

ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ АЛГОРИТМІВ КЛАСТЕРИЗАЦІЇ В АНАЛІЗІ МЕДИЧНИХ ДАНИХ

Мірошниченко Н. С.

Науковий керівник – д. т. н., проф. Перова І. Г.

Харківський національний університет радіоелектроніки, каф. СТ
м. Харків, Україна

Clustering algorithms utilized in medical data analysis segment entities or patients based on shared attributes, facilitating the identification of patient subgroups with similar characteristics and the discovery of comparable treatment outcomes. Moreover, cluster analysis enhances understanding of the data's organizational structure, potentially reflecting the inherent organization of particular diseases or risk factors. The utilization of clustering algorithms encourages the exploration of innovative approaches to diagnosis, treatment, and disease management.

У медичному аналізі даних для виявлення структурних груп та створення класифікаційних схем застосовуються різноманітні алгоритми кластеризації. Серед них важливим є агломеративний метод ієрархічної кластеризації, який базується на послідовному об'єднанні подібних об'єктів у групи. Цей підхід дозволяє поетапно формувати ієрархічну структуру кластерів, де кожен об'єкт спочатку розглядається як окремий кластер, а потім пари кластерів поступово об'єднуються, досягаючи певного рівня взаємодії [1].

Крім того, для аналізу медичних даних також використовують спектральну кластеризацію, яка базується на спектральній теорії графів. Цей метод перетворює дані у графову структуру, де вузли відображають об'єкти, а ребра вказують на ступінь взаємодії між ними. Після цього застосовуються методи аналізу власних значень для визначення структури кластерів.

Для проведення експериментального дослідження зазначених алгоритмів кластеризації було використано вибірку даних, що містить показники, пов'язані з виникненням посттравматичних стресових розладів у бійців АТО. У цій вибірці представлені бінарні, дискретні, категоріальні та рангові дані. Загальна кількість об'єктів у вибірці становить 83, а кількість ознак – 199.

Після отримання даних вони піддавалися процедурі компресії, за допомогою методу головних компонент (РСА), для зменшення обсягу даних перед застосуванням алгоритмів кластеризації. Даний метод дав можливість зменшити обсяг даних, втративши при цьому найменшу кількість інформації. Дані, які були попередньо оброблені та підготовлені, були введені у відповідний алгоритм кластеризації для подальшого аналізу та групування відповідно до їхніх схожих характеристик чи властивостей.

Отримані результати кластеризації оцінювались візуально за допомогою графіків, де можна побачити кількість кластерів та якість групування [2].

В результаті аналізу вибірки даних за допомогою агломеративного методу ієрархічної кластеризації та побудови дендрограми було виявлено наявність викидів у початкових даних. У контексті медичних даних виявлення викидів може мати значимі наслідки, оскільки вони можуть вказувати на несподівані аномалії або відхилення в здоров'ї пацієнтів або в результатах досліджень. Це може включати виявлення непередбачуваних патологій, реакцій на терапію чи інших незвичайних станів.

Інтерпретація викидів може потребувати подальшого клінічного аналізу для розкриття їхньої природи, що може включати проведення додаткових медичних тестів, перегляд медичних історій або консультації з медичними фахівцями.

Після аналізу медичної вибірки даних за допомогою спектральної кластеризації було виявлено трійку чітко виокремлених кластерів, що може свідчити про існування трьох різних підгруп або фенотипів бійців АТО з варіаціями в характеристиках стресових розладів. Це вказує на необхідність розробки індивідуалізованих підходів до діагностики та лікування для кожної з цих підгруп. Аналіз кластерів може допомогти виявити фактори, що мають найбільший вплив на ризик розвитку посттравматичних стресових розладів у бійців АТО, що вказує на потребу подальшого дослідження потенційних факторів ризику або прогностичних маркерів. Також можна припустити, що розподіл на різні групи може вказувати на необхідність розробки різних стратегій лікування, психологічної підтримки та терапевтичних втручань для кожної з цих підгруп.

За результатами експериментального дослідження алгоритмів кластеризації в аналізі медичних даних можна дійти висновку, що виявлення структурних груп або підгруп в цих даних може мати значимі наслідки для клінічної практики та наукових досліджень. Результати цього дослідження підтверджують, що застосування цих алгоритмів дозволяє ефективно ідентифікувати ризикові фактори, прогностичні маркери та розробляти індивідуалізовані стратегії діагностики та лікування різних медичних станів.

Список використаних джерел:

1. Tymoshyk, M. V., Strel'nykov, V. I. Applying user data mining using clustering algorithms. // *Connectivity*. 2020. Vol. 143, №1. P. 20–23.
2. Мірошніченко Н.С., Перова І.Г, Черненко І.О. Огляд існуючих підходів до діагностування посттравматичних стресових розладів / Стан, досягнення та перспективи інформаційних систем і технологій. – 2021. – С. 229.