

РЕКОМЕНДАЦІЙНА СИСТЕМА ВИБОРУ ВІДЕОФІЛЬМІВ

Вінницький національний технічний університет

Анотація

Проаналізовано актуальність використання рекомендаційних систем, досліджено стратегії створення рекомендаційних систем. Розроблено системні вимоги до рекомендаційної системи вибору відеофільмів, а також побудовано її структурну схему, яка відображає основні програмні модулі.

Ключові слова: рекомендаційні системи, колаборативна фільтрація, інтелектуальний аналіз даних, врахування вподобань користувача.

Abstract

The relevance of using recommendation systems has been analyzed. Strategies of creating recommendation systems have been explored. System requirements of the recommendation system of choosing movies and thus, the structural scheme which shows main program units have been designed.

Keywords: recommendation systems, collaborative filtering, data mining, account of the user's preferences.

Вступ

В наш час найпоширенішим і найдоступнішим джерелом інформації є всесвітня мережа Інтернет, яка дає можливість зручного збору інформації про вподобання користувачів онлайн-ресурсів.

В Інтернеті налічується величезна кількість пошукових систем, каталогів, рейтингів, файлообмінників і т.п., а тому кількість контенту постійно збільшується, оскільки кожен сайт або сервіс надає певну інформацію, яка в процесі свого функціонування породжує ще більші обсяги інформації. Основне, але не єдине завдання рекомендаційних сервісів, - допомогти споживачеві контенту зорієнтуватися у результатах пошуку.

Рекомендаційна система - підклас системи фільтрації інформації, яка будує рейтинговий перелік об'єктів (фільми, музика, книги, новини, веб-сайти), яким користувач може надати перевагу. Для цього може використовуватися інформація з профілю користувача. Рекомендаційні системи віднайшли широке застосування в таких сферах як електронна комерція, соціальні мережі, веб-додатки тощо, де акцент робиться на користувача даних [1].

Метою роботи є аналіз методів формування рекомендацій, визначення системних вимог до рекомендаційної системи вибору відеофільмів.

Аналіз підходів до формування рекомендацій в інтелектуальних системах

У процесі роботи рекомендаційні системи збирають дані про користувачів, використовуючи поєднання явних і неявних методів.

Приклади явного збору даних [2]:

- користувач оцінює запропонований об'єкт за диференційованою шкалою;
- користувач ранжує групу об'єктів від найкращого до найгіршого;
- користувач вибирає кращий з двох запропонованих об'єктів;
- користувачу пропонують створити список його улюблених об'єктів.

Приклади неявного збору даних:

- спостереження за тим, що користувач оглядає в інтернет-магазинах або базах даних іншого типу;
- ведення записів про поведінку користувача онлайн;
- відстеження вмісту комп'ютера користувача.

Існують дві основні стратегії створення рекомендаційних систем: фільтрація вмісту і

колаборативна фільтрація. При фільтрації вмісту створюються профілі користувачів і об'єктів. Профілі користувачів можуть містити демографічну інформацію або відповіді на певний набір питань. Профілі об'єктів можуть містити назви жанрів, імена акторів, імена виконавців, тощо, або ж якусь іншу інформацію в залежності від типу об'єкта. Цей підхід застосований у проєкті Music Genome Project: музичний аналітик оцінює кожен композицію за сотнями різних музичних характеристик, які можна використати для виявлення музичних уподобань користувача [3].

При колаборативній фільтрації використовується інформація про поведінку користувачів у минулому — наприклад, інформація про придбання або оцінки. В цьому разі не має значення, з якими типами об'єктів ведеться робота, але при цьому можна брати до уваги неявні характеристики, які складно було б врахувати при створенні профілю. Основна проблема цього типу рекомендаційних систем — «холодний старт»: відсутність даних про користувачів чи об'єкти, які нещодавно з'явилися у системі [3].

Заснований на сусідстві підхід є історично першим в колаборативній фільтрації і використовується в багатьох рекомендаційних системах. У даному підході для активного користувача підбирається підгрупа користувачів схожих з ним. Комбінація ваг і оцінок підгрупи використовується для прогнозування оцінок активного користувача. У даного підходу можна виділити наступні основні кроки [4]:

- присвоїти вагу кожному користувачеві з урахуванням схожості його оцінок і активного користувача;
- вибрати кількох користувачів, які мають максимальну вагу, тобто максимально схожі на активного користувача. Дана група користувачів і називається «сусідами»;
- здійснити прогнозування оцінок активного користувача для неоцінених їм предметів з урахуванням ваг і оцінок сусідів.

Заснований на моделях підхід надає рекомендації, вимірюючи параметри статистичних моделей для оцінок користувачів, побудованих за допомогою таких методів як, метод баєсівських мереж, кластеризації, латентно-семантичної моделі, такі як сингулярний розклад, імовірнісний латентно-семантичний аналіз, прихований розподіл Діріхле і марківський процес прийняття рішень на основі моделей. Моделі розробляються з використанням інтелектуального аналізу даних, алгоритмів машинного навчання, щоб знайти закономірності на основі навчальних даних. Число параметрів в моделі може бути зменшено в залежності від типу за допомогою методу головних компонент [4].

Цей підхід є більш комплексним і дає більш точні прогнози, оскільки допомагає розкрити латентні фактори, що пояснюють спостережувані оцінки. Даний підхід має ряд переваг. Він обробляє розріджені матриці краще, ніж підхід заснований на сусідстві, що в свою чергу допомагає з масштабністю великих наборів даних. Недоліки цього підходу полягають в «дорогому» створенні моделі. Необхідний компроміс між точністю і розміром моделі, тому що можна втратити корисну інформацію у зв'язку із скороченням моделей [4].

Колаборативна фільтрація широко використовується в комерційних сервісах і соціальних мережах. Перший сценарій використання — це створення рекомендації щодо цікавої і популярної інформації на основі врахування «голосів» спільноти. Такі сервіси, як Reddit, Digg або DiCASTA — це типові приклади систем, що використовують алгоритми колаборативної фільтрації.

Інша сфера використання полягає у створенні персоналізованих рекомендацій для користувача, на основі його попередньої активності і даних про переваги інших, схожих з ним користувачів. Даний спосіб реалізації можна знайти на таких сайтах, як YouTube, Last.fm і Amazon, а також в таких геосоціальних сервісах, як Gvidi і Foursquare.

Розробка системних вимог до рекомендаційної системи вибору відеофільмів

Рекомендаційна система вибору відеофільмів, що розробляється, дозволить звичайним користувачам зручно, швидко і якісно підібрати відеоматеріал для перегляду на основі його особистих вподобань. Отримати рекомендації від даної системи зможуть як зареєстровані користувачі, так і гості, проте останні матимуть обмежені можливості у користуванні системою (наприклад, гості не матимуть можливість вести облік уже переглянутих відеофільмів). Системою буде керувати група адміністраторів, які слідкуватимуть за коректністю її роботи.

Рекомендаційна система вибору відеофільмів буде поділятися на 3 основні частини: android-

додаток, панель адміністратора, система управління базою даних. Кожна з категорій користувачів, зазначених вище матиме свої права доступу до відповідних розділів системи.

Практичність системи полягатиме в наданні користувачам таких можливостей:

- перегляд інформації про відеофільми;
- аналіз вподобань користувача;
- надання конкретних рекомендацій відеофільмів.

Основною вимогою реалізації є використання операційної системи Android 4.0 та мови програмування Java Android.

Структурно система буде поділятися на декілька модулів:

- інтерфейс користувача;
- модуль пошуку рекомендацій;
- база даних відеоматеріалів;
- база даних користувачів;
- модуль статистики.

Структурна схема рекомендаційної системи вибору відеофільмів зображена на рис.1.

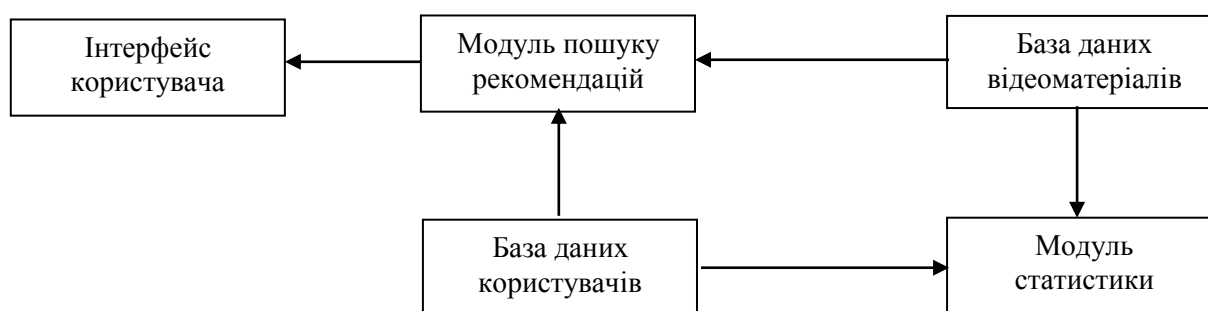


Рис. 1. Структурна схема рекомендаційної системи вибору відеофільмів

Висновки

Проаналізовано актуальність використання рекомендаційних систем, досліджено методи формування рекомендацій, а також переваги і недоліки методу колаборативної фільтрації як основи для створення рекомендаційної системи вибору відеофільмів. Визначено актуальність досліджень в даній області, що обумовлюється наявністю невирішених проблем в колаборативній фільтрації: розрідженість даних, проблема холодного старту, синонімія, білі ворони. Розроблено системні вимоги до рекомендаційної системи вибору відеофільмів, а також побудовано її структурну схему, яка відображає основні модулі: інтерфейс користувача, модуль пошуку рекомендацій, модуль статистики, бази даних відеоматеріалів та користувачів.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Recommender Systems Handbook / R. Francesco, R. Lior, S. Bracha, K. B. Paul. – Dordrecht: Springer, 2015. – 1009 p.
2. McLeod D. Collaborative Filtering for Information Recommendation Systems [Електронний ресурс] / D. McLeod, A. Y. Chen // Research Reports. – 2009. – Режим доступу до ресурсу: http://research.create.usc.edu/cgi/viewcontent.cgi?article=1101&context=nonpublished_reports.
3. Koren Y. Matrix Factorization Techniques for Recommender Systems / Y. Koren, B. Robert, V. Chris. // Journal Computer. – 2009. – №42 (8). – pp. 30–37.
4. Su X. A Survey of Collaborative Filtering Techniques / X. Su, T. M. Khoshgoftaar. // Advances in Artificial Intelligence. – 2009. – №2009 (4). – pp. 1–19.

Щербань Вадим Сергійович — студент групи 1-КН-126, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: vadim_scherban@mail.ru;

Гайдейчук Юрій Анатолійович — студент групи 1-КН-126, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: yuriy.haydeychuk@gmail.com;

Науковий керівник: **Яровий Андрій Анатолійович** — д-р техн. наук, професор, професор кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця, email: a.yarovyy@gmail.com.

Vadym S. Shcherban — Department of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: vadim_scherban@mail.ru;

Yuriy A. Haideichuk — Department of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: yuriy.haydeychuk@gmail.com;

Supervisor: **Andriy A. Yarovyy** — Dr. Sc. (Eng.), Professor, Professor of Computer Science Chair, Vinnytsia National Technical University, Vinnytsia, e-mail: a.yarovyy@gmail.com.