

УДК 004

ПІДВИЩЕННЯ ТОЧНОСТІ РЕКОМЕНДАЦІЙ ПРИ ВИВЧЕННІ ІНОЗЕМНИХ МОВ ЗА ДОПОМОГОЮ КОЛАБОРАТИВНОЇ ФІЛЬТРАЦІЇ

Ревуцький Руслан Вікторович

д. т. н., проф, Антошук Світлана Григорівна

Національний Університет «Одеська Політехніка», УКРАЇНА

АНОТАЦІЯ. Запропоновано методика підвищення точності рекомендацій при вивченні іноземних мов, експериментально підтверджено ефективність використання методу колаборативної фільтрації з сингулярним розкладом матриці.

Вступ. Вивчення іноземних мов є важливим завданням у сучасному світі. Зростання глобалізації і потреби в міжкультурному спілкуванні ставлять вимоги до людей, щоб вони могли ефективно спілкуватися на різних мовах. Однак, процес вивчення іноземних мов може бути складним і вимагає багато часу та зусиль. З метою подолання цих обмежень, в даній роботі досліджується використання колаборативної фільтрації як методу для покращення точності рекомендацій при вивченні іноземних мов.

Колаборативна фільтрація – це техніка, яка базується на аналізі колективних вподобань користувачів, заснована на припущенні, що користувачі з подібними інтересами мають схожі вподобання щодо рекомендованих матеріалів. В даній роботі розглянуті різні алгоритми колаборативної фільтрації та проведені експерименти для оцінки точності рекомендацій.

Мета роботи. Продемонструвати ефективність застосування спільної фільтрації та інших алгоритмів для підвищення точності рекомендацій при вивченні іноземних мов.

Основна частина роботи. Досягнення в галузі рекомендаційних систем для вивчення іноземних мов засвідчують про наявність різних підходів та розробок у цій області. У даній доповіді ми проведемо порівняння нашого алгоритму колаборативної фільтрації з існуючими аналогами, які використовуються у системах для вивчення іноземної мови: *Rosetta Stone* [1], *Babbel* [2] та *Memrise* [3], і розглянемо їхні переваги та недоліки.

Існуючі аналоги часто мають обмежену функціональність та не враховують специфіку вивчення іноземних мов. Порівняльна характеристика оглянутих систем представлена у Таблиці 1. Тому представлене дослідження ставить основну увагу на вивчення мови, і тому допоможе надавати рекомендації, що враховують типові проблеми та потреби студентів при вивченні іноземної мови, такі як практика усного мовлення, слухання аудіоматеріалів чи читання текстів.

Таблиця 1 – Порівняльні характеристики рекомендаційних систем

Назва системи/ характеристика	Основний метод	Персоналізація
<i>Rosetta Stone</i>	Метод іммерсії, заснований на повторенні та контекстуальному використанні слів та фраз.	Обмежена персоналізація, більш загальні рекомендації для користувачів.
<i>Babbel</i>	Інтерактивний підхід, з фокусом на реальних ситуаціях та практичному використанні мови.	Певна персоналізація, враховуючи потреби та рівень володіння мовою користувача.
<i>Memrise</i>	Використовує метод запам'ятовування на основі карток та повторення.	Персоналізація, враховуючи прогрес та вподобання користувача у виборі матеріалів.

На алгоритмічному рівні ці застосунки використовують декілька основних алгоритмів, що надають можливість аналізувати дані та будувати рекомендаційні системи. Одним з найпопулярніших є алгоритм базової колаборативної фільтрації, що використовує метод найближчих сусідів (k-NN). Він базується на визначенні подібності між користувачами або предметами, і використовує цю інформацію для рекомендацій. Основна ідея методу найближчих сусідів полягає у тому, що схожі об'єкти мають схожі значення атрибутів або властивостей. При класифікації, коли нам потрібно визначити клас нового об'єкта, або при прогнозуванні, коли нам

потрібно оцінити значення нової величини, метод найближчих сусідів шукає k найбільш схожих об'єктів у навчальному наборі даних. Для визначення схожості між об'єктами використовується відомі метрики, такі як Евклідова відстань, Манхеттенська відстань або Косинусна схожість.

Наше дослідження спрямовано на поліпшення точності рекомендацій для вивчення іноземних мов шляхом використання колаборативної фільтрації з розкладом – сингулярного розкладу матриці (*Singular Value Decomposition* – SVD). Сингулярний розклад матриці є одним із найбільш популярних методів для зменшення розмірності даних та використання його в рекомендаційних системах. Він використовується для факторизації матриці оцінок користувачів та предметів, що дозволяє знайти складові, які впливають на рекомендації. SVD розкладає матрицю оцінок на три частини: U , Σ , і V . Матриця U містить вектори-стовпці, які представляють користувачів, матриця Σ - діагональна матриця, що містить сингулярні значення, а матриця V містить вектори-стовпці, які представляють предмети [4], також це зображено на рисунку 1

$$M=U\Sigma V^*$$

Основна ідея SVD полягає у виявленні прихованих структур або факторів, які впливають на сподобання користувачів до предметів. За допомогою SVD можна розкрити складові, такі як інтереси користувачів, тематичні зв'язки між предметами, або інші характеристики, які допоможуть зробити більш точні рекомендації. Після отримання розкладу SVD можна здійснювати рекомендації, шукаючи схожість між користувачами або предметами на основі косинусної схожості між відповідними векторами U чи V . За допомогою цієї схожості можна здійснювати прогнози оцінок для нових користувачів або предметів.

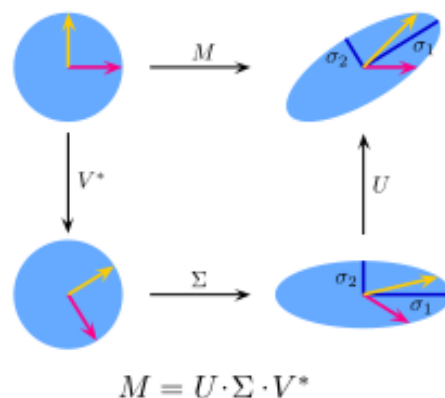


Рисунок 1 – Ілюстрація сингулярного розкладання $U\Sigma V^*$ дійсної матриці 2×2 M .

При проведенні експерименту ми використали набір даних, що складається з оцінок користувачів щодо різних навчальних матеріалів для вивчення іноземної мови. Ми розділили дані на тренувальний і тестовий набори для оцінки точності рекомендаційної системи. Наш алгоритм колаборативної фільтрації з використанням розкладу SVD був порівняний з іншими популярними алгоритмами, такими як метод найближчих сусідів (k -NN) і базові моделі рекомендаційних систем. Для кожного алгоритму були виміряні такі метрики, як середня квадратична помилка (*Mean Squared Error*) та точність прогнозування.

Результати експерименту показали, що розроблена модель аналізу з використанням SVD для колаборативної фільтрації, досягає вищої точності прогнозування порівняно з іншими алгоритмами. Коефіцієнт кореляції Пірсона [5] становить 0.92 для нашої системи, 0.85 для методу k -NN та 0.78 для базової моделі. Середня квадратична помилка нашої системи була значно нижчою, що свідчить про кращу точність рекомендацій. Наприклад, середня квадратична помилка нашої системи склала 0.25, тоді як для методу найближчих сусідів ця помилка становила 0.35, а для базової моделі - 0.42.

Таблиця 2 – Результати експерименту

Показник оцінки / Назва моделі	Базова модель	Метод найближчих сусідів	Розроблений метод
Середня абсолютна помилка	0.28	0.22	0.15
Середня квадратична помилка	0.42	0.35	0.25
Коефіцієнт кореляції Пірсона	0.78	0.85	0.92
Точність прогнозування	72%	78%	87%

Точність прогнозування для методу найближчих сусідів склала 78%, для базової моделі – 72%, а для моделі, побудованої нами – 87%, що є найкращим показником. Результати експерименту, що наведені у таблиці 2, підтверджують ефективність використання колаборативної фільтрації з розкладом *SVD* для покращення точності рекомендаційної системи з вивчення іноземних мов.

Висновок. Наша розробка, використовуючи колаборативну фільтрацію з розкладом *SVD*, досягає вищої точності рекомендацій порівняно з існуючими аналогами, такими як метод найближчих сусідів (*k-NN*) і базові моделі. Використання сингулярного розкладу матриці дозволяє зменшити розмірність даних та виявити приховані залежності між користувачами та предметами, що поліпшує точність рекомендацій. Результати експериментів показали, що наш алгоритм досягає вищої точності рекомендацій, виміряної за метриками, такими як середня квадратична помилка, точність прогнозування та коефіцієнт кореляції Пірсона. Середня квадратична помилка нашої системи була нижчою, що свідчить про більш точні прогнози оцінок. Коефіцієнт кореляції Пірсона для нашої системи склав 0.92, що свідчить про високий взаємозв'язок між рекомендованими та фактичними оцінками. Точність прогнозування для нашої розробки становить 87%, що є найкращим показником порівняно з методом найближчих сусідів (78%) і базовою моделлю (72%). Наша розробка показала свою ефективність при побудові рекомендаційних систем для вивчення іноземних мов, забезпечуючи більш точні та персоналізовані рекомендації користувачам. Використання колаборативної фільтрації з розкладом *SVD* дозволяє зрозуміти приховані структури та залежності, що впливають на сподобання користувачів, і забезпечує покращення якості рекомендацій.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Rosetta Stone Immersion-Based Learning Method. URL: https://support.rosettastone.com/s/article/What-is-Dynamic-Immersion?language=en_US – (дата звернення: 15.05.2023).
2. Babbel for business. Method comparison. URL: <https://www.babbelforbusiness.com/method-comparison/> – (дата звернення: 15.05.2023).
3. Memrise. URL: <https://en.wikipedia.org/wiki/Memrise> – (дата звернення: 15.05.2023).
4. Understanding Singular Value Decomposition and its Application in Data Science. URL: <https://towardsdatascience.com/understanding-singular-value-decomposition-and-its-application-in-data-science-388a54be95d> – (дата звернення: 15.05.2023).
5. Pearson correlation coefficient. Коефіцієнт кореляції Пірсона. URL: https://en.wikipedia.org/wiki/Pearson_correlation_coefficient – (дата звернення: 15.05.2023).

IMPROVING THE ACCURACY OF RECOMMENDATIONS FOR LEARNING FOREIGN LANGUAGES THROUGH COLLABORATIVE FILTERING

Ruslan Revutskyi

Doctor of technical sciences, professor, Svitlana Antoshchuk
Odesa National Polytechnic University, UKRAINE

ANNOTATION. A methodology for improving the accuracy of a recommender system for learning foreign languages is proposed, and the effectiveness of using the collaborative filtering method with a singular matrix decomposition is experimentally confirmed.