

# ДОСЛІДЖЕННЯ МЕТОДІВ СТВОРЕННЯ РЕКОМЕНДАЦІЙНИХ СИСТЕМ ДЛЯ ПЛАТФОРМИ ОГЛЯДУ КНИГ

Вінницький національний технічний університет

## *Анотація*

*Доведено актуальність задачі створення рекомендаційних систем для платформи огляду книг. Здійснено огляд основних методів створення рекомендацій. Проаналізовано їх основні переваги та недоліки. Виконано порівняння та вибір підходу для розв'язання даної задачі.*

**Ключові слова:** колаборативне фільтрування, фільтрування на основі вмісту, рекомендаційна система

## *Abstract*

*The relevancy of the creating recommendation systems task for the book review platform is proved. An overview of the main methods of creating recommendations is provided. Their main advantages and disadvantages are analyzed. The comparison and selection of the approach for solving this problem is performed.*

**Keywords:** collaborative filtering, content-based filtering, recommendation system

## **Вступ**

В останні роки були розроблені різні типи рекомендаційних систем на основі текстового огляду, порівняльної думки, рейтингів користувачів, моделей покупок, профілів користувачів і т. д. Ці системи змінили спосіб ведення електронної онлайн-комерції та функціонування соціальних медіа – від рекомендацій друзів у Facebook до придбання продуктів на Amazon та вибору фільму та музики на Netflix.

Система рекомендацій виступає як сімейство систем фільтрації інформації, що надають рекомендації користувачам на основі їх симпатій та антипатій. Актуальність рекомендацій стає ще більш значущою в сучасному світі завдяки великій кількості інформації та можливостей. Оскільки обсяг інформації в наш час стрімко збільшується, це породжує дилему у користувачів під час вибору товарів, які вони насправді хочуть придбати, або послуги, на які вони насправді хочуть підписатися. У такому випадку система рекомендацій набуває величезного значення.

До таких систем можна віднести і платформи для огляду книг. Наразі вже існують подібні платформи для огляду, але вони вбудовані в системи електронної комерції (Amazon, eBay). Суттєвим недоліком таких платформ є те, що вони в першу чергу спрямовані на підтримку бізнесу, а не на інтереси користувачів. Крім того, має місце обмеженість інструментів оцінювання книг.

Система рекомендацій, що пропонується у даній роботі має вдосконалити існуючі системи рекомендацій, надавши користувачам ширший функціонал, та додавши характеристики, що можуть впливати на оцінку книги. Крім того, пропонується відв'язати таку платформу від чинників, які вносять певні бізнес фактори.

Нижче стисло проаналізовано методи впровадження системи рекомендацій, а також техніки, що використовуються цими методами.

## **Основна частина**

Існують різні методи для розробки ефективної системи рекомендацій, два з яких складають основу для розробки інших підходів. Це методи фільтрування вмісту та спільного фільтрування [1].

Методи, що базуються на вмісті (content-based), використовують перелік функцій елемента та порівнюють його з елементами, яким раніше віддавали перевагу певні користувачі. Елементи, що відповідають схожості, рекомендуються користувачеві. Основна функція фільтрування на базі вмісту працює у два етапи. Спочатку відбувається збереження профілю користувача на основі функцій елементів, яким користувач найчастіше надає перевагу. Ці ознаки використовуються для відображення подібності одного предмета з іншим за допомогою рівняння подібності. Після цього відбувається порівняння характеристики кожного елемента з профілем користувача та рекомендуються ті, що мають

високий ступінь подібності [2]. Для системи, що базується на вмісті, потрібно побудувати профіль елемента, який є записом основних характеристик цього елемента. Такі характеристики легко виявити. Наприклад, у книзі, запис може містити список жанрів, авторів, рік та країну випуску тощо.

Content-based фільтрація є найбільш простим і природним методом, що можна застосувати як рекомендацію, адже він не вимагає від користувача зворотного зв'язку. Іноді одного вибору достатньо, щоби рекомендувати користувачеві багато елементів системи. Цей підхід також розширюється природним шляхом для випадків, коли інформація про товари добре організована та доступна, наприклад, книги, фільми, пісні, продукти і т. д. Але у той же час це також є обмеженням фільтрування на основі вмісту, оскільки опис товару не завжди присутній і це створює певні труднощі під час вимірювання подібності між предметами. Такі рекомендаційні системи мають обмеження для отримання подібних результатів і є статичними протягом часу.

Колаборативне фільтрування є найпопулярнішим і найпоширенішим методом рекомендацій. Його основою є те, що користувачі, що мають однакові інтереси, схильні надавати однакові переваги новим та майбутнім елементам. Цей прийом працює за двома принципами. По-перше, він служить критерієм вибору групи подібних людей, думки яких будуть накопичені як основа для рекомендації (найближчих сусідів). По-друге, він також використовує ці думки, щоб сформувати більшу групу та мати більший вплив на рекомендацію. Прийоми спільного фільтрування передбачали дуже великі масиви даних та непрямі сфери застосування, такі як фінанси, прогнозування погоди, зондування довілля, електронна комерція тощо [3].

Методи спільного фільтрування використовують набір даних переваг/оцінок, наданих користувачами для елементів, щоби передбачити нові елементи, які можуть сподобатися активному користувачеві. Модель може бути виражена у вигляді преференцій / рейтингової матриці порядку  $m \times n$ , де  $m$  – кількість користувачів ( $U_1, U_2, U_3, \dots, U_m$ ), а  $n$  – кількість елементів ( $I_1, I_2, I_3, \dots, I_n$ ), оцінених користувачами. Значення комірки  $r_{ij}$  матриці – це рейтинг елемента  $j$ , який було надано користувачем  $i$ . Ці оцінки можуть бути неявними (наприклад, придбання товару) або явними (відгуки користувача за шкалою  $k$ ). Результат спільних прийомів може бути двох типів: перший – прогнозування  $r_{ij}$ , де числове значення показує, що користувач надає перевагу елементу  $j$ , а друге – список рекомендацій з  $N$  найпопулярніших елементів, які користувачеві можуть сподобатися найбільше [4].

Основною функцією рекомендації є передбачення корисності елемента для користувача. Система рекомендацій характеризує, як користувач  $U$  зацікавлений у пункті з певним ступенем уподобання або рейтингу  $r(u, i)$ . Кожен користувач має свій профіль користувача, що описує його смак, симпатії, антипатії або оцінку чи відгук по певному предмету. Кожен елемент характеризується набором функцій, наприклад, для книги набір функцій може містити ідентифікатор книги, жанр, авторів, дату та країну виходу.

Фільтрування, засноване на знаннях, використовує знання або інформацію користувачів, предметів та їх взаємозв'язки. Системи, що використовують даний метод описують, як конкретний предмет відповідає вимогам користувача. Це вимагає окремих знань про користувачів та елементи, специфічних для предметної області (домену) [5].

Гібридні методи фільтрування поєднують переваги двох або більшої кількості методів фільтрування та долає їх обмеження. Такі методи забезпечують доволі високу ефективність і покращують результати рекомендацій [6]. Гібридні методи можуть використати одну з нижченаведених стратегій для розроблення методу гібридного фільтрування:

- 1) використання колаборативних та content-based фільтрів для вироблення окремих рекомендацій, а потім створення лінійної комбінації цих двох рекомендацій, для надання єдиної рекомендації;

- 2) колаборативне фільтрування може бути використано разом з характеристиками фільтрування на основі вмісту для обчислення подібності між користувачами та знайденими сусідами з метою прогнозування рекомендацій;

- 3) методи, засновані на вмісті, можуть бути додані до спільних характеристик фільтрування, таких як модель прихованих факторів із підходом, заснованим на вмісті;

- 4) звичайний імовірнісний метод поєднання спільної та колаборативної техніки для прогнозування рекомендацій.

Проаналізувавши усі вищенаведені методи фільтрування даних для системи рекомендацій, було вирішено зупинитися на гібридному методі, скомбінувавши метод колаборативного фільтрування та

метод фільтрування, що заснований на вмісті. Це дозволить уникнути недоліків кожного методу та можливість отримати рекомендації, що якнайбільше збігатимуться з інтересами користувачів.

Також з метою прогнозування множини книг, що можуть бути цікавими користувачам, можна врахувати підходи, запропоновані у роботах [7, 8].

## Висновки

Обґрунтовано актуальність розробки рекомендаційних систем для платформи огляду книг. Здійснено аналіз основних методів створення рекомендаційних систем, зокрема основні переваги та недоліки кожного з них. Запропоновано, з метою підвищення якості рекомендації книг здійснити комбінування методу колаборативного фільтрування та методу фільтрування, що заснований на вмісті.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Goldberg D. Using collaborative filtering to weave an information tapestry / D. Goldberg, D. Nichols, B. M. Oki., 1992. – 72 с.
2. Chen L. A user-centric evaluation framework for recommender systems. In: Proceedings of the fifth ACM conference on Recommender Systems / L. Chen, P. Pu, R. Hu. – Нью Йорк: ACM, 2011.
3. Breese J. S. Емпіричний аналіз прогнозних алгоритмів робочий фільтр. В: Матеріали чотирнадцятої щорічної конференції з питань невизначеності в ШІ [Електронний ресурс] / J. S. Breese, D. Heckerman, C. Kadie. – 1998.
4. Чебанюк О. В. Методика розпізнавання рукописного тексту на основі аналізу векторів руху за допомогою сенсорних пристроїв / О. В. Чебанюк, Д. А. Долотов – Вісник східноукраїнського національного університету імені Володимира Даля 2015, С. 230 – 236.
5. Claypool O. Combining content-based and collaborative filters in an online newspaper./ O. Claypool, / ACM SIGIR Workshop on Recommender Systems: algorithms and evaluation. – 2011. – 6 с.
6. Гудфеллоу Я. Методы фильтрации / Я. Гудфеллоу, И. Бенджио, А. Курвилль пер. с англ. А. А. Слинкина – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с.
7. Солоний М. А. Порівняльний аналіз підходів розробки модулю прогнозування для мобільного додатку інтернет-магазину / М. А. Солоний, І. Р. Арсенюк // Матеріали XLIX науково-технічної конференції підрозділів ВНТУ. Вінниця, 2020. URL: <http://ir.lib.vntu.edu.ua/bitstream/handle/123456789/29506/8989.pdf?sequence=3>
8. Арсенюк І. Р. Інтелектуальний модуль інтернет-парсингу контенту веб-ресурсів / Арсенюк І. Р., Кратасюк В. // Матеріали ІХ Міжнародної науково-практичної конференції “Інтернет-Освіта-Наука” (ІОН-2014). – Вінниця: УНІВЕРСУМ-Вінниця, 2014. – С. 65 – 66.

**Зайчик Владислав Олександрович** – студент групи 2КН-17Б, факультет інформаційних технологій та комп’ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: [vladyslav.zaichyk@gmail.com](mailto:vladyslav.zaichyk@gmail.com)

**Арсенюк Ігор Ростиславович** – к. т. н., доцент, доцент кафедри комп’ютерних наук, Вінницький національний технічний університет, м. Вінниця.

**Vladyslav V. Zaichyk** – student of group 2КН-17Б, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: [vladyslav.zaichyk@gmail.com](mailto:vladyslav.zaichyk@gmail.com)

**Igor R. Arsenyuk** – Cand. Sc., Assistant Professor of the Chair of Computer Science, Vinnytsia National Technical University, Vinnytsia.