та надійності систем.

4. Висновки

Під час дослідження було проведено аналіз існуючих підходів до обробки запитів у системах на основі подій, зокрема зворотного індексу та проекцій. Було з'ясовано, що ці методи, хоча й широко використовуються, мають суттєві обмеження, пов’язані з продуктивністю, гнучкістю та витратами на підтримку актуальності даних. Виявлено, що навіть перевірені підходи потребують вдосконалення або комбінування для досягнення оптимальних результатів.

Зазначено брак альтернативних методів, які могли б доповнити або замінити класичні підходи, що вказує на необхідність подальших досліджень, спрямованих на пошук нових стратегій, здатних підвищити гнучкість і масштабованість систем на основі подій. Особливо актуальним є врахування специфіки динамічних середовищ і змінних бізнес-процесів, які вимагають адаптивних рішень.

Підкреслено важливість використання інструментів для оцінки продуктивності впроваджених рішень. Такі інструменти надають можливість об'єктивного порівняння підходів, визначення їх переваг і недоліків, а також адаптації рішень до специфічних умов використання, сприяючи створенню стабільних і ефективних систем.

В результаті проведеного аналізу, виявлено значний потенціал для вдосконалення існуючих рішень та розробки нових підходів, які відповідають вимогам сучасних систем на основі подій.

Література

[1] S. Lima, J. Correia, F. Araujo, and J. Cardoso, "Improving observability in Event Sourcing systems," Journal of Systems and Software, vol. 181, pp. 111015, Jun. 2021, doi: 10.1016/j.jss.2021.111015.

[2] R. Vecera, S. Rozsnyai, and H. Roth, "Indexing and search of correlated business events," in Proc. Second Int. Conf. Availability, Reliability and Security (ARES), 2007, pp. 1124-1134, doi: 10.1109/ARES.2007.100.

[3] E. J. Gallas, G. Dimitrov, P. Vasileva, Z. Baranowski, L. Canali, A. Dumitru, A. Formica, and the ATLAS Collaboration, "An Oracle-based event index for ATLAS," Journal of Physics: Conference Series, vol. 898, Track 2: Offline Computing, 2017, doi: 10.1088/1742-6596/898/4/042033.

[4] M. Overeem, M. Spoor, and S. Jansen, "The dark side of event sourcing: Managing data conversion," in Proc. 2017 IEEE 24th Int. Conf. Software Analysis, Evolution and Reengineering (SANER), 2017, pp. 193-204, doi: 10.1109/SANER.2017.7884621.

[5] M. Overeem, M. Spoor, S. Jansen, and S. Brinkkemper, "An empirical characterization of event sourced systems and their schema evolution—Lessons from industry," *Journal of Systems and Software*, vol. 178, pp. 110970, 2021, doi: 10.1016/j.jss.2021.110970.

[6] R. Malyi and P. Serdyuk, "Developing a performance evaluation benchmark for event sourcing databases," Journal of Lviv Polytechnic National University "Information Systems and Networks, vol. 15, pp. 159-168, Jul. 2024, doi: 10.23939/sisn2024.15.159.

[7] H. Cabane and K. Farias, "On the impact of event-driven architecture on performance: An exploratory study," Future Gener. Comput. Syst., vol. 153, pp. 52-69, 2024, doi: 10.1016/j.future.2023.10.021.

DOI: https://doi.org/10.30837/IST-2024.P1.169

Semantic Image Segmentation on Autonomic Devices: Prospects for Implementing Deep Learning Models

Mykhailo Trusov[1*†] та Dmitro Uzlov[1†]

[1] *V.N. Karazin Kharkiv National University*

Abstract

The implementation of deep learning models for semantic segmentation on autonomic devices is a promising direction in developing intelligent systems capable of analyzing visual information independently. The present study explores the possibilities and challenges of using such models, employing theoretical analysis, systematization, and generalization of their usage in autonomic devices. Our research reveals significant potential for this technology, identifying key parameters influencing model efficiency on resource-constrained devices and proposing implementation recommendations. Furthermore, we highlight contemporary approaches to model encryption, emphasizing the need for a comprehensive security approach combining hardware and software solutions. The study concludes that advancements in this field will contribute to more efficient autonomic systems across various applications, including computer vision and robotics. By addressing technical challenges and security considerations, this research paves the way for robust autonomic visual processing systems, potentially revolutionizing industries reliant on intelligent image analysis. The relevance of this study lies in its practical insights into implementing deep learning models for semantic segmentation in resource-constrained environments, contributing to the evolving field of autonomic intelligent systems.

Keywords

Semantic segmentation, deep learning, computer vision, model optimization, autonomic devices

1. Introduction

Computer vision is a rapidly evolving field of artificial intelligence (AI) focused on creating technologies that replicate human visual perception. It combines elements from computer science, optics, mechanics, and neuroscience to develop algorithms for analyzing and interpreting visual data [1]. A key area within this field is semantic image segmentation, which classifies each pixel in an image according to specific categories, providing deeper insights as compared to simple object detection [2]. Semantic segmentation models generate a segmentation map where each pixel is color-coded based on its semantic class, allowing for the creation of masks that highlight distinct areas of an image [3]. This technique is crucial for applications in satellite imagery, UAV data analysis, robotics, medicine, and automotive industries. Current approaches to semantic segmentation include traditional methods that rely on feature extraction and pixel classification, which can be time-consuming and dependent on domain expertise [4]. To overcome these limitations, deep learning techniques using neural networks have been developed, significantly enhancing accuracy and adaptability. Prominent architectures for semantic segmentation include Fully Convolutional Networks (FCN), U-Net, DeepLab, and PSPNet. However, integrating these neural networks into autonomic devices with limited or no Internet access represents the challenging task. This work aims to explore the

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
* Corresponding author.
† These authors contributed equally.
mykhailo.trusov@karazin.ua (M. Trusov); dmytro.uzlov@karazin.ua (D. Uzlov)
0009-0001-4390-5307 (M. Trusov); 0000-00Q3-3308-424X (D. Uzlov)

application of deep learning models in such devices and analyze effective integration methods without constant connectivity.

2. Utilization of deep learning models in autonomic devices for semantic image segmentation

Recent research in semantic segmentation demonstrates that deep learning models, particularly Fully Convolutional Networks (FCNs) and other neural networks, can be effectively deployed on autonomic devices without Internet access. This is achieved through the on-device inference, where trained models are deployed directly on the device for local predictions [5]. This approach reduces latency, enhances privacy, and allows operation in areas with limited or no Internet connectivity.

Edge computing is another concept that places computational resources closer to data sources, enabling local processing. For instance, smart cameras can analyze video streams directly on the device using pre-trained models, reducing dependence on constant internet connections and facilitating real-time data processing [6].

Specialized embedded systems like NVIDIA Jetson, Google Coral, and Intel Movidius are widely used for efficient neural network execution on autonomic devices. These platforms are optimized for neural network operations, significantly accelerating computations compared to standard processors while maintaining low power consumption [7].

Various optimization techniques are employed to ensure efficient neural network operation on resource-constrained devices. These include model compression methods such as quantization, pruning, and knowledge distillation, which reduce computational complexity while maintaining accuracy. Hardware acceleration using GPUs, TPUs, or AI-specific chips is another crucial optimization approach, leveraging specialized computing devices optimized for neural network operations [8].

Combining model compression and hardware acceleration methods achieves a synergistic effect, significantly enhancing neural network efficiency on embedded systems. This comprehensive optimization approach opens up broad prospects for implementing complex AI algorithms in mobile and embedded devices, expanding their application scope and increasing the autonomy and intelligence of edge devices.

Offline data processing is another aspect of modern embedded systems that allows devices to perform complex tasks without constant internet connectivity. This approach is particularly useful in situations where devices collect and process data in batches, performing segmentation tasks autonomicly and only occasionally connecting to the internet for updates or additional data transmission.

3. Key determinants of trained model storage requirements

The memory requirements for deploying trained models on devices are influenced by several key factors. Model complexity plays a significant role, with more intricate architectures generally demanding more storage space. The specific neural network architecture chosen also impacts memory usage, as some designs are inherently more efficient than others.

Optimization techniques are crucial in managing memory consumption. Quantization, which reduces the precision of model weights, can substantially decrease memory requirements, often compressing models by up to 75%. Pruning, another effective method, removes less critical weights or neurons, further reducing model size while maintaining performance. Knowledge distillation, where a smaller model learns to emulate a larger one, can result in compact models with comparable effectiveness.

The choice of model is also pivotal. Lightweight architectures like MobileNetV2, designed specifically for mobile applications, typically occupy only 10-15 MB after optimization [9]. In

contrast, more complex models such as full-sized Fully Convolutional Networks (FCNs) or advanced architectures like ResNet can require hundreds of megabytes or even several gigabytes [10].

Additional considerations include the memory needed for runtime environments (e.g., TensorFlow Lite, ONNX Runtime) and auxiliary data such as class labels or configuration files. These components contribute to the overall memory footprint and must be accounted for in deployment planning.

Broadly speaking, memory requirements can be categorized as follows: 10-50 MB for simple tasks and lightweight models, 50-200 MB for moderately complex models with some optimization, and 200 MB to several GB for sophisticated, high-performance models.

By leveraging efficient architectures and applying appropriate optimization techniques, it is possible to deploy neural networks on devices with limited memory resources. This capability enables many automated devices to operate autonomicly, even in scenarios with limited or no Internet connectivity.

4. Implementation of trained models in custom software products

To use trained models in custom applications, specialized libraries or frameworks are necessary [11]. Here are some popular options for various platforms:

- TensorFlow Lite which is designed for mobile and embedded devices, supporting Python, C++, Java, and Swift. It offers model conversion, inference APIs, and hardware acceleration support.
- ONNX Runtime which represents a cross-platform environment for ONNX format models, supporting multiple languages and optimized for various hardware accelerators.
- PyTorch Mobile which is useful for deploying PyTorch models on mobile devices, supporting Python, Java, and C++.
- Core ML which stands for the framework for iOS and macOS applications, primarily used with Swift and Objective-C.
- NVIDIA TensorRT which represents a toolkit for high-performance deep learning inference on NVIDIA GPUs, supporting Python and C++.
- OpenCV which is characterized as an open-source computer vision and machine learning library, supporting C++, Python, Java, and MATLAB.

The choice of library or framework depends on project specifics, target platform, and performance requirements. Here's a brief example of integrating with TensorFlow Lite in Python:

```
import tensorflow as tf
import numpy as np

# Load the TFLite model and allocate tensors.
interpreter = tf.lite.Interpreter(model_path="model.tflite")
interpreter.allocate_tensors()

# Get input and output tensors.
input_details = interpreter.get_input_details()
output_details = interpreter.get_output_details()

# Prepare input data
input_data = np.array(your_input_data, dtype=np.float32)

# Run inference
interpreter.set_tensor(input_details[0]['index'], input_data)
```

```
interpreter.invoke()

# Get output data
output_data = interpreter.get_tensor(output_details[0]['index'])
print(output_data)
```

5. Security of deep learning models for autonomic devices

In the context of using deep learning models for semantic image segmentation on autonomic devices, ensuring the security of trained models is crucial. Modern approaches to security in this domain focus on hardware-based solutions and comprehensive encryption strategies.

One key component of this security framework is the use of Hardware Security Modules (HSMs) [12]. These specialized physical devices play a vital role in protecting deep learning models by performing cryptographic processing and providing secure key storage. By integrating HSMs into autonomic devices for semantic image segmentation, encryption and decryption operations can be conducted in a secure environment, preventing unauthorized access to trained models and protecting intellectual property.

Complementing HSMs, Trusted Platform Modules (TPMs) offer another layer of hardware-based security [13]. These secure cryptoprocessors store cryptographic keys and perform cryptographic operations, creating a reliable hardware foundation for protection. TPMs ensure that models can only be decrypted through authorized operations, significantly enhancing security even if physical access to the device is obtained.

While hardware solutions provide a strong foundation, a truly comprehensive approach to protecting deep learning models combines robust encryption with stringent access control. Models should be encrypted using advanced algorithms like AES-256, with encryption keys stored securely in HSMs or TPMs. This ensures data integrity even if models are extracted from the device. Building upon this encryption, access control should implement multi-factor authentication (MFA) and a multi-layered authentication and authorization system, restricting access to verified users only [14].

6. Conclusions

Overall, the present study highlights the potential of deep learning models for semantic image segmentation on autonomic devices, enabling visual analysis without constant external computing resources. Key to implementation is the optimization of model size and efficiency through techniques like quantization, pruning, and knowledge distillation, making them suitable for resource-constrained devices. Deployment requires specialized frameworks such as TensorFlow Lite or PyTorch Mobile, with optimization for specific hardware crucial. Security is paramount, necessitating a comprehensive approach combining hardware and software solutions. Future developments focus on improving neural network architectures, developing new optimization methods, and creating specialized hardware accelerators. While challenges persist, the field promises significant advancements in intelligent visual processing systems, opening new possibilities for autonomic devices across various applications.

References

[1] R. Szeliski, Computer vision: algorithms and applications. Springer Nature, 2022. 925 p. doi: 10.1007/978-3-030-34372-9.

[2] Hao S., Zhou Y., Guo Y. A brief survey on semantic segmentation with deep learning. Neurocomputing 406 (2020) 302-321. doi:10.1016/j.neucom.2019.11.118.

[3] Guo Y., Liu Y., Georgiou T., Lew M. A review of semantic segmentation using deep neural networks. Int. J. Multimed. Info Retr. 7 (2018) 87-93. doi: 10.1007/s13735-017-0141-z.

[4] O'Mahony N., Campbell S., Carvalho A., et al. Deep learning vs. traditional computer vision, Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Vol. 11, Springer International Publishing, 2020. P. 128-144. doi:10.1007/978-3-030-17795-9.
[5] Mairittha N., Mairittha T., Inoue S. On-device deep learning inference for efficient activity data collection. Sensors (Basel) 19 (2019) 3434. doi: 10.3390/s19153434.
[6] Cui T. Review of deep learning and mobile edge computing in autonomous driving. Вісник Львівської політехніки 12 (2022) 208-218. doi:10.23939.
[7] Zhang Z., Li J. A review of artificial intelligence in embedded systems. Micromachines 14 (2023) 897. doi:10.3390/mi14050897.
[8] Merone M., Graziosi A., Lapadula V., Petrosino L., d'Angelis O., Vollero L. A practical approach to the analysis and optimization of neural networks on embedded systems. Sensors 22 (2022) 7807. doi:10.3390/s22207807.
[9] Dong K., Zhou C., Rian Y., Li Y. MobileNetV2 model for image classification. 2nd International Conference on Information Technology and Computer Application (ITCA). IEEE, 2020. P. 476-480. doi:10.1109/ITCA52113.2020.00106.
[10] Berthelier A., Chateau T., Duffner S., et al. Deep model compression and architecture optimization for embedded systems: a survey. J. Signal Processing Systems 93 (2021) 863-878. doi:10.1007/s11265-020-01596-1.
[11] Hadidi R., Cao J., Xie Y., et al. Characterizing the deployment of deep neural networks on commercial edge devices. IEEE International Symposium on Workload Characterization (IISWC), IEEE. 2019. P. 35-48. doi:10.1109/IISWC47752.2019.9041955.
[12] Mavrovouniotis S. Ganley M. Hardware security modules. Secure Smart Embedded Devices, Platforms and Applications. New York, NY: Springer New York, 2013. P. 383-405. doi:10.1007/978-1-4614-7915-4_17.
[13] Ezirim K., Khoo W., Koumantaris G., et al. Trusted platform module – a survey. The Graduate Center of The City University of New York, 11. 2012.
[14] Köylü T.Ç., Wedig Reinbrecht C.R., Gebregiorgis A., et al. A survey on machine learning in hardware security. ACM J. Emerg. Techn. Comput. Syst. 19 (2023) 1-37. doi:10.1145/3589506.

DOI: https://doi.org/10.30837/IST-2024.P1.174

Застосування алгоритмів машинного навчання для верифікації та валідації інформаційних систем через виявлення аномалій

Олександр Неліпа[1*] та Надія Калита[1]

[1] *Харківський національний університет радіоелектроніки, Проспект Науки, 14, Харків, 61166, Україна*

Анотація
У статті досліджується інтеграція алгоритмів машинного навчання в процеси верифікації та валідації інформаційних систем. У роботі розглядається взаємозв'язок аномалій із верифікацією та валідацією як ідентифікаторів дефектів та акцентується увага на обмеженнях традиційних методів виявлення аномалій.

Ключові слова
Функціональний дефект, виявлення аномалій, верифікація, валідація, машинне навчання.

1. Вступ

Інформаційні системи є основою критично важливих операцій у різних галузях, зокрема у фінансових послугах, охороні здоров'я, виробництві та державному секторі. Забезпечення надійності, безпеки та операційної стабільності цих систем є надзвичайно важливим, оскільки аномалії в таких середовищах можуть призвести до серйозних наслідків, таких як витоки даних [1], фінансові збитки [2] та переривання надання послуг [3]. Виявлення аномалій є аналітичним процесом, спрямованим на ідентифікацію точок даних, об'єктів або подій, що значно відхиляються від типових моделей або норм [4, 5]. Ця методологія відіграє критичну роль в автоматизованих системах моніторингу, особливо у виявленні потенційно шкідливих відхилень, що дозволяє забезпечити цілісність і безпеку даних [6].

Традиційні підходи до верифікації [7] та валідації [8] зосереджуються на тестуванні та моніторингу систем за допомогою попередньо визначених методів виявлення аномалій на основі правил або статичного аналізу. Хоча ці підходи є фундаментальними, вони стають дедалі менш ефективними для управління великими обсягами динамічно змінюваних даних.

Підхід, що пропонується, ґрунтується на використані алгоритмів машинного навчання для подолання обмежень, властивих методам верифікації та валідації на основі правил і статичного аналізу. Завдяки використанню таких методів, системи можуть не лише виявляти аномалії, що відхиляються від відомих шаблонів, але й адаптуватися та навчатися на основі еволюції даних. Методи, засновані на машинному навчанні, забезпечують гнучкий, адаптивний та проактивний підхід до виявлення аномалій [9], підвищуючи ефективність процесів для складних інформаційних систем.

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
* Corresponding author.
oleksandr.nelipa1@nure.ua (O. Nelipa); nadiia.kalyta@nure.ua (N. Kalyta)
0000-0003-1387-1414 (O. Nelipa); 0000-0001-6181-732X (N. Kalyta)

2. Виявлення аномалій в процесах верифікації та валідації

У контексті процесів верифікації та валідації інформаційних систем «дефект» означає будь-яку непередбачену поведінку, функціональну помилку чи відхилення від очікуваних результатів у системі [10]. У ході таких процесів дефекти ідентифікуються та класифікуються для розуміння їхнього впливу, джерела походження та необхідних дій.

Аномалії – це несподівана поведінка або відхилення від типових патернів системи, що вказують на наявність дефектів [11]. Не всі аномалії є дефектами, але багато з них свідчать про приховані проблеми, які можуть вплинути на надійність, безпеку або продуктивність системи. Аномалії зазвичай поділяються на дві основні категорії:

Статистичні аномалії – це відхилення від усталених патернів чи порогових значень, наприклад, незвично високий час відгуку, пікове використання ресурсів або несподівані патерни мережевого трафіку [11]. Такі аномалії можуть свідчити про проблеми з продуктивністю, помилки в конфігурації або загрози безпеці.

Структурні аномалії – це відхилення від нормальних структур даних, наприклад відсутність полів у записі бази даних або невідповідності у схемах даних [11]. Структурні аномалії часто виявляють проблеми з цілісністю даних або функціональні помилки в системі.

Певні аномалії класифікуються як дефекти, якщо вони впливають на функціональність, продуктивність або надійність системи. Найбільш поширені типи аномалій, які можуть бути визнані дефектами, включають:

Сплески продуктивності – несподівані стрибки часу відгуку, використання пам’яті або навантаження на процесор, які можуть сигналізувати про вузькі місця продуктивності або витоки ресурсів.

Відхилення даних – значення, що виходять далеко за межі очікуваних діапазонів, наприклад, аномально високі суми транзакцій, незвичайні часи входу в систему або неочікувані зміни активності користувачів. Такі аномалії можуть свідчити про пошкодження даних або порушення безпеки.

Несанкціоновані патерни доступу – численні невдалі спроби входу з невідомих IP-адрес або доступ поза звичайними годинами. Такі аномалії часто вказують на дефекти безпеки або спроби несанкціонованого доступу.

Відхилення конфігурації – зміни в конфігураціях систем, такі як модифікації налаштувань безпеки, правил доступу до ресурсів або політик мережі. Конфігураційні аномалії можуть призвести до погіршення продуктивності чи вразливостей безпеки.

3. Обмеження існуючих підходів в процесах верифікації та валідації

Попри те, що традиційні методи верифікації та валідації є ключовими для забезпечення якості та безпеки інформаційних систем, із зростанням масштабу та складності систем вони стикаються зі значними обмеженнями.

Методи виявлення аномалій на основі правил покладаються на заздалегідь визначені умови та статичні порогові значення для ідентифікації аномалій у системі [12]. Хоча вони ефективні для виявлення добре відомих патернів аномальної поведінки, ці підходи мають низку недоліків, такі як негнучкість до змін, висока частка хибнопозитивних результатів, та проблеми з масштабованістю.

Інструменти статичного аналізу оцінюють код і структуру системи для виявлення потенційних вразливостей або відповідності встановленим стандартам безпеки та якості. Подібним чином, попередньо визначені підходи використовують тестові сценарії, розроблені для оцінки очікуваної поведінки системи [13, 14]. Однак ці методи також мають свої обмеження, такі як відсутність динамічного контексту, обмежений обсяг тестування та високі витрати.

Крім того, класичні методи часто залежать від людського досвіду для інтерпретації результатів і адаптації правил. Ця залежність створює вузькі місця у швидко змінюваних середовищах [15], де для обробки аномалій потрібне майже миттєве виявлення та реагування.

4. Виявлення аномалій на основі машинного навчання як удосконалене рішення

Машинне навчання відкриває революційні можливості для процесів верифікації та валідації, дозволяючи системам самостійно навчатися на історичних даних, розпізнавати складні патерни та адаптуватися до нових потенційних аномалій без потреби у заздалегідь визначених і прописаних правилах. Переваги такого підходу включають:

Адаптивність – алгоритми постійно навчаються на вхідних даних [9], що дозволяє їм виявляти як відомі, так і невідомі типи аномалій.

Масштабованість – алгоритми здатні обробляти великі набори даних і складні, багатовимірні структури даних, що робить їх більш придатними для сучасних, орієнтованих на дані систем.

Вибір конкретного алгоритму залежить від природи аномалій, типу вхідних даних та доступності маркованих даних [16, 17]. Найбільш підходящі типи алгоритмів для виявлення аномалій у процесах верифікації та валідації:

Навчання з учителем – доцільне, коли в історичних даних є марковані приклади нормальних і аномальних подій. Наприклад, у фінансових системах відомі патерни шахрайських транзакцій можуть бути використані як марковані дані для навчання моделей з учителем, що дозволить їм виявляти подібні патерни у майбутніх даних.

Навчання без учителя – ідеальне для виявлення нових аномалій без потреби у маркованих даних. Наприклад, у середовищах із динамічними патернами даних, таких як моніторинг мережевого трафіку, де моделі без учителя можуть виявляти раніше невідомі загрози.

Гібридні моделі – ці підходи використовують обмежений набір маркованих даних для покращення точності моделі у сценаріях, коли аномалії є рідкісними, а маркування – дорогим [17].

5. Запропонований підхід впровадження машинного навчання в процеси верифікації та валідації

Для інтеграції машинного навчання у процеси верифікації та валідації з метою виявлення аномалій пропонується структурована методологія, що складається з чотирьох основних етапів.

Збір та попередня обробка даних – ефективне виявлення аномалій залежить від наявності комплексних і якісних даних. Дані мають збиратися з різних джерел, включаючи системні журнали, активність користувачів, записи транзакцій і зовнішню інформацію про загрози. Попередня обробка є важливим кроком для видалення шуму, стандартизації ознак і нормалізації значень, що робить дані придатними для використання у моделях машинного навчання [16, 18].

Вибір відповідної моделі є критично важливим і залежить від доступності маркованих даних та типу аномалій, які потрібно виявити [9]. Для відомих типів аномалій використовуються моделі з учителем, які навчаються на маркованих даних, тоді як у середовищах, де аномалії виникають непередбачувано, застосовуються моделі без учителя.

Після навчання моделі розгортаються у реальних умовах, де вони постійно моніторять вхідні дані. Для підтримання точності використовується механізм зворотного зв’язку,

який дозволяє регулювати чутливість виявлення для мінімізації хибнопозитивних результатів [16], а також періодично перенавчати моделі на основі нещодавніх даних для врахування нових патернів або можливих змін у поведінці системи.

Оцінювання продуктивності є необхідним для забезпечення ефективності систем на основі машинного навчання у виявленні аномалій. Регулярне застосування метрик оцінки моделі, таких як точність, повнота та F1-score [19], допомагає оцінити здатність моделі ідентифікувати аномалії у системі. Для адаптивного навчання важливим є постійний моніторинг даних, оскільки зміни в основних патернах даних можуть вимагати перенавчання моделі для підтримання достатнього рівня точності [17].

Для верифікації та валідації на основі великих мовних моделей для мультимодального виявлення аномалій застосовуються BERT, GPT, LLaMA2, які виконувати широкий спектр завдань – від прогнозування до виявлення аномалій – із залученням різних джерел даних, ефективно інтегруючи мультимодальні потоки для підвищення точності виявлення аномалій.

6. Висновки

Процеси верифікації та валідації є ключовими для забезпечення надійності, точності та безпеки інформаційних систем. Традиційні методи, засновані на правилах, мають обмеження, такі як негнучкість, висока частка хибнопозитивних результатів та проблеми з масштабованістю. Інтеграція машинного навчання дозволяє подолати ці виклики завдяки автоматизованому виявленню аномалій. Моделі здатні навчатися на нових даних, адаптуватися до змін та виявляти складні, непередбачені аномалії.

Запропонований підхід передбачає класифікацію дефектів та обробку багатовимірних потоків даних за допомогою великих мовних моделей.

Таким чином, інтеграція машинного навчання в процеси верифікації та валідації усуває недоліки підходів на основі правил, забезпечуючи гнучкість, масштабованість і точність у виявленні дефектів, що є важливим кроком у розвитку сучасних інформаційних систем.

Література

[1] X. Jin, C. Katsis, F. Sang, J. Sun, A. Kundu, and R. Kompella, "Edge security: Challenges and issues," arXiv preprint, arXiv:2206.07164, 2022.

[2] M. Ahmed, A. N. Mahmood, and M. R. Islam, "A survey of anomaly detection techniques in financial domain," Future Generation Computer Systems, vol. 55, pp. 278–288, 2016.

[3] Z. Sun, C. E. Rubio-Medrano, Z. Zhao, T. Bao, A. Doupé, and G.-J. Ahn, "Understanding and predicting private interactions in underground forums," in Proceedings of the Ninth ACM Conference on Data and Application Security and Privacy, 2019, pp. 303–314.

[4] L. Zhang, D. Lai, and A. V. Miranskyy, "The impact of position errors on crowd simulation," Simulation Modelling Practice and Theory, vol. 90, pp. 45–63, 2019.

[5] W. Dang, B. Zhou, W. Zhang, and S. Hu, "Time Series Anomaly Detection Based on Language Model," in Proceedings of the Eleventh ACM International Conference on Future Energy Systems, E-Energy '20, Association for Computing Machinery, New York, NY, USA, June 2020, pp. 544–547.

[6] W. Dang, B. Zhou, L. Wei, W. Zhang, Z. Yang, and S. Hu, "TS-Bert: Time Series Anomaly Detection via Pre-training Model Bert," in M. Paszynski, D. Kranzlmüller, V. V. Krzhizhanovskaya, J. J. Dongarra, and P. M. A. Sloot (Eds.), Computational Science – ICCS 2021, Lecture Notes in Computer Science, Springer International Publishing, Cham, 2021, pp. 209–223.

[7] ISTQB Standard Glossary of Terms Used in Software Testing: Verification. Available: https://glossary.istqb.org/en_US/term/verification

[8] ISTQB Standard Glossary of Terms Used in Software Testing: Validation. Available: https://glossary.istqb.org/en_US/term/validation
[9] S. K. Devineni, S. Kathiriya, and A. Shende, "Machine learning-powered anomaly detection: Enhancing data security and integrity," Journal of Artificial Intelligence & Cloud Computing, vol. 184, no. 2, pp. 2–9, 2023. doi:10.47363/JAICC/2023
[10] S. Apel, A. von Rhein, P. Wendler, A. Größlinger, and D. Beyer, "Strategies for product-line verification: Case studies and experiments," in Proceedings of the 2013 35th International Conference on Software Engineering (ICSE), San Francisco, CA, USA, 2013, pp. 482–491. doi:10.1109/ICSE.2013.6606594.
[11] Z. Li, X.-Y. Jing, and X. Zhu, "Progress on approaches to software defect prediction," IET Software, vol. 12, 2018. doi:10.1049/iet-sen.2017.0148.
[12] G. Costa, A. Forestiero, and R. Ortale, "Rule-Based Detection of Anomalous Patterns in Device Behavior for Explainable IoT Security," IEEE Transactions on Services Computing, vol. PP, pp. 1-13, 2023. doi:10.1109/TSC.2023.3327822.
[13] J. Zheng, L. Williams, N. Nagappan, W. Snipes, J. P. Hudepohl, and M. A. Vouk, "On the value of static analysis for fault detection in software," IEEE Transactions on Software Engineering, vol. 32, no. 4, pp. 240-253, Apr. 2006. doi:10.1109/TSE.2006.38.
[14] N. Ayewah, W. Pugh, D. Hovemeyer, J. D. Morgenthaler, and J. Penix, "Using Static Analysis to Find Bugs," IEEE Software, vol. 25, no. 5, pp. 22-29, Sept.-Oct. 2008. doi:10.1109/MS.2008.130.
[15] W. L. Oberkampf and T. G. Trucano, "Verification and validation benchmarks," Nuclear Engineering and Design, vol. 238, no. 3, pp. 716-743, 2008. doi:10.1016/j.nucengdes.2007.02.032.
[16] R. Ahamed, A. Habeeb, F. Nasaruddin, A. Gani, I. A. Targio Hashem, E. Ahmed, and M. Imran, "Real-time big data processing for anomaly detection: A Survey," International Journal of Information Management, vol. 45, pp. 289-307, 2019. doi:10.1016/j.ijinfomgt.2018.08.006.
[17] E. Quatrini, F. Costantino, G. Di Gravio, and R. Patriarca, "Machine learning for anomaly detection and process phase classification to improve safety and maintenance activities," Journal of Manufacturing Systems, vol. 56, pp. 117-132, 2020. doi:10.1016/j.jmsy.2020.05.013.
[18] M. Hosseinzadeh, E. Azhir, O. Ahmed, M. Ghafour, S. Ahmed, A. Rahmani, and B. Vo, "Data cleansing mechanisms and approaches for big data analytics: a systematic study," Journal of Ambient Intelligence and Humanized Computing, vol. 14, pp. 1-13, 2021. doi:10.1007/s12652-021-03590-2.
[19] B. J. Erickson and F. Kitamura, "Magician's Corner: 9. Performance Metrics for Machine Learning Models," Radiology: Artificial Intelligence, vol. 3, no. 3, p. e200126, 2021. doi:10.1148/ryai.2021200126.
[20] J. Su, C. Jiang, X. Jin, Y. Qiao, T. Xiao, H. Ma, R. Wei, Z. Jing, J. Xu, and J. Lin, "Large Language Models for Forecasting and Anomaly Detection: A Systematic Literature Review," arXiv, abs/2402.10350, 2024.

Сучасні рекомендаційні системи. Аналіз алгоритмів персоналізації в соціальних мережах та медіаплатформах

Єгор Нестеренко[1*] та Ірина Кириченко[1]

[1] *Харківський національний університет радіоелектроніки, пр. Науки, 14, м. Харків, 61166, Україна*

Анотація

У роботі розглянуто методи та моделі, що застосовуються у рекомендаційних системах популярних медіаплатформ, із фокусом на інтеграцію колаборативної фільтрації, фільтрації на основі вмісту та алгоритмів глибокого навчання. На прикладі платформ TikTok, YouTube та Netflix досліджено ефективність гібридних підходів і архітектур нейронних мереж для підвищення персоналізації та залучення користувачів.

Ключові слова

рекомендаційні системи, аналіз поведінки користувачів, колаборативна фільтрація, фільтрація на основі вмісту, гібридні моделі, глибоке навчання, персоналізований маркетинг

1. Вступ

Технології обробки великих даних та машинного навчання змінюють підходи до персоналізації рекомендацій у маркетингових інформаційних системах. Традиційні методи, зокрема колаборативна та контентна фільтрація, широко застосовуються, однак вони мають обмеження, особливо при роботі з великими та різноманітними наборами даних. Це обмежує точність і релевантність рекомендацій, що особливо помітно у середовищах, де користувацькі вподобання швидко змінюються.

Колаборативна фільтрація дозволяє будувати рекомендації на основі дій схожих користувачів, проте стикається з проблемою “холодного старту” для нових користувачів та продуктів. Фільтрація на основі вмісту, своєю чергою, орієнтована на індивідуальні характеристики контенту, але має схильність до одноманітності рекомендацій, що звужує діапазон пропозицій для користувачів. Для подолання цих обмежень розроблено гібридні методи, які поєднують переваги обох підходів і забезпечують підвищену точність та різноманітність рекомендацій.

Сучасні рекомендаційні системи, побудовані на основі глибокого навчання, здатні враховувати складні залежності між даними, що недоступні для традиційних моделей. Використання нейронних мереж, таких як згорткові та рекурентні, дозволяє враховувати контекст і послідовність дій користувачів, що підвищує точність прогнозування їхніх інтересів. У цій роботі розглядаються архітектури та підходи, які використовують платформи TikTok, YouTube та Netflix для побудови ефективних рекомендаційних систем, здатних адаптуватися до змін у поведінці користувачів і підвищувати їхню взаємодію з платформою.

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
* Corresponding author.
yehor.nesterenko@nure.ua (Є. Нестеренко); iryna.kyrychenko@nure.ua (І. Кириченко)
0009-0007-0263-1323 (Є. Нестеренко); 0000-0002-7686-6439 (І. Кириченко)

2. Методологія роботи рекомендаційних систем

Для розробки ефективних рекомендаційних систем застосовують декілька основних підходів, кожен з яких має свої переваги та обмеження, що впливають на якість персоналізованих рекомендацій.

2.1. Колаборативна фільтрація

Колаборативна фільтрація є одним із найбільш поширених підходів у рекомендаційних системах, що будує рекомендації на основі схожих дій інших користувачів. Вона використовує або підхід на основі користувачів (user-based) або на основі елементів (item-based) для передбачення вподобань [1]. Користувачам із подібними вподобаннями можуть пропонуватися схожі фільми чи продукти, що ґрунтується подібності інтересів користувачів. Колаборативна фільтрація має обмеження, зокрема проблему холодного старту – відсутність достатніх даних для нових користувачів або продуктів, що ускладнює побудову точних рекомендацій [2].

2.2. Фільтрація на основі вмісту

Метод фільтрації на основі вмісту враховує характеристики контенту, який сподобався користувачу, для створення персоналізованих рекомендацій. Цей підхід дозволяє системі пропонувати новий контент на основі подібності характеристик (жанр, ключові слова, категорії), що присутні в улюблених фільмах чи товарах користувача [3]. Хоча такий підхід корисний для індивідуальної персоналізації, він має обмеження у вигляді потенційної одноманітності рекомендацій. Оскільки система орієнтована лише на поточні інтереси користувача, вона часто не виходить за межі його звичного спектру вподобань, що може знизити варіативність контенту. Це явище, відоме як "проблема вузького тунелю" (narrow tunnel problem) [4].

2.3. Гібридні методи

Гібридні рекомендаційні системи об’єднують елементи колаборативної та контентної фільтрації, що дозволяє зменшити обмеження кожного з методів і підвищити точність рекомендаційних пропозицій. Такі системи комбінують кілька методів, щоб отримати максимально релевантний результат для користувача [5]. Наприклад, Netflix використовує гібридний підхід, що поєднує колаборативну фільтрацію, фільтрацію на основі вмісту та додаткові алгоритми машинного навчання, такі як глибоке навчання для врахування історичних даних користувача та інших неявних вподобань [6]. Гібридні методи дозволяють знизити проблему холодного старту, підвищуючи адаптивність системи для нових користувачів і продуктів, а також забезпечують більш різноманітний контент за рахунок об'єднання декількох стратегій [7].

3. Огляд використання методів глибокого навчання в рекомендаційних системах популярних платформ

Рекомендаційні системи застосовують технології глибокого навчання для аналізу поведінки користувачів. Це дозволяє створювати точні персоналізовані рекомендації. Розглянемо особливості таких систем на прикладі TikTok, YouTube та Netflix.

3.1. TikTok

TikTok поєднує колаборативну та контентну фільтрацію. Колаборативна фільтрація аналізує взаємодії користувачів, а контентна – метадані відео, такі як хештеги, підписи та

музичні треки. Це забезпечує динамічність рекомендацій і дозволяє новим авторам досягати аудиторії навіть без великої бази підписників. Водночас, низька прозорість алгоритмів викликає критику через утворення "ехо-камер", що обмежує різноманітність контенту [8].

3.2. YouTube

YouTube реалізує двоступеневу архітектуру: генерацію кандидатів та їх ранжування. На першому етапі система відбирає сотні релевантних відео, використовуючи ембеддинги відео та історії переглядів, а на другому – сортує їх за прогнозованим часом перегляду (expected watch time). Це дозволяє уникати поширення "клікбейтного" контенту, збалансувати рекомендації між новим і популярним відео та враховувати унікальні поведінкові патерни користувачів [9].

3.3. Netflix

Netflix застосовує гібридний підхід, який інтегрує Netflix використовує гібридний підхід із матричною факторизацією, згортковими нейронними мережами (CNN) та алгоритмами ранжування, такими як Personalized Video Ranker (PVR). Матрична факторизація допомагає виявляти латентні вподобання, CNN аналізують візуальний контент, а PVR формує персоналізовані списки, враховуючи історію переглядів та контекст (час доби, місцезнаходження тощо). Така архітектура дозволяє платформі забезпечувати високу релевантність рекомендацій і підвищувати лояльність користувачів.[10]

4. Порівняння підходів і вибір оптимальної системи

Вибір рекомендаційної системи залежить від потреб користувачів, цілі платформ. Наприклад, короткострокові інтереси ефективно обробляє підхід, реалізований на платформі TikTok, де використовуються механізми уваги та рекурентні нейронні мережі, які дозволяють швидко адаптуватися до змін у вподобаннях користувачів. Такий підхід забезпечує високу динаміку рекомендацій, що важливо для платформ з великим обсягом користувацького контенту.

YouTube використовує двоступеневу архітектуру, де перший етап охоплює широкий набір релевантного контенту, а другий етап (ранжування) підбирає найбільш релевантні відео для користувача, враховуючи його довгострокові інтереси. Ця стратегія є особливо корисною для відеоплатформ, де потрібне точне поєднання короткострокових і довгострокових вподобань для підтримки залученості користувачів.

Netflix застосовує гібридні методи, що включають матричну факторизацію та згорткові нейронні мережі. Це дозволяє створювати детальні профілі користувачів і покращувати рекомендації, спираючись на як явні, так і неявні сигнали (наприклад, повторні перегляди). Підхід є оптимальним для платформ, що мають високу частоту повторного використання та вимагають підвищеного рівня персоналізації.

Підходи свою специфіку, і вибір залежить від потреб платформи. Гібридні системи, що поєднують кілька методів, дозволяють компенсувати недоліки окремих підходів, але потребують значних обчислювальних ресурсів та витрат на розробку.

5. Висновки

У роботі було розглянуто ключові методи та підходи, які використовуються у побудові рекомендаційних систем на популярних платформах. Стратегії колаборативної та контентної фільтрації, глибокого навчання, пропонують різні переваги для персоналізації рекомендацій, залежно від типу платформи та її цілей. Сучасні рекомендаційні системи, такі як у TikTok, YouTube та Netflix, демонструють можливості

використання складних моделей глибокого навчання [11] для підвищення точності рекомендацій та утримання користувачів.

Розвиток рекомендаційних систем у маркетингу залежить від здатності обробляти великі обсяги даних у реальному часі, а також від подальшого вдосконалення технологій глибокого навчання, таких як RNN, CNN та механізми уваги. Питання інтерпретованості моделей, етичні аспекти персоналізованих рекомендацій та захист даних користувачів залишаються актуальними й потребують особливої уваги в майбутніх дослідженнях.

Література

[1] Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. Advances in Artificial Intelligence, 2009.

[2] Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th International Conference on World Wide Web (pp. 285-295).

[3] Lops, P., de Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In Recommender Systems Handbook (pp. 73-105). Springer.

[4] Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In The Adaptive Web (pp. 325-341). Springer.

[5] Burke, R. (2002). Hybrid recommender systems: Survey and experiments. User Modeling and User-Adapted Interaction, 12(4), 331-370.

[6] Gomez-Uribe, C. A., & Hunt, N. (2015). The Netflix recommender system: Algorithms, business value, and innovation. ACM Transactions on Management Information Systems (TMIS), 6(4), 1-19.

[7] Jannach, D., & Adomavicius, G. (2016). Hybrid recommender systems. In Recommender Systems Handbook (pp. 331-367).

[8] Zeng, W., Li, H., Chen, G., & Guo, C. (2022). Understanding the TikTok Recommendation System: A Survey and Evaluation. IEEE Transactions on Multimedia.

[9] Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube recommendations. Proceedings of the 10th ACM Conference on Recommender Systems, 191-198.

[10] Gomez-Uribe, C. A., & Hunt, N. (2015). The Netflix recommender system: Algorithms, business value, and innovation. ACM Transactions on Management Information Systems (TMIS), 6(4), 1-19.

[11] Proniuk, G., Geseleva, N., Tereshchenko, G., & Kyrychenko, I. (2019). Spatial interpretation of the notion of relation and its application in the system of artificial intelligence. CEUR-WS, 2362, 266–276.

DOI: https://doi.org/10.30837/IST-2024.P1.183

Оптимізація вагових коефіцієнтів у гібридній моделі користувача за допомогою генетичного алгоритму

Микита Гребенюк[1] та Поліна Ситнікова[1]

[1]*Харківський національний університет радіоелектроніки, пр. Науки, 14, м. Харків, 61166, Україна*

Анотація

Сучасні системи рекомендацій стикаються з проблемою адаптації до динамічних і мінливих уподобань користувачів. Традиційні моделі часто покладаються на фіксовані ваги характеристик, припускаючи, що кожна характеристика однаково впливає на обчислення подібності. Однак це надто спрощує різноманітний і мінливий характер уподобань користувачів, що призводить до неоптимальних рекомендацій. У нашій попередній роботі ми представили гібридну модель користувача, яка поєднує демографічні, колаборативні та контентні для підвищення точності рекомендацій. Ця стаття базується на цій моделі, пропонуючи рішення для динамічного коригування ваг характеристик за допомогою генетичного алгоритму (GA). GA пропонує ефективний підхід для оптимізації вагових показників за допомогою еволюції популяції потенційних рішень шляхом відбору, кросинговеру та мутації. Уточнюючи ваги функцій з часом, система може краще фіксувати індивідуальні пріоритети та вподобання користувачів, підвищуючи релевантність рекомендацій. Функція придатності розроблена для оцінки якості рекомендацій, а GA ітеративно коригує коефіцієнти на основі оцінок користувачів. Наш підхід усуває розрив між статичним і динамічним вираховуванням коефіцієнтів системах рекомендацій, створюючи більш персоналізовані та чутливі моделі. Запропонована модель, керована GA, пропонує значний потенціал для підвищення точності рекомендацій і задоволеності користувачів, що робить її цінним внеском у сферу систем рекомендацій.

Ключові слова

Генетичний алгоритм, система рекомендацій, гібридна модель користувача, ваговий коефіцієнт

1. Вступ

Сучасні системи рекомендацій все частіше використовують гібридні моделі, щоб охопити переваги користувачів шляхом поєднання різноманітних характеристик, таких як демографічна інформація, історія оцінок та метадані вмісту. У нашій попередній роботі ми представили гібридну модель рекомендацій [1], яка використовує ці багатоаспектні характеристики для підвищення персоналізації та точності рекомендацій. Проте невід'ємною проблемою для таких моделей є однакова вага всіх характеристик, яка не завжди відображає реальні переваги користувачів. Користувачі часто по-різному оцінюють різні аспекти, і ці переваги змінюються з часом. Щоб вирішити цю проблему, оптимізація вагових показників має важливе значення.

Модель, представлена в нашому попередньому дослідженні, разом з іншими подібними гібридними моделями, добре підходить для оптимізації за допомогою генетичного алгоритму [2] (GA). GA пропонує потужний механізм для динамічного коригування вагових коефіцієнтів характеристик, більш точно узгоджуючи рекомендації з мінливими уподобаннями користувачів. Застосовуючи GA, ми можемо неодноразово уточнювати важливість кожної характеристики, підвищуючи гнучкість і адаптивність

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
mykyta.hrebeniuk@nure.ua (M. Hrebeniuk); polina.sytnikova@nure.ua (P. Sytnikova)
0009-0008-0989-7957 (M. Hrebeniuk); 0000-0002-6688-4641 (P. Sytnikova)

моделі рекомендацій. Цей підхід є перспективним для підвищення задоволеності користувачів і релевантності рекомендацій, оскільки дозволяє адаптувати важливість характеристик відповідно до індивідуальних уподобань.

2. Постановка проблеми

У звичайних системах рекомендацій переваги користувача часто моделюються з використанням фіксованих вагових коефіцієнтів, призначених різним характеристикам, припускаючи, що кожна з них однаково впливає на обчислення подібності. Це припущення надто спрощує поведінку користувачів і не враховує унікальний і динамічний характер індивідуальних уподобань. На практиці користувачі визначають пріоритетність різних аспектів на основі своїх особистих інтересів і контекстуальних факторів, які можуть змінюватися з часом. Ігнорування цих різноманітних пріоритетів може призвести до неоптимальних рекомендацій, які не повністю відповідають очікуванням користувачів, що зрештою знижує ефективність системи.

Основна проблема полягає в динамічному налаштуванні вагових коефіцієнтів [3] таким чином, щоб відображати поточні пріоритети кожного користувача. Моделі з фіксованою вагою не можуть врахувати цю мінливість, особливо коли уподобання користувачів змінюються поступово або навіть раптово з часом. Наприклад, на платформах медіа-рекомендацій користувачі можуть тимчасово зосередитися на певному жанрі чи темі, що потребує гнучкої моделі, яка може адаптуватися до цих тенденцій. Крім того, точне обчислення подібності між користувачами зі статичною моделлю часто призводить до погано узгоджених рекомендацій, оскільки воно не може врахувати відносну важливість окремих характеристик, що призводить до узагальненого та менш персоналізованого результату.

Щоб вирішити цю проблему, ми пропонуємо використовувати генетичний алгоритм (GA) для динамічної оптимізації ваги характеристик у моделі користувача. GA — це адаптивні евристичні алгоритми пошуку, засновані на принципах природного відбору та генетики. Розвиваючи популяцію потенційних наборів ваг за допомогою відбору, схрещування та мутації, GA здатні визначати оптимальні або майже оптимальні рішення в складних, багатовимірних просторах пошуку. Застосування цього підходу до систем рекомендацій може дати нам можливість точно налаштувати вагу кожної характеристики, створивши більш чуйну модель, яка відповідає вподобанням користувачів у реальному світі.

Основними цілями цього дослідження є:

- Розробка гібридної моделі користувача для систем рекомендацій, де вагові коефіцієнти характеристик адаптуються на до поведінки та вподобань користувача.
- Запровадження механізму оптимізації на основі GA для повторного уточнення вагових коефіцієнтів характеристик, покращуючи узгодженість моделі з мінливими інтересами користувачів.
- Визначення ефективної функції придатності для GA, яка оцінює якість рекомендацій шляхом вимірювання подібності між прогнозованими та фактичними уподобаннями користувача.

Це дослідження спрямоване на усунення ігнорування переваг користувача в існуючих моделях рекомендацій шляхом впровадження адаптивного підходу до оптимізації ваги за допомогою GA. Усунення обмеження статичних ваг і інтеграція гнучкої системи, яка розвивається з часом, допоможе підвищити точність рекомендацій.

3. Генетичний алгоритм для оптимізації коефіцієнтів

Запропоноване рішення інтегрує GA для динамічного коригування вагових коефіцієнтів для кожної характеристики в моделі користувача. Процес GA починається з випадково згенерованої популяції індивідуумів, де кожен представляє потенційне рішення [4] — у цьому випадку набір вагових коефіцієнтів характеристик. Кожен індивідуум потім оцінюється на основі функції придатності, розробленої для вимірювання ефективності заданих значень у покращенні точності рекомендацій.

Під час кожної ітерації або покоління індивідууми оцінюються, і більш «придатні» вибираються для створення наступного покоління за допомогою генетичних операторів, таких як схрещування і мутація. Схрещування поєднує ознаки двох індивідуумів для створення нових, тоді як мутація вносить невеликі зміни для збільшення різноманітності в популяції. Процес продовжується ітеративно, доки не буде виконано умову завершення, як правило, коли знайдено оптимальний або майже оптимальний набір ваг.

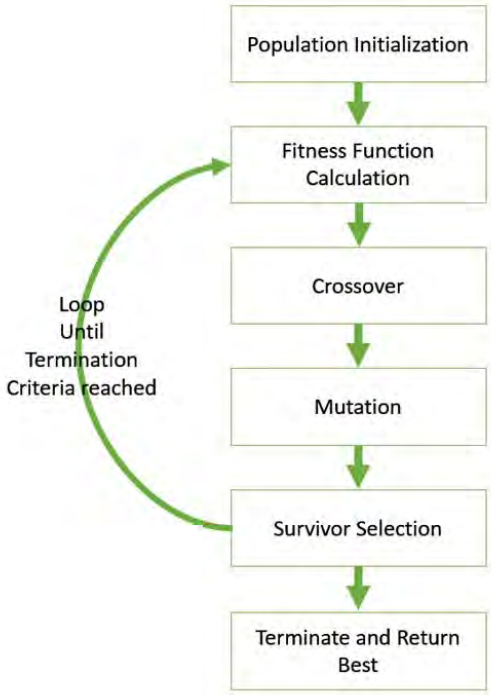


Рисунок 3. Діаграма роботи генетичного алгоритму.

У цій системі функція придатності має вирішальне значення, оскільки вона визначає, наскільки кожен набір ваг покращує точність рекомендацій. Щоб спростити процес, завдання переформульовано як контрольоване навчання. Фактичні рейтинги користувачів поділяються на тестовий і навчальний набір. Система рекомендацій генерує прогнози для навчального набору, а оцінка придатності обчислюється як середня різниця між фактичним і прогнозованим рейтингом для всіх елементів у навчальному наборі.

Математично, оцінка придатності для активного користувача u_a обчислюється як:

$$fitness(u_a) = \frac{\sum_{j=1}^{|S_a^{TE}|} |r_{a,j} - pr_{a,j}|}{|S_a^{TE}|} \quad (1)$$

де $|S_a^{TE}|$ – потужність навчального набору для активного користувача, $r_{a,j}$ – фактична оцінка, і $pr_{a,j}$ прогнозована оцінка для елемента j. Цей показник служить мірою придатності для набору ваги, причому менші відмінності вказують на кращу прогнозовану ефективність.

Глобальна функція нечіткої відстані [1], що включає вагові коефіцієнти, визначені GA, має такий вигляд:

$$fdis(\boldsymbol{u_x}, \boldsymbol{u_y}) = \frac{\sum_{i=1}^{N} w_i \times fdis(x_i, y_i)}{N} \quad (2)$$

де w_i - ваговий коефіцієнт для характеристики i, і N – кількість характеристик. За допомогою цієї формули (2) GA адаптує вагові коефіцієнти кожної характеристики, щоб динамічно відображати пріоритети користувачів. Набір вагових коефіцієнтів користувача

$$weight(u_a) = (w_1, \dots, w_n) \quad (3)$$

де n – кількість характеристик, представлена у вигляді вектора дійсних коефіцієнтів. Призначаючи низьку вагу для менш відповідних характеристик, система ефективно адаптує вибір функцій відповідно до індивідуальних уподобань користувача, гарантуючи, що система рекомендацій залишається чутливою до зміни смаків.

Ця адаптивна модель, керована GA, підвищує точність рекомендацій шляхом постійного вдосконалення вагових показників, що призводить до більш точних персоналізованих рекомендацій.

4. Висновки

У цій роботі ми представили підхід на основі генетичного алгоритму (GA) для динамічної оптимізації ваг характеристик у гібридних системах рекомендацій. Долаючи обмеження моделей із фіксованою вагою, наш метод дозволяє системі рекомендацій краще адаптуватися до мінливих уподобань користувачів. Використання GA для коригування ваги характеристик дає більш персоналізовані та точні рекомендації, покращуючи загальну взаємодію з користувачем. Наш підхід демонструє потенціал покращення релевантності рекомендацій шляхом постійного вдосконалення системи на основі індивідуальних уподобань користувачів. Майбутні дослідження можуть бути зосереджені на оптимізації параметрів GA та тестуванні моделі в різних областях застосування, щоб оцінити її ширшу застосовність.

Список літератури

[1] N. Khairova, N. Sharonova, D. Sytnikov, M. Hrebeniuk, P. Sytnikova, Recommendation System Based on a Compact Hybrid User Model Using Fuzzy Logic Algorithms, Modeling, Optimization, and Controlling in Information and Technology Systems Workshop (MOCITSW-CoLInS 2024) at 8th COLINS 2024 COMPUTATIONAL LINGUISTICS AND INTELLIGENT SYSTEMS, Volume II, Lviv, Ukraine, April 12-13, 2024, Vol-3668 48-63 urn:nbn:de:0074-3668-5.

[2] D. Roy, M. Dutta, Hybrid Measuring the Similarity Value Based on Genetic Algorithm for Improving Prediction in A Collaborative Filtering Recommendation System.: Recommendation System, Advances in Distributed Computing and Artificial Intelligence

[3] Journal Regular Issue, Vol. 10 N. 2 (2021), 165-182. doi: 10.14201/ADCAIJ2021102165182

[4] Longquan Tao, Jinli Cao, Fei Liu, Dynamic feature weighting based on user preference sensitivity for recommender systems, Knowledge-Based Systems, Vol. 149, 2018, 61-75. doi: 10.1016/j.knosys.2018.02.019.

[5] Savio D Immanuel; Udit Kr. Chakraborty, Genetic Algorithm: An Approach on Optimization, 2019 International Conference on Communication and Electronics Systems (ICCES), 2019, 700-708. doi: 10.1109/ICCES45898.2019.9002372.

DOI: https://doi.org/10.30837/IST-2024.P1.187

Визначення факторів які впливають на рівень професійного вигорання медичних працівників*

Данило Чигрин[1,*,†], Ігор Гребеннік [1,†], Ігор Завгородній[2,†], Олена Літовченко[2,†] та Марина Лисак[2,†]

[1] *Харківський національний університет радіоелектроніки, пр. Науки, 14, 61166, Харків, Україна*
[2] *Харківський національний медичний університет, пр. Науки, 4, 61022, Харків, Україна*

Abstract
Ця робота присвячена визначенню інформативних показників при аналізі рівня професійного вигорання серед робітників, з особливим акцентом на медичних працівників. Професійне вигорання визначається як стан емоційного, фізичного та ментального виснаження, що виникає в результаті тривалої взаємодії з високим рівнем стресу на робочому місці. В роботі розглядаються основні аспекти та шкали професійного вигорання, методи його діагностування, особливості вигорання серед лікарів різних спеціалізацій та експериментальний розрахунок рівня вигорання на лікарях трьох напрямів.

Keywords
професійне вигорання, медичні працівники, стрес на робочому місці, медичні спеціалізації, емоційне виснаження, психометричні тести

1. Загальні відомості про професійне вигорання

Професійне вигорання — це стан емоційного, фізичного та ментального виснаження, викликаний тривалою та інтенсивною стресовою роботою. Поняття професійного вигорання було вперше введено американським психологом Гербертом Фройденбергером у 1970-х роках. Він визначив вигорання як «стан розумового та фізичного виснаження, спричиненого професійним життям» [1]. Усесвітня організація охорони здоров'я підсумовує вигорання як «синдром, концептуалізований як результат хронічного стресу на робочому місці, із яким не вдалося успішно впоратися» [2].

Існують три ключові шкали професійного вигорання:

- Емоційне виснаження (ee) — відчуття, що емоційні ресурси вичерпані, що може призвести до апатії та втрати інтересу до роботи.
- Деперсоналізація (zy) — розвиток цинічного ставлення до виконуваної роботи та людей, з якими працює особа.
- Зниження професійної ефективності (ef) — відчуття некомпетентності та неефективності у професійній діяльності [3].

Причини професійного вигорання можуть включати:

- Високі вимоги до роботи та мало контролю над робочим процесом.

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
* Corresponding author.
† These authors contributed equally.
danylo.chigrin@gmail.com (Д. Чигрин); igor.grebennik@nure.ua (І. Гребеннік); zavnikua@gmail.com (І. Завгородній); latyshkaelena@gmail.com (О. Литовченко); marynalysak11@gmail.com (М. Лисак)
0009-0003-3228-4503 (Д. Чигрин); 0000-0003-3716-9638 (І. Гребеннік); 0000-0001-7803-3505 (І. Завгородній); 0000-0002-5286-1705 (О. Литовченко); 0000-0002-5891-1531 (М. Лисак)

- Недостатність винагороди (емоційної, фінансової, соціальної).
- Відсутність справедливості на робочому місці.
- Конфлікти у відносинах з колегами або керівництвом.
- Невизначеність ролі та обов'язків.

Розуміння та визнання цих факторів є важливим кроком у запобіганні та лікуванні професійного вигорання. Важливо створити середовище, де працівники можуть відчувати підтримку та визнання своїх зусиль, а також мати можливість для професійного розвитку та особистісного зростання. Але оскільки професійне вигорання має індивідуальні прояви і від людини до людини проявляється по-різному, то насамперед постає задача визначення факту наявності професійного вигорання у робітника задля відокремлення цього стану від інших та підтвердження наявності такої проблеми.

2. Діагностування професійного вигорання

Діагностування професійного вигорання є важливим кроком у виявленні та лікуванні цього стану. Існує кілька методів та інструментів, які можуть бути використані для оцінки ступеня вигорання. Професійне вигорання часто визначається через тривале спостереження за поведінкою та емоційним станом особи. Основні симптоми включають емоційне виснаження, цинізм або деперсоналізацію, та зниження професійної ефективності. Ці ознаки можуть бути виявлені через безпосереднє спілкування або за допомогою психологічних тестів.

Одним з найпопулярніших інструментів для вимірювання професійного вигорання є інвентар вигорання Маслаха (Maslach Burnout Inventory, MBI). Цей тест включає різні шкали для оцінки трьох вищезгаданих шкал вигорання. Крім того, існують інші тести, такі як Copenhagen Burnout Inventory (CBI) та Oldenburg Burnout Inventory (OLBI), які також широко використовуються для діагностики вигорання.

MBI вважається золотим стандартом у діагностиці професійного вигорання і має високу надійність та валідність. В роботі [4] цей метод було допрацьовано для визначення не лише групи вигорання, але і так званої групи препатології, яка дозволяє визначити ті шкали професійного вигорання, по значенню яких людина знаходиться на початковому етапі цього процесу. CBI та OLBI є більш доступними і легкими у використанні, але можуть не завжди відображати всі аспекти вигорання, як MBI [5].

Важливо обрати метод, який найкраще відповідає конкретним потребам дослідження та доступним ресурсам. Також необхідно звертати увагу на культурну адаптацію тестів, оскільки різні культурні групи можуть сприймати та виражати симптоми вигорання по-різному.

3. Особливості професійного вигорання лікарів

Професійне вигорання серед медичних працівників становить значну проблему, яка впливає на ефективність медичного обслуговування та добробут самого медичного персоналу. Лікарі, які ведуть активну медичну практику, знаходяться під високим ризиком виникнення професійного вигорання через інтенсивність робочого процесу та емоційне навантаження.

Статистика та особливості вигорання серед медичних працівників: За даними досліджень, професійне вигорання може торкнутися до половини медичних працівників у деяких регіонах та спеціалізаціях. Це явище частіше спостерігається серед лікарів вищих спеціалізацій, таких як хірурги та анестезіологи, де рівень стресу та відповідальності особливо високий.

Особливості вигорання за медичними спеціалізаціями:

- Хірурги: Високий рівень стресу, пов'язаний з необхідністю проведення складних операцій та відповідальністю за життя пацієнтів.
- Анестезіологи: Емоційне навантаження через постійну участь у критичних медичних ситуаціях та управління болем пацієнтів.
- Терапевти: Стрес від тривалого спілкування з великою кількістю пацієнтів та необхідність управління хронічними захворюваннями.
- Педіатри: Емоційне виснаження, пов'язане з роботою з дітьми, особливо у випадках серйозних захворювань.

Кожна з цих спеціалізацій має свої унікальні виклики, які можуть сприяти розвитку професійного вигорання. Важливо враховувати ці особливості при аналізі загального стану професійного здоров'я медичних працівників [6].

Розуміння специфіки професійного вигорання в різних медичних спеціалізаціях дозволяє краще оцінити ризики та потреби медичних працівників. Це знання є ключовим для розробки стратегій підтримки та зниження рівня вигорання серед медичного персоналу.

4. Експериментальні дослідження

Для аналізу було проведено опитування трьох груп лікарів різних напрямків: онкологи (37 людей), анестезіологи (73 людини) та лікарі швидкої допомоги (). Для кожної з груп було визначено рівень вигорання згідно таблицею балів [4]. Загальну візуалізацію в просторі головних компонент (PCA – Principal Component Analysis) представлено на рисунку 1. Слід зазначити, що групи перетинаються в просторі ознак і найбільший інтерес представляє група препатології, для якою вирішено визначити ті інформативні питання та шкали, що призводять до початку вигорання в залежності від спеціалізації лікаря.

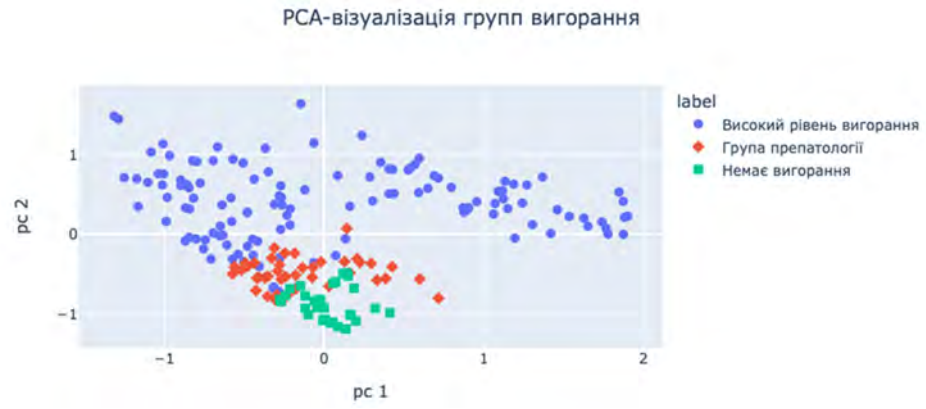


Рисунок 4: PCA-візуалізація груп вигорання серед лікарів.

Для цього було проведено аналіз однорідності груп за віком, стажем роботи та статтю, який показав, що всі три групи є однорідними. Далі було проведено розробку моделі класифікації для кожної групи медичних працівників. Навчання моделі проводилось за допомогою мови програмування Python 3.9 (бібліотека sklearn.neural_network) шляхом кроссвалідації (k-fold=5). Інформативні показники було визначено шляхом використання пояснювача SHAP (SHapley Additive exPlanations) [7]. В результаті було отримано:

- Онкологи: точність моделі класифікації склала 0.83 ± 0.08. Інформативні питання: 6 (шкала ee), 11 (шкала ef), 12 (шкала ef), 14 (шкала zy).

- Анестезіологи: точність моделі класифікації склала 0.85 ± 0.12. Інформативні питання: 11 (шкала ef), 12 (шкала ef), 14 (шкала zy), 16 (шкала ef).
- Лікарі швидкої допомоги: 0.80 ± 0.14. Інформативні питання: 1 (шкала ee), 8 (шкала zy), 14 (шкала zy), 13 (шкала zy).

При дослідженні групи медичних працівників без конкретизації спеціальності було отримано точність моделі на рівні 0.89 ± 0.08 та до найбільш інформативних питань відносились питання зі шкал емоційного виснаження та деперсоналізації: 6 ee (я відчуваю себе вигорілим від роботи); 8 zy (я став менше цікавитися своєю роботою з тих пір, як почав цю роботу); 13 zy (я просто хочу робити свою роботу і щоб мене не турбували); 14 zy (я став більш цинічним щодо того, чи впливає на щось моя робота).

5. Висновки

Запропонована модель визначення інформативних ознак, що призводять до професійного вигорання, яка уможливила раннє виявлення, надаючи цінну інформацію про комплекс факторів, що сприяють його розвитку. Наші висновки показали, що загальна група препатології складалася з 43 медичних працівників, які демонстрували виражений феномен цинізму – вигорання за шкалою деперсоналізації, при тому окремі групи медичних працівників мали свої специфічні ознаки вигорання. Аналізуючи розподіл загальної вибірки за спеціальностями, ми визначили специфічні критерії оцінки для кожної групи та визначили унікальні характеристики професійного вигорання серед лікарів анестезіологів, онкологів та працівників швидкої медичної допомоги.

References

[1] Freudenberger, H. J. (1989). Burnout: Past, Present, and Future Concerns. Loss, Grief & Care, 3(1–2), 1–10. https://doi.org/10.1300/J132v03n01_01.

[2] Mental Health, Brain Health and Substance Use (MSD). (2024, March 8). Clinical descriptions and diagnostic requirements for ICD-11 mental, behavioural and neurodevelopmental disorders (CDDR). https://www.who.int/publications/i/item/9789240077263.

[3] Maslach, Christina & Jackson, Susan & Leiter, Michael. (1997). The Maslach Burnout Inventory Manual.

[4] Böckelmann, I., Perova, I.G., Lalimenko, O.S. et al. Neue Wege in der Früherkennung des Burnout-Risikos bei Notärzten und Feldscherern in einem ukrainischen Rettungsdienst. Zbl Arbeitsmed 73, 160–169 (2023). https://doi.org/10.1007/s40664-023-00501-4

[5] Williamson, Kelly et al. “Comparing the Maslach Burnout Inventory to Other Well-Being Instruments in Emergency Medicine Residents.” Journal of graduate medical education vol. 10,5 (2018): 532-536. doi:10.4300/JGME-D-18-00155.1.

[6] Oliinyk P, Chaplyk V, Timchenko Y. PROFESSIONAL BURNOUT IN HEALTHCARE WORKERS: SIGNS, CAUSES, PREVENTION AND TREATMENT. Proc Shevchenko Sci Soc Med Sci [Internet]. 2022 Jun.27 [cited 2024Nov.17]; 66(1). Available from: https://mspsss.org.ua/index.php/journal/article/view/650

[7] SHAP documentation, 2018. URL: https://shap.readthedocs.io/en/latest/ .

DOI: https://doi.org/10.30837/IST-2024.P1.191

JSBSim як інструмент моделювання динаміки польотів із використанням методів машинного навчання

Олексій Винокур[1] та Ірина Перова[1]

[1] *Харківський університет радіоелектроніки, Харків, Україна*

Анотація

JSBSim є сучасною платформою з відкритим кодом для моделювання динаміки польотів літальних апаратів із шістьма ступенями свободи (6DOF). Робота присвячена оцінці можливостей JSBSim як симулятора для розробки та тестування систем автоматичного приземлення літальних апаратів із застосуванням методів машинного навчання. У дослідженні використано алгоритми підкріплювального навчання (Reinforcement Learning), зокрема DQN і PPO, а також рекурентні нейронні мережі LSTM для обробки часових рядів. Результати дослідження демонструють ефективність обраного підходу, підкреслюючи значний потенціал JSBSim у дослідженнях автономних систем керування польотами.

Ключові слова

машинне навчання, літальні апарати, симуляція польоту, динаміка польоту, нейронні мережі, LSTM, JSBSim, 6DOF

1. Вступ

Розвиток автономних літальних апаратів є одним із ключових напрямів сучасної авіації та космічної галузі. Особливе значення має автоматизація критичних етапів, таких як приземлення, що вимагає високої точності та стабільності. Моделювання цих процесів у реальних умовах часто є дорогим та небезпечним. Тому симуляційні середовища, такі як JSBSim, стають важливим інструментом для тестування нових алгоритмів та систем керування.

JSBSim — це платформа для моделювання динаміки польотів із відкритим кодом, що дозволяє відтворювати аеродинамічні характеристики літальних апаратів із шістьма ступенями свободи (6DOF) [1]. Її гнучкість та підтримка зовнішніх бібліотек дають змогу інтегрувати сучасні алгоритми машинного навчання, зокрема підкріплювального навчання та рекурентні нейронні мережі, для створення автономних систем керування польотами [3; 6]

Мета цієї роботи — дослідити можливості JSBSim у контексті розробки алгоритмів для автоматичного приземлення з використанням методів підкріплювального навчання.

2. Методологія дослідження

Для моделювання динаміки польоту літальних апаратів використовувався JSBSim. Симуляція враховувала вплив зовнішніх сил (гравітації, турбулентності) та роботу двигунів. Аеродинамічні параметри задавалися через конфігураційні файли JSBSim, що дозволяло налаштовувати симуляцію для різних моделей літальних апаратів [1].

Information Systems and Technologies (IST-2024), November 26-28, 2024, Kharkiv, Ukraine
oleksii.vynokur@nure.ua (O. Vynokur); rikywenok@gmail.com (I. Perova);
0009-0001-4328-3886 (O. Vynokur); 0000-0003-2089-5609 (I. Perova);

JSBSim інтегрувався з бібліотекою OpenAI Gym для створення тренувального середовища [4]. Використовувалися такі алгоритми:

- Deep Q-Network (DQN) для роботи з дискретними діями (зміна тяги або кута нахилу) [3].
- Proximal Policy Optimization (PPO) для моделювання безперервних дій, що дозволяє точніше керувати маневрами під час посадки [5].
- Для обробки часових рядів (вхідних даних, таких як висота, швидкість, кут нахилу, тяга двигунів та вплив вітру) застосовувалися рекурентні нейронні мережі (LSTM). Мережа моделює динамічні процеси з довготривалими залежностями, що є критичними для точного управління літальними апаратами.

Нейронна мережа складалася з двох LSTM-шарів з 128 нейронами кожен, за якими слідували повнозв'язні шари [7]. Мережа передбачала оптимальні параметри для приземлення літального апарата на основі поточних і попередніх станів системи.

3. Результати дослідження

Результати симуляції показали, що використання алгоритмів підкріплювального навчання дозволяє досягти стабільного і точного приземлення літального апарата в умовах симуляції.

- Алгоритм PPO показав ефективність у виконанні маневрів, таких як зміна тяги. Зокрема, модель досягла високих значень швидкості (близько 180 вузлів) при максимальній тязі, що дозволило стабільно підтримувати швидкість на рівні, необхідному для посадки [5].
- LSTM ефективно враховувала залежності між послідовними станами літального апарата, що допомогло підтримувати стабільний процес приземлення, незважаючи на коливання висоти та зміну зовнішніх параметрів, таких як турбулентність [7].

Layer (type)	Output Shape	Param #
lstm_46 (LSTM)	(None, 10, 128)	68,608
lstm_47 (LSTM)	(None, 128)	131,584
dense_23 (Dense)	(None, 5)	645

Рисунок: 1 Структура нейронної мережі LSTM

Низка тестів, у яких використовувалася максимальна тяга (throttle = 1.0), показала, що літальний апарат підтримував стабільний режим на висоті близько 1000 футів з помірним зниженням висоти. Швидкість літального апарата (vc-kts) постійно зростала від початкових значень до 180 вузлів, що є критично важливим для ефективного контролю.

Платформа JSBSim забезпечила реалістичне моделювання польоту з урахуванням складних фізичних процесів, таких як взаємодія з атмосферними умовами та робота системи керування. Тренування моделей вимагало значних обчислювальних ресурсів, особливо для LSTM-мереж, але результати тренування PPO та LSTM підтвердили доцільність цього підходу для автономних систем керування польотами [1; 6] [8].

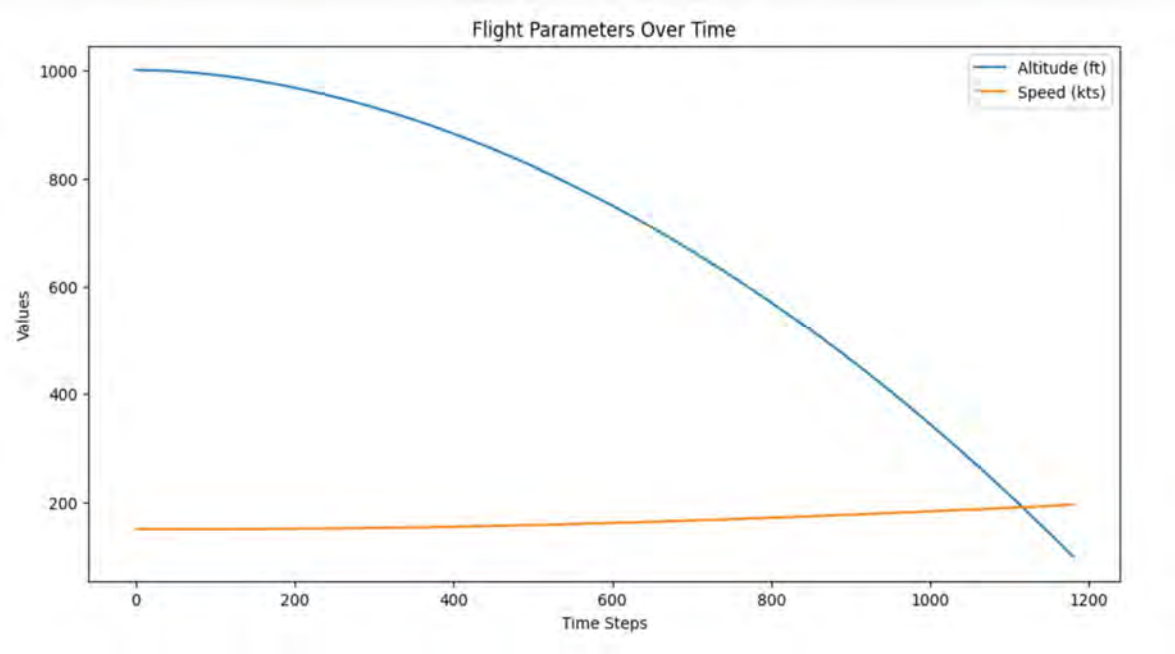


Рисунок: 2 Зміна параметрів висоти та каліброваної швидкості протягом приземлення

4. Обговорення

Результати дослідження підкреслили важливість інтеграції методів машинного навчання з точними симуляційними платформами, такими як JSBSim. Використання LSTM дозволило врахувати часові залежності в поведінці літального апарата, що є критичним для точного управління в реальних умовах [7].

Основні виклики включають:

- Потрібно більше даних для тренування нейронних мереж, що вимагає значних обчислювальних ресурсів.
- Складність налаштування параметрів симуляції для конкретних задач (наприклад, зміни умов вітру або турбулентності).

У подальших дослідженнях планується використання інших функцій активації в нейронних мережах, а також модифікація кількості шарів і кількості нейронів у мережі для підвищення ефективності моделювання [5; 9]. Це дозволить покращити точність та стабільність автономного управління в більш складних умовах.

5. Висновки

JSBSim є потужним інструментом для дослідження автономних систем польотів. Інтеграція з методами машинного навчання дозволяє створювати високоточні системи автоматизації, зокрема для приземлення літальних апаратів. Подальші дослідження мають бути спрямовані на розробку більш ефективних алгоритмів та покращення обчислювальної ефективності, що дозволить реалізувати ці підходи в реальних умовах.

Успішна інтеграція алгоритмів підкріплювального навчання (PPO) та рекурентних нейронних мереж (LSTM) в єдину програму дозволила створити автономну систему керування польотом, що забезпечує стабільне та точне приземлення, що є основним досягненням цього дослідження.

Список використаних джерел та посилань

[1] JSBSim Documentation. JSBSim Flight Dynamics Model: The Open Source Flight Dynamics Model for Use in Aerospace Engineering, Research, and Education. [Електронний ресурс]. – Доступно за посиланням: http://jsbsim.sourceforge.net.

[2] Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction. – 2nd Edition, MIT Press. – 552 с.

[3] Lillicrap, T. P., Hunt, J. J., Pritzel, A., et al. (2016). Continuous control with deep reinforcement learning. – arXiv preprint, arXiv:1509.02971.

[4] OpenAI. (2018). OpenAI Gym: Toolkit for Developing and Comparing Reinforcement Learning Algorithms. [Електронний ресурс]. – Доступно за посиланням: https://gym.openai.com.

[5] Kochenderfer, M. J., & Wheeler, T. A. (2019). Algorithms for Optimization. – MIT Press.

[6] FlightGear. (2024). Flight Simulation with JSBSim. [Електронний ресурс]. – Доступно за посиланням: https://www.flightgear.org.

[7] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. – MIT Press.

[8] Tieleman, T., & Hinton, G. (2012). Lecture 6.5—RMSProp: Divide the gradient by a running average of its recent magnitude. – COURSERA: Neural Networks for Machine Learning.

[9] Копитин, О. В., Кравченко, І. В. (2022). Розробка системи керування БПЛА на основі методів машинного навчання. – Наукові праці Національного авіаційного університету, 2(61), 123-129.

[10] Nguyen, Q., Nguyen, T., & Hwang, D. (2020). Autonomous landing of UAVs using reinforcement learning. – IEEE Transactions on Aerospace and Electronic Systems, 56(4), 2902-2913.

[11] Munos, R., et al. (2016). Safe and efficient off-policy reinforcement learning. – arXiv preprint, arXiv:1606.02647.

Зміст

Наукове видання

«Інформаційні системи та технології» ІСТ-2024

МАТЕРІАЛИ
13-ї Міжнародної науково-технічної конференції
Частина 1

26 – 28 листопада року

Харків, Україна

«INFORMATION SYSTEMS AND TECHNOLOGIES»
IST-2024

Proceedings
of the 13th International Scientific and Technical Conference
Part 1

November 26 – 28, 2024

Kharkiv, Ukraine

Наукові редактори: Ю.О. Романенков, В.В. Безкоровайний, S. Yakovlev, L. Petryshyn, З.В. Дудар, В.Г. Кобзєв
Коректори – М.С. Широкопетлєва, О.О. Олійник
Комп'ютерна верстка – Ю.В. Міщеряков, А.Ю. Міщеряков

Харківський національний університет радіоелектроніки
61166, Харків, пр. Науки, 14,
ХНУРЕ

Електронне видання