

## **NLP- ТА LLM-ОРІЄНТОВАНА СИСТЕМА ПЕРСОНАЛІЗОВАНИХ РЕКОМЕНДАЦІЙ ФІЛЬМІВ**

Салтан С.О., Гриньова О.Є.

e-mail: serhii.saltan.cpe@nure.ua, olena.hrynova@nure.ua

Харківський національний університет радіоелектроніки, каф. ШІ  
м. Харків, Україна

This text discusses the development of a recommendation system for movies based on text analysis, aiming to simplify the process of finding suitable films. The system aggregates reviews from reliable sources like Metacritic and Rotten Tomatoes and uses user profile data to suggest movies, especially during the «cold start» phase. It also customizes recommendations based on user preferences and allows for joint recommendations for couples by analyzing both users' tastes.

Користувачі інтернет-ресурсів стикається зі стрімким зростанням обсягу даних, що значно ускладнює пошук необхідної інформації. Надмірність контенту ускладнює орієнтацію користувачів, особливо у специфічних сферах, таких як вибір фільмів. Критичні огляди експертів містять великі обсяги текстових даних, аналіз яких є трудомістким та вимагає значних часових ресурсів. У цьому контексті актуальним є розробка інтелектуальної системи, здатної автоматизувати процес узагальнення оцінок критиків та формувати персоналізовані рекомендації для користувачів, так званої рекомендаційної системи [1]. Така система має забезпечувати ефективну фільтрацію інформації, враховуючи як об'єктивні оцінки експертів, так і суб'єктивні уподобання аудиторії.

Подібні підходи вже активно застосовуються в різноманітних сферах споживання контенту. Наприклад, Blinkist – це сервіс, що спеціалізується на аналізі книг та створенні коротких резюме, відомих як «блінки», які містять основні ідеї творів. Система рекомендацій Blinkist функціонує на основі детального аналізу уподобань користувачів, враховуючи прочитані блінки, їх оцінки та інтереси, з одного боку, та рекомендацій лідерів думок з іншого. Це дозволяє пропонувати нові книги, які можуть бути цікавими для конкретного користувача. Blinkist використовує два основні підходи: контентну фільтрацію, яка визначає рекомендації на основі тематичного та жанрового збігу, а також колаборативну фільтрацію, що ґрунтується на вподобаннях інших користувачів з подібними інтересами та інфлюенсерів. Завдяки цим методам, Blinkist забезпечує персоналізований підхід до вибору контенту, що сприяє більш ефективному задоволенню інформаційних потреб користувачів.

Рекомендаційні системи класифікуються за кількома основними типами – контентно-орієнтовані, колаборативні, гібридні та контекстні [2]. Система, запропонована в роботі, належить до категорії гібридних

рекомендаційних систем, оскільки інтегрує методи контентного аналізу (аналіз характеристик об'єктів, зокрема, жанр, режисер, акторський склад, індивідуальні вподобання користувача) та персоналізованого підходу, заснованого на даних користувача (вік, стать, регіон, тощо). Такий підхід дозволяє забезпечити високу точність рекомендацій завдяки синтезу різних технологій.

До характеристик пропонованої рекомендаційної системи слід віднести наступні – узагальнення, тональність рецензій та рейтинг, вподобання. Узагальнення – система забезпечує користувачам єдину платформу для ознайомлення з результатами опрацювання рецензій професійних критиків із таких ресурсів, як Metacritic та Rotten Tomatoes (можливо налаштувати список джерел, звідки брати критику для аналізу), а також здійснює, без спойлерів, неупереджений аналіз критичних статей, що дозволяє уникнути необхідності переглядати численні сторінки з відгуками до фільму на різних платформах. Користувач отримує узагальнений аналіз рецензій із декількох джерел у зручному форматі, відповідно до свого профілю.

Тональність рецензій та рейтинг. Застосовуються методи обробки природної мови (NLP) для оцінювання тональності рецензій, виділення ключових тем та виявлення прихованих закономірностей у відгуках. Використання NLP для рекомендацій фільмів має кілька суттєвих переваг - дозволяє більш точно визначати загальну оцінку фільму та відповідність уподобанням конкретного користувача. Застосування методів NLP забезпечить можливість автоматизованого аналізу поточних тенденцій, трендів та тематичних напрямів у сфері кінематографу, що сприятиме оптимізації та підвищенню релевантності рекомендаційних систем.

Аналіз рецензій здійснюватиметься з урахуванням індивідуальних потреб користувача за допомогою попередньо навченої моделі BART. Зокрема, мова текстів буде адаптована до вікової категорії користувача - для дітей та підлітків будуть застосовані спрощені тексти, порівняно з дорослими. В рамках системи передбачено окремий параметр налаштування, що дозволяє користувачеві індивідуально визначати стиль викладення узагальненої рецензії, яка є результатом інтеграції множинних експертних оцінок.

Формування інтегрованих рекомендацій здійснюється на основі гібридної моделі, що поєднує контентно-орієнтований підхід з колаборативною фільтрацією. Колаборативна фільтрація передбачає ідентифікацію подібностей між користувачами та формування рекомендацій на основі узгоджених переваг, що забезпечує підвищення точності та персоналізації при задоволенні потреб користувачів [3].

На початковому етапі використання, так званий «холодний старт», коли ще відсутні історичні дані про перегляди та вподобання користувача, система ґрунтується виключно на інформації, наданій у процесі реєстрації.

Зокрема, при першому зверненні до системи алгоритм рекомендацій зважає на анкетні дані користувача та формує список фільмів відповідно до його категорії: жінки (драми, мелодрами, романтична комедія); чоловіки (бойовики, вестерн, воєнний фільм, історичний фільм); діти (фентезі, мультики, казки); сімейні (комедія, фантастика) тощо.

На основі накопичених даних щодо історії переглядів користувача, система здатна адаптувати рекомендації відповідно до його індивідуальних вподобань, використовуючи метод  $k$ -найближчих сусідів ( $k$ -Nearest Neighbors, KNN). Це сприяє підвищенню релевантності відповідей системи за рахунок динамічної адаптації профілю користувача.

З метою надання користувачеві рекомендацій щодо фільмів, що належать до жанрово близьких категорій переглянутого контенту, застосовується алгоритм контентної фільтрації (Content-Based Filtering). Для категоризації фільмів використовується алгоритм фільтрації на основі знань (Knowledge-Based Filtering). Конструювання нейронної мережі, здатної аналізувати контекст та емоційне забарвлення рецензій, включаючи наявність спойлерів, реалізується з використанням трансформеру BERT (Bidirectional Encoder Representations from Transformers), що є дуже ефективними для розуміння контексту та нюансів мови. BERT може використатись для класифікації тексту, вилучення ознак та інших завдань.

Етапи обробки відгуків. Спочатку здійснюється збір відгуків з різних платформ, що дозволяє отримати широкий спектр думок та оцінок щодо фільмів (рецензії з соціальних мереж, кінопорталів та інших джерел). Потім відгуки класифікуються за ключовими характеристиками, такими як жанр, позитивні та негативні оцінки, а також тональність рецензій. Далі, на основі класифікованих даних, генерується підсумок (квінтесенція) відгуків, який враховує зазначені критерії. Це узагальнення є основою для подальшого формування рекомендацій.

Запропонована система є комплексним рішенням, що забезпечує високу точність рекомендацій. Поєднання контентної фільтрації, фільтрації на основі знань та глибокого навчання надає нові можливості для ефективного підбору контенту, адаптованого до індивідуальних потреб користувачів, забезпечуючи високу адаптивність і зручність використання системи.

#### Список використаних джерел:

1. Дукельська К.Б., Гриньова О.Є. Інтелектуальний інтерфейс застосунку «Connectify». 28-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у XXI столітті». Зб. матеріалів форуму. Т. 6.1 – Харків: ХНУРЕ. 2024. – С.24-25. <https://doi.org/10.30837/IYF.IIS.2024.024> (дата звернення: 03.05.2025).

2. Falk K. Practical Recommender Systems. New York: Manning Publications, 2019. 432 p.

3. Aggarwal C. Recommender Systems. New York: Textbook, 2016. 448 p.