

## ПРОГНОЗУВАННЯ ПОВЕДІНКИ СПОЖИВАЧІВ З ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ

*Скринник Д. В., магістрант, daniil.skrynnyk.kita@donntu.edu.ua;  
Донецький національний технічний університет, Луцьк, Україна*

У сучасному світі цифрової комерції, де споживачі зіткнулися з небувалою різноманітністю вибору, виникає складність у ефективному навігуванні по великих каталогах товарів. Це ставить перед користувачами задачу знаходити продукти, що відповідають їх індивідуальним потребам та вподобанням. І хоча ця різноманітність надає більше можливостей, вона також створює певні труднощі. Однією з ключових проблем є питання, як допомогти користувачам ефективно орієнтуватися у великих каталогах, щоб знайти те, що їм потрібно. Для вирішення цієї проблеми розглядається застосування системи рекомендацій, що використовує гібридний метод фільтрації.

У рамках проекту здійснюється збір та попередня обробка даних, що включають інформацію про взаємодію користувачів із товарами, а також атрибути самих товарів. Особливу увагу приділено механізму «багаторукого бандита», який дозволяє експериментувати з різними варіантами рекомендацій та вибирати найефективніші, тим самим покращуючи точність рекомендацій та загальну корисність системи для користувачів.

### Структура пропонованої системи розпізнавання

У процесі рекомендації в веб-додатку використовується гібридна система рекомендацій, що зображена на рис. 1.



Рисунок 1. Загальна схема роботи  
гібридної системи з впровадженням  
стратегії багаторукого бандита

Спочатку використовується колаборативна фільтрація для аналізу взаємодій між користувачами і елементами. Паралельно проводиться контент-орієнтована фільтрація, що аналізує властивості елементів. Результати обох підходів інтегруються та комбінуються, після чого система видає кінцеві рекомендації користувачам. Після

видачі рекомендацій, стратегія багаторукого бандита оптимізує процес розподілу рекомендацій.

Для першої частини гібридної системи використовується фільтрація вмісту. Механізм рекомендацій працює шляхом порівняння характеристик елементів із профілем користувача та ранжування елементів на основі їхньої схожості з уподобаннями користувача. Кінцевим результатом є список рекомендованих елементів, які, швидше за все, зацікавлять користувача.

Одним із основних способів досягти цього, особливо з текстовими даними, є косинусна подібність, яку часто отримують із представлень вмісту TF-IDF [1].

Основою для цього підходу є метрика TF-IDF (1) та (2). Вона кількісно визначає важливість слова в конкретному документі порівняно з його поширеністю в колекції документів. Обчислення складається з двох частин:

$$TF = \frac{n_i}{\sum_k n_k}, \quad (1)$$

де  $n_i$  — число входжень слова в документ;

$n_k$  — кількість слів в документ.

$$IDF = \log \frac{|D|}{|(d_i \ni t_i)|}, \quad (2)$$

де  $|D|$  — кількість документів колекції;

$|(d_i \ni t_i)|$  — кількість документів, в яких зустрічається слово  $t_i$ .

Після визначення ваги TF-IDF для всіх термінів у всіх документах кожен документ можна представити як вектор у багатовимірному просторі, де кожен вимір відповідає вазі TF-IDF терміна [1].

Далі, щоб оцінити подібність між будь-якими двома документами, використовується міра косинусної подібності. Ця метрика обчислює косинус кута між двома векторами (3).

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}, \quad (3)$$

де  $A \cdot B$  — скалярний добуток векторів;

$\|A\|$  та  $\|B\|$  — скалярні значення цих векторів [2].

Для реалізації фільтрації вмісту використовується бібліотека аналізу тексту scikit-learn. Вона використовує інструменти TfidfVectorizer для створення матриці TF-IDF на основі текстових описів товарів та linear\_kernel для обчислення косинусної подібності між товарами на основі цієї матриці.

Другою важливою частиною гібридної системи є колаборативна фільтрація яка базується на ідеї, що користувачі, які мають схожі зацікавлення чи вподобання в минулому, ймовірно, матимуть схожі вподобання й у майбутньому. Цей метод працює на основі аналізу взаємодії користувачів з продуктами, для цього зручно використовувати сингулярне розкладання матриці [3].

Припустимо, що є матриця  $R$  (4):

$$R = \begin{bmatrix} r_{11} & r_{11} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix}, \quad (4)$$

Де  $R$  — матриця з розміром  $m \times n$ , де  $m$  це кількість користувачів,  $n$  це кількість ознак товару;

$r_{ij}$  — оцінка, яку користувач  $i$  дав товару  $j$ .

Декомпозицію матриці  $R$  можна представити наступним чином (5):

$$R = U\Sigma V^T, \quad (5)$$

де  $U$  — ортогональна матриця розміром  $m \times m$ , що представляє вектори ознак, які відповідають користувачам у прихованому просторі ознак;

$\Sigma$  — це діагональна матриця розміром  $m \times n$  з невід'ємними дійсними числами, відомими як сингулярні значення;

$V^T$  — ортогональна матриця розміром  $n \times n$ , що представляє вектори ознак, що відповідають елементам у прихованому просторі ознак.

Після декомпозиції, можна виділити рейтинг  $r_{ij}$  (6) користувача  $i$ , товару  $j$ :

$$\hat{r}_{ij} = \sum_{k=1}^K u_{ik} \sigma_k u_{jk}, \quad (6)$$

де  $k$  — кількість сингулярних значень, які необхідно зберегти [3].

Для колаборативної фільтрації використовується бібліотеки `scikit-learn` і `pandas`. Вона використовує метод головних компонент для зменшення розмірності матриці, обчислює схожість користувачів за допомогою `cosine_similarity` та рекомендує книги на основі замовлень схожих користувачів.

Ключовим етапом є об'єднання обох рекомендацій, що дозволяє вирішити обмеження, характерні для окремих типів систем, забезпечуючи більш високу якість та точність рекомендацій. Для гібридної системи було обрано важелеве комбінування результатів. Це означає, що якщо одна модель має більшу вагу, інша модель відповідно матиме меншу вагу.

Комбінований рейтинг  $R$  (7) для колаборативної фільтрації та фільтрація на основі вмісту може бути обчислений за формулою:

$$R = w_1 R_{colab} + w_2 R_{content}, \quad (7)$$

де  $w_1$  — вага колаборативного метода;

$R_{colab}$  — рекомендаційний рейтинг колаборативної фільтрації;

$w_2$  — вага фільтрації на основі вмісту;

$R_{content}$  — рекомендаційний рейтинг фільтрації на основі вмісту;

Для забезпечення коректності, зазвичай ваги нормалізують, тобто:  $w_1 + w_2 = 1$ .

Але в гібридній системі рекомендацій, досить складно обрати правильні ваги, тому використовується ідея «багаторукогого бандита», де кожна «рука» відповідає певній комбінації ваг, що визначає баланс між колаборативними та контентними рекомендаціями. Для оптимального вибору цих ваг використовується стратегія верхньої довірчої межі (8), яка адаптивно регулює ці ваги залежно від накопиченого досвіду.

$$R_{ui}^{(s)} = w_1^{(s)} R_{colab_{ui}} + w_2^{(s)} R_{content_{ui}}, \quad (8)$$

де  $w_1^{(s)}$ ,  $w_2^{(s)}$  представляють коефіцієнти користувачів  $u$  та товару  $i$  для вибраної рук  $s$ .

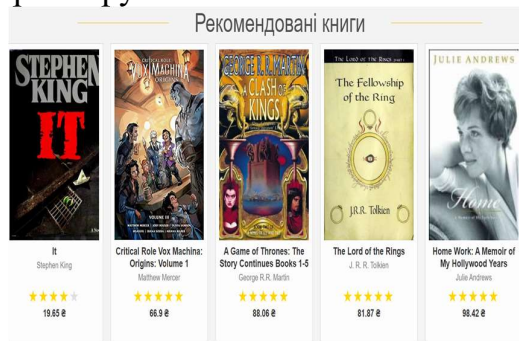


Рисунок 2. Результат роботи рекомендаційної системи

На рис. 2 представлено приклад роботи інтернет-книгарні, макету сайту, розробленого з використанням Flask. На цьому рисунку можна побачити, як застосовується стратегія багаторукого бандита для рекомендації книг. Для перевірки ефективності цієї системи створено тестового користувача, який купує різні

книги. На основі його вибору система пропонує схожі товари, щоб запропонувати користувачеві релевантні рекомендації.

## Література

1. Recommender System Series, Part 1: Introduction & Content-Based Filtering [Електронний ресурс]. — Режим доступу: <https://medium.com/encora-technology-practices/recommender-system-series-part-1-introduction-content-based-filtering-785ade61d2dc> (дата звернення: 28.11.2023). — Назва з екрана.
2. Understanding Cosine Similarity in Recommendation Engines [Електронний ресурс]. — Режим доступу: <https://medium.com/@TechTalkWithAlex/understanding-cosine-similarity-in-recommendation-engines-3d98068f2b73> (дата звернення: 28.11.2023). — Назва з екрана.
3. Introduction to Recommender System [Електронний ресурс]. — Режим доступу: <https://medium.com/hackernoon/introduction-to-recommender-system-part-1-collaborative-filtering-singular-value-decomposition-44c9659c5e75> (дата звернення: 28.11.2023). — Назва з екрана.

## Анотація

Представлено розробку системи рекомендацій для веб-додатку, яка використовує гібридний підхід, комбінуючи фільтрацію на основі вмісту та колаборативну фільтрацію. Процес розробки включає збір та попередню обробку даних про взаємодію користувачів з товарами та атрибутами самих товарів, з особливим акцентом на механізмі «багаторукого бандита» для оптимізації рекомендацій.

Ключові слова: система рекомендацій, гібридна фільтрація, колаборативна фільтрація, фільтрація на основі вмісту, багаторукий бандит, Python, Flask.

## Abstract

Presented is the development of a recommendation system for a web application that employs a hybrid approach, combining content-based and collaborative filtering. The development process includes the collection and preliminary processing of data on user interactions with products and the attributes of the products themselves, with a special focus on the "multi-armed bandit" mechanism for optimizing recommendations.

Keywords: recommendation system, hybrid filtering, collaborative filtering, content-based filtering, multi-armed bandit, Python, Flask.