

Технічні науки

УДК 004.67

Вараниця С. А.

магістрант

Львівський національний університет

"Львівська політехніка"

Білас О. Є.

професор кафедри програмного забезпечення

Львівський національний університет

"Львівська політехніка"

Вараниця С. А.

студент

Львовский национальный университет

"Львовская политехника"

Билас О. Е.

профессор кафедры программного обеспечения

Львовский национальный университет

"Львовская политехника"

Varanytsia S. A.

student

Lviv National University "Lviv Polytechnic"

Bylas O. E.

Professor of Software Department

Lviv National University "Lviv Polytechnic"

**ГІБРИДНІ РЕКОМЕНДАЦІЙНІ СИСТЕМИ НА ОСНОВІ МЕДІЙНИХ
ВПОДОБАНЬ КОРИСТУВАЧА**

**ГИБРИДНЫЕ РЕКОМЕНДАЦИОННЫЕ СИСТЕМЫ НА БАЗЕ
МЕДИЙНЫХ ПРЕДПОЧТЕНИЙ ПОЛЬЗОВАТЕЛЯ**

**HYBRID RECOMMENDER SYSTEM BASED ON USER MEDIA
PREFERENCES**

Анотація: Висвітлено основні проблеми та методи рішень даних проблем на основі створення гібридної рекомендаційної системи зі спрямуванням на використанні інформації по медійних вподобаннях користувачів.

Ключові слова: рекомендаційна система, алгоритми, рекомендації, матриці, медіа.

Аннотация: Освещены основные проблемы и методы их решений на базисе создания гибридной рекомендационной системы с ориентировкой на использовании информации по медиа предпочтениях пользователей.

Ключевые слова: рекомендационная система, алгоритмы, рекомендации, матрицы, медиа.

Summary: Covering main problems and managing their solving methods on the base on creating the hybrid recommender system with using the data about user media preferences.

Key words: Recommender system, recommendation system, algorithms, media, SVD.

Вступ

Швидке розростання веб-мереж стало причиною утворення баз великої кількості інформації, використовуючи яку ми можемо створювати більш ніж точні рекомендації для користувачів і, що головне, брати до уваги не тільки їхні попередні відгуки та оцінки, а й вчитись на теперішніх відгуках. На даний момент більше розвинуті контентно-спрямовані рекомендаційні системи. Це системи, рекомендації яких базуються на одному типі інформації. Тобто якщо ми рекомендуємо користувачу певний фільм, то рекомендація створюється виключно на алгоритмах, які беруть до уваги смаки цього та інших користувачів виключно по кінострічках. Даний варіант підвищує початкову точність рекомендацій, і користувач отримує інформацію, яка базується на попередніх даних того ж типу.

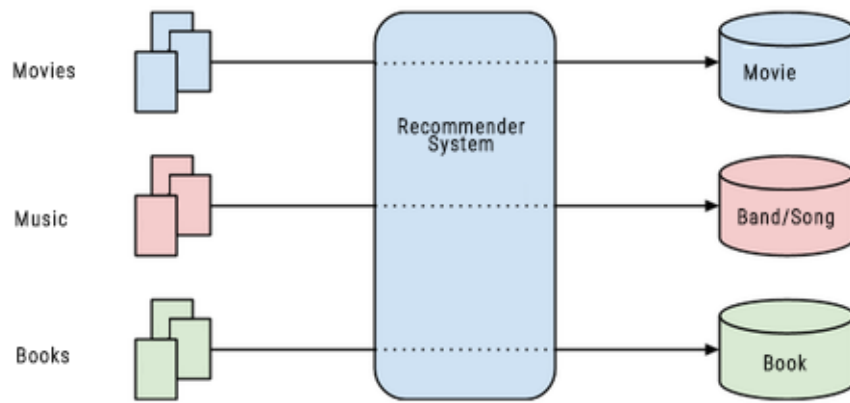


Рис.1. Загальний вигляд рекомендаційної системи [3]

Але маючи доступ до інших медійних даних (напр. до музичних вподобань), гібридна рекомендаційна система здатна створювати і надавати рекомендації, які сприймають різнотипні вхідні дані. В результаті базуючись на інформації по музичних вподобаннях користувача ми можемо не тільки надавати рекомендації по музичних гуртах, але й точніше вгадувати вірогідність вподобання користувача того чи іншого фільму, або тої чи іншої гри.

Виклад основного матеріалу

Розглянемо один з основних алгоритмів для роботи з великою кількістю даних, алгоритм SVD (сингулярного представлення матриці). Нехай у нас буде матриця, у якій є дані по фільмах, музиці, книгах (оцінки, рейтинги, факти вподобання та відмов тощо). Відповідно отримуємо матрицю

$$R = (r_{i,q})_{i=1, q=1}^{N, M}, \text{ в якій вже записані відомі нам дані.}$$

Ця матриця в більшості випадків буде сильно розріджена, з цією проблемою потрібно боротись підтягуванням даних сусідніх вподобань, так як один користувач наврядчи зможе оцінити велику кількість медійних товарів (крім випадків багатомісячного, або й багаторічного користування).

Відповідно для того щоб мати змогу спрямовано працювати з

подібною розрідженою матрицею використовується так зване сингулярне представлення:

$$R = UDV^T$$

R — великорозмірна матриця розміру $N \times M$

Головна перевага даного представлення в тому, що вона дає оптимальне приближення, навіть якщо в матриці D залишити тільки F перших діагональних елементів, а всі інші обнулити.

У випадку з крос-медійними даними, де нам потрібно зважати не тільки на однотипні оцінки та вподобання нам необхідно представити кожного користувача системи вектором з f факторів U_i , а кожну категорію (тип) – вектором з f факторів V_j . Відповідно, щоб передбачити рейтинг користувача i типу j ми беремо їх скалярне представлення $U_i V_j = U_i^T V_j$.

Це виділяє проблему коли по вже відомим оцінкам та реакціям необхідно передбачити сприйняття користувачем наступної рекомендації по різнотиповій інформації.

Після введення предикатів, для початку тільки базових, нам потрібно знайти найкращі з них. Тобто знайти ті предикати, які дають найменшу долю похибки. Це ми знаходимо наступним чином:

$$L(\mu, b_i, b_a, v_a, u_i) = \sum_{(i, a) \in D} (r_{i, a} - \mu - b_i - b_a - v_a u_i)^2$$

Функцію $L(\mu, b_i, b_a, v_a, u_i)$ мінімізуємо градієнтним спуском, а для компенсування ефекту перенавчання системи варто додати параметр регулювання.

В результаті функція виглядає наступним чином:

$$b^*, q^*, p^* = \arg \min_{b, p, q} \sum_{(i, a) \in D} (r_{i, a} - \mu - b_i - b_a - q_a^T p_i)^2 + \lambda (\sum_i b_i^2 + \sum_a b_a^2 + \sum_a q_a^2 + \sum_i p_i^2),$$

Де λ — параметр регуляції.

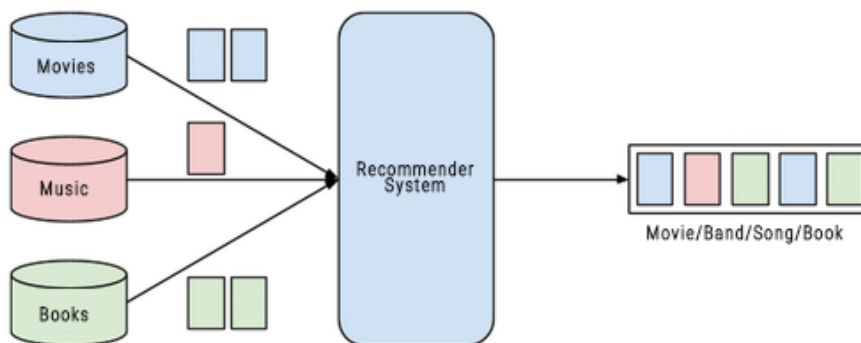


Рис.2. Вигляд гібридної медійної рекомендаційної системи [3].

В результаті виконання доволі простих маніпуляцій з однієї сторони ми збільшуємо ризик початкових помилкових рекомендацій, особливо якщо система не здатна відразу проаналізувати схожі початкові запити, але якщо вирішити проблему початкового рішення шляхом рішення методу холодного старту, аналізуючи дані користувачів схожих метрик, то точність рекомендацій нормалізується, але різноманітність збільшиться в декілька разів, так як будуть братись до уваги дані з декількох типів, і за рахунок перенавчання системи ми можемо навіть пропонувати елемент з однієї категорії навіть не зважаючи на оцінки користувача в даному типі (перевіряючи тільки чи він бачив/чув даний медійний продукт), але відсоток успішності рекомендації буде високим.

Висновки

Алгоритми генерування рекомендацій мають проходити не тільки перевірки на швидкодію, якість та кількість, але й на перевірку на крос-медійну варіацію, цим загальна цінність рекомендаційної системи підвищується не тільки на самонавчальному рівні системи, але й вдосконалюється якість наданих рекомендацій.

Алгоритм SVD продемонстрував гнучкість і легкість у використанні, відповідно навіть якщо елементів одного типу ми не мали взагалі (наприклад інформації по фільмах, які користувач вподобав), то базуючись

на інших даних (музичних смаках та книгах) рекомендаційна система запропонувала фільми, 3 з 5 яких користувач вподобав, і 2 з яких бачив раніше.

Даний метод вимагає багато доопрацювань, та все ж переваги його використання видні ще на самому початку його впровадження та аналізу його роботи, так як за допомогою простих видозмін можна отримати крос-контентну самодостатню рекомендаційну систему.

Література:

1. *N. Rubens, D. Kaplan, M. Sugiyama. Recommender Systems Handbook: Active Learning in Recommender Systems (eds. F. Ricci, P.B. Kantor, L. Rokach, B. Shapira). Springer, 2011.*
2. *Zan Huang, Xin Li, Hsinchun Chen Link Prediction Approach to Collaborative Filtering (англ.)// University of Arizona, USA : Матеріали конф. / JCDL'05, Denver, Colorado, USA, June 7–11, 2005. — 2005.*
3. *Jin R., Si L., and Zhai C. Preference-Based Graphic Models for Collaborative Filtering // Proc. 19th Conf. Uncertainty in Artificial Intelligence (UAI 2003), August 2003.*
4. *Koren, Y.; Bell, R.; Volinsky, C. (07 August 2009). Matrix Factorization Techniques for Recommender Systems.. Computer (IEEE) 42 (8): 30–37.*