

СУЧАСНІ МОВНІ МОДЕЛІ ДЛЯ АНАЛІЗУ ТЕКСТУ

Дебре В.С.

Науковий керівник – д.т.н., проф. Гороховатський В.О.

Харківський національний університет радіоелектроніки, каф. ІНФ,

м. Харків, Україна

тел.: (057) 702-14-19, e-mail: viktor.debre@nure.ua

The research materials concern the formation and use of a vector representation for textual data in modern language models. Therefore, the use of such an approach when solving the problems of obtaining answers to questions about professional activity.

Велика мовна модель – це алгоритм нейронних мереж, який застосовує методи глибокого навчання та використовує обширні набори даних для осмислення, узагальнення, генерування та передбачення нового змісту. Властивості цих моделей виражаються у вивченні статистичних зв'язків у текстових документах під час інтенсивного обчислювального процесу, який відбувається під самоконтролем. Вони представляють собою штучні нейронні мережі, при цьому найбільші та потужніші з них базуються на трансформаторній архітектурі [1–3].

Семантичний плагін – це набір функцій, які можуть бути доступні для мовних моделей та сервісів. У фреймворку Semantic Kernel цей підхід реалізовано за допомогою виклику планувальників напряму з плагінів на основі аналізу текстових інструкцій під час виконання застосунків. Робота цього рушію продемонстрована на рис. 1. На вхід Semantic Kernel отримує текстове запитання, після чого проходить через етапи послідовного виклику компонентів мовної моделі, визначення наміру виконання, отримання необхідної інформації з бази знань, та формування висновків на основі всієї інформації. На виході формується текстова відповідь на запитання чи запит та виконується набір інструкцій користувача.

При використанні звичайного підходу передається історія розмови та контекст, який стосується прикладного запитання чи завдання. Але такий підхід дуже обмежений в кількості можливих символів. При використанні підходу з ланцюгом послідовного виклику запитів до системи на базі окремих моделей для планера та зв'язування тексту – все зводиться до векторного представлення та формування відповіді.

За допомогою цього можна передавати значно більшу кількість інформації для обробки запиту користувача. Крім того, оскільки використовується модель для перетворення текстової інформації в вигляд числових векторів, це дозволяє використовувати великі об'єми тексту, які можуть зберігатись в векторних базах даних. Близькі принципи застосовують у системах аналізу візуальних образів [4, 5].

При умові поєднання методів зв'язування тексту з векторними базами знань та підходів рушію Semantic Kernel можна досягти достатній рівень швидкодії та розуміння контексту тексту для виконання практичних завдань.



Рисунок 1 – Схема роботи Semantic Kernel

Векторна база даних – це структура, яка зберігає дані у вигляді векторів великої розмірності, які є математичними представленнями ознак або атрибутів. Кожен вектор має певну кількість вимірів, яка може коливатися від десятків до тисяч залежно від складності та деталізації даних. В даній реалізації вектори формуються на основі контексту та намірів тексту, за допомогою розділення різних характеристик по різних вимірах для векторів.

Для реалізації пошуку у векторних базах даних доцільно застосовувати методи HNSW (Navigable Small World Graphs) для графу пошуку по ній, або IVFFlat (Inverted File with Flat Compression) у разі необхідності забезпечення більшої швидкодії при завантаженні знань в базу та зменшенні точності.

Список використаних джерел:

1. Gorokhovatskyi, V., Vlasenko, N. (2021). Редукція опису зображення у складі множини дескрипторів на основі метричного критерію інформативності. *Advanced Information Systems*, 5(4), pp. 10–16.
2. Gorokhovatskyi V., Tvoroshenko I. (2023) Identification of visual objects by the search request. *Int. scientific symp. «Intelligent Solutions-S». Computational intelligence. Decision making theory: proceedings of the international symposium, September 28, 2023, Kyiv-Uzhorod, Ukraine*, 25-27.
3. Gorokhovatsky V.A. Efficient Estimation of Visual Object Relevance during Recognition through their Vector Descriptions. *Telecommunications and Radio Engineering*. – 2016, Vol. 75, No 14. – P. 1271–1283.
4. Daradkeh Y.I., Gorokhovatskyi V., Tvoroshenko I., Gadetska S., and Al-Dhaifallah M. (2023) Statistical data analysis models for determining the relevance of structural image descriptions, *IEEE Access*, 11, 126938–126949.
5. Gorokhovatskyi V., Tvoroshenko I., Yakovleva O. (2024) Transforming image descriptions as a set of descriptors to construct classification features, *Indonesian Journal of Electrical Engineering and Computer Science*, 33 (1), 113–125.