

РЕКОМЕНДАЦІЙНА СИСТЕМА НА ОСНОВІ МОДЕЛІ TWO TOWER

Вінницький національний технічний університет;

Анотація

У роботі розглянуто гібридну рекомендаційну систему на основі моделі Two Tower. Виділено основні принципи роботи даної моделі та її переваги порівняно з традиційними підходами. Проаналізовано можливості покращення функціональних можливостей рекомендаційних систем шляхом застосування сучасних нейронних мереж. Показано, що модель Two Tower дозволяє враховувати інтереси користувачів та надавати їм релевантні рекомендації в режимі реального часу.

Ключові слова: рекомендаційна система, гібридні системи, модель Two Tower, нейронні мережі.

Abstract

The paper considers a hybrid recommender system based on the Two Tower model. The main working principles of this model and its advantages compared to traditional approaches are highlighted. The possibilities of improving the functionality of recommender systems by using modern neural networks are analyzed. It is shown that the Two Tower model allows taking into account the interests of users and providing them with relevant recommendations in real time.

Keywords: recommendation system, hybrid systems, Two Tower model, neural networks.

Вступ

Сучасні рекомендаційні системи відіграють ключову роль у різноманітних галузях, допомагаючи користувачам знаходити релевантний контент, продукти чи послуги серед величезної кількості варіантів. Традиційні методи, такі як колаборативна фільтрація та контент-базована фільтрація, мають певні обмеження, пов'язані з проблемою холодного старту, нездатністю враховувати неявні ознаки та складні залежності в даних. Тому виникає потреба у розробці більш досконалих підходів із застосуванням сучасних технологій.

Метою роботи є покращення та розширення функціональних можливостей рекомендаційної системи.

Результати дослідження

Система рекомендацій може бути персоналізованою або неперсоналізованою. Неперсоналізована система може бути простішою, але персоналізована система працює краще, оскільки вона задовольняє потреби кожного окремого користувача

Сучасні системи рекомендацій в більшості є персоналізованими і їх можна умовно поділити на три категорії:

1. Колаборативна фільтрація (Collaborative Filtering) - рекомендує елементи на основі показників подібності між користувачами або елементами. Основне припущення алгоритму полягає в тому, що користувачі зі схожими інтересами мають спільні переваги. Існує два основних типи колаборативної фільтрації[1]:

- На основі користувачів: тут ми шукаємо користувачів, які оцінили різні елементи однаково, а потім знаходимо оцінку відсутнього елемента за допомогою цих користувачів.
- На основі предметів: тут ми досліджуємо зв'язок між парою предметів (користувач, який купив А, також купив Б). Ми знаходимо відсутню оцінку за допомогою оцінок, наданих іншим користувачам.

2. Контент-базована фільтрація (Content-based Filtering): Працює на схожості між продуктами. Спочатку ми повинні створити вектор, що представляє всі характеристики продукту. Потім ми обчислюємо схожість між цими векторами за допомогою таких методів, як[2]:

- Евклідова відстань
- Манхеттенська відстань

- Відстань Жаккара
- Косинусна відстань (або косинусна схожість)

3. Гібридні системи

В більшості своїй новітні гібридні системи є кращими бо вирішують низку проблем, які наявні в колаборативних або контент-базованих системах, такі як ефект “холодного старту” і тд. Все це відбувається завдяки використанню нейронних мереж.

Існує кілька архітектур рекомендацій на основі нейронних мереж. Однак, якщо контекстна інформація про наших користувачів доступна лише під час запиту. Це означає, що ми повинні обчислювати рекомендації в реальному часі без великих затримок. Отже, нам потрібна дуже ефективна система.

На щастя, є одна архітектура, яка відповідає всім цим вимогам – Two Tower.

Ідея архітектури Two Tower полягає в тому, що є дві окремі нейронні мережі, які називаються вежами. Можемо інтерпретувати ці дві вежі як окремі моделі, одна з яких представляє користувачів, відома як вежа запитів, а інша - предмети, відома як вежа кандидатів. Під час навчання кожна вежа вчиться перетворювати довільний набір вхідних ознак у вектори[3]. Архітектура моделі зображена на рис. 1.

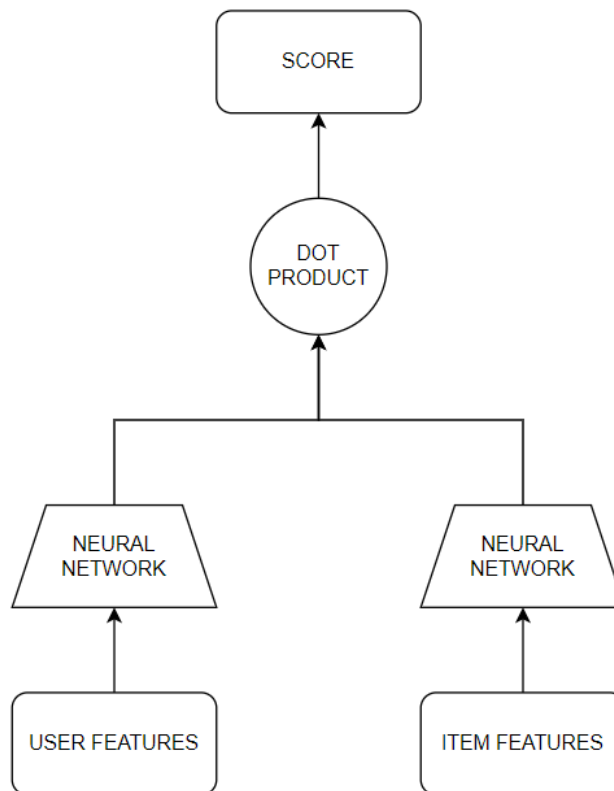


Рис. 1. Архітектура моделі Two Tower

Як ви можете бачити на рис. 1, інформація про користувача та предмети взаємодіє виключно на фінальній стадії роботи системи. Ця характеристика дозволяє ефективно навчати модель і робить Two Tower мережі ідеальним кандидатом для виведення в реальному часі.

Під час виведення рекомендацій ми можемо використовувати той факт, що вбудовані вектори предметів (embeddings) є фіксованими. Після того, як модель була натренована, вбудовані вектори для всіх об'єктів в системі не змінюються. Їх обчислює "вежа кандидатів" на етапі тренування, і ці векторні подання предметів фіксуються. Завдяки цій властивості, під час виведення рекомендацій для користувача, ми можемо заздалегідь обчислити всі вбудовані вектори предметів відразу після тренування і просто завантажити їх у пам'ять, що означає, що нам не потрібно запускати нейронну мережу для елементів під час виведення рекомендацій.

Вбудовані вектори користувача, однак, залежать від контексту. Тому нам доводиться обчислювати їх на ходу. Однак ми повинні робити це лише для одного користувача за раз, що знову ж таки швидко

обчислюється. Ми можемо досягти більшої ефективності якщо вирішимо обчислювати їх з певною регулярністю, однак це означає по жертвування врахуванням історії та контексту в реальному часі.

Нарешті, ми обчислюємо точковий добуток між завантаженими вбудовуваннями елементів і обчисленими вбудовуваннями користувачів і впорядковуємо результати в порядку спадання на основі оцінок. Знову ж таки, дуже ефективна і швидка операція. Для отримання кінцевого результату можливі інші математичні операції окрім точкового добутку або навіть і прості нейронні мережі.

Висновки

У роботі були розглянуті існуючі три підходи по створенню рекомендаційних систем і в результаті проведеного аналізу вибрані гібридні системи, крім того були розглянуті архітектури та моделі рекомендацій на основі нейронних мереж з яких вибрана модель Two Tower.

Визначено, що використання Two Tower моделі для надання рекомендацій є доцільним та ефективним підходом, що розширяє можливості надання рекомендацій за рахунок нейронних мереж та дозволяє враховувати інтреси користувачів та надавати їм рекомендації в реальному часі. Таким чином досягається мета дослідження.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Collaborative Filtering in Machine Learning [Електронний ресурс]. – Режим доступу: <https://www.geeksforgeeks.org/collaborative-filtering-ml/>
2. Building a Content-Based Recommendation System [Електронний ресурс]. – Режим доступу: <https://www.analyticsvidhya.com/blog/2022/08/building-a-content-based-recommendation-system/>.
3. Video Recommendations at Joyn: Two Tower or Not to Tower, That Was Never a Question [Електронний ресурс]. – Режим доступу: <https://medium.com/tech-p7s1/video-recommendations-at-joyn-two-tower-or-not-to-tower-that-was-never-a-question-6c6f182ade7c>

Колісник Сергій Миколайович — студент групи ЗКН-20б, кафедра комп'ютерних наук, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м.Вінниця, e-mail: siroga.03.03@gmail.com

Сілагін Олексій Віталійович—канд. техн. наук, доцент кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця, e-mail: avsilagin@vntu.edu.ua

Kolisnyk Sergiy Mykolayovych - Department of Computer Science, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: siroga.03.03@gmail.com

Silagin Oleksii Vitaliiovich - Cand. Sc. (Eng), Associate Professor of the Department of Computer Science, Vinnytsia National Technical University, Vinnytsia, e-mail: avsilagin@vntu.edu.ua