

**Шарабан О.Ю.**

*студент;*

*Науковий керівник: Лєсна Н.С.*

*доктор технічних наук, професор,*

*Харківський національний університет радіоелектроніки*

## **РОЗРОБКА ТА ДОСЛІДЖЕННЯ МЕТОДІВ ФОРМУВАННЯ МУЗИЧНИХ РЕКОМЕНДАЦІЙ**

Рекомендаційні системи стали основними додатками в електронній комерції та доступі до інформації, надаючи пропозиції, які ефективно скорочують великі інформаційні простори, щоб користувачі отримували ті об'єкти, які найкращим чином відповідають їхнім потребам і перевагам. За час існування інформаційних технологій було запропоновано безліч методів для формування рекомендацій, включаючи методи, засновані на контенті, колаборативні методи, методи, засновані на знаннях і інші. Для підвищення продуктивності ці методи іноді комбінуються в гібридні рекомендаційні методи.

У даній роботі розглядаються сфери застосування методів формування рекомендацій, зокрема, в музичному середовищі, а також досліджуються методи формування рекомендацій, такі як колаборативна фільтрація (КФ) [1] та фільтрація контенту [2], що створюють рекомендації музичних треків для користувачів. Також автором пропонується новий метод, який зможе удосконалити існуючі та підвищити якість формування рекомендацій.

Все частіше використовуються системи колаборативних рекомендацій, які аналізують поведінку користувачів в минулому і знаходять відносини між елементами і користувачами [3]. Наприклад, в системі рекомендацій музики [www.ok.ru](http://www.ok.ru), другій за значимістю соціальної платформі в Росії, була запропонована модель, в якій всі дані, отримані з історії активності користувачів, метаданих контенту і соціальних мереж, об'єднуються у «смаковому графі». На цьому графі вершини представляють різні музичні об'єкти, такі як користувачі, треки і виконавці, а дуги представляють відносини між сутностями (користувачеві подобається трек або виконавець, який схожий на інших виконавців). Для виконання різних завдань рекомендації користувач переміщається по графу, вважаючи за краще дуги з великою вагою. Для створення рекомендацій пропонується випадковий шлях, де елементи з більш високою ймовірністю стійкого стану, найбільш ймовірно, будуть доречні для користувача. Ще одна рекомендаційна задача – вибрати з набору елементів найбільш відповідні для користувача. Цього можна домогтися, проходячи по графу переваг користувача, підраховуючи всі відвідування вершин, вважаючи вершини з більш високими значеннями більш придатними для користувача.

Фільтрація на основі контенту залежить від контенту об'єктів, представлених конкретними дескрипторами, і заснована на схожості об'єктів. Перш ніж застосовувати цю техніку, об'єкти повинні бути описані автоматичними або ручними методами. В області музики автоматичні методи виконуються алгоритмами, які витягують низький рівень (частота дискретизації

аудіо файлу, спектральний центр звукового кадру), середній рівень (тональність, ритм, гармонія, інтенсивність, структура) і високорівневий (аналіз схожих гітарних соло). Ручні методи надаються користувачами (створення тегів для музичних треків) або експертами (надання точної інформації про виробничий процес, біографії виконавця або жанрової інформації). Після збору цієї інформації фільтрація на основі контенту прогнозує, які об'єкти сподобаються користувачу, виходячи з його переваг, а не покладається на рейтинги інших користувачів. Для розрахунку подібності між двома об'єктами розглядається відстань між елементами. Коли об'єкти описуються числовими атрибутами, використовуються прості метрики, такі як евклідова відстань.

Методи колаборативної фільтрації – це дуже ефективний спосіб створення рекомендацій в соціальних мережах, однак вони також не позбавлені певних недоліків. Для їх коректної роботи потрібен великий обсяг даних. Але чим він більше, тим імовірніше знайти елементи, які не оцінені користувачами, тобто існує проблема з розрідженістю даних. Іншою проблемою є так звана проблема холодного старту. Оскільки колаборативна фільтрація заснована на оцінках, які користувач ставить конкретним об'єктам, нових користувачів і нові елементи складніше класифікувати, оскільки вони ще не мають пов'язаної інформації.

Методи фільтрації на основі контенту дуже ефективні, коли немає великого набору даних, доступних користувачам [4], однак вони також створюють певні проблеми. Найбільш актуальною є труднощі з рекомендаціями нових предметів, якщо припустити, що користувачеві завжди будуть радити об'єкти, дуже схожі на інші об'єкти, які вже є в його профілі. Крім того, ці системи все ще стикаються з безліччю проблем з отриманням дескрипторів високого рівня, які можуть мати велике значення для користувача.

У надії уникнути обмежень будь-якої системи і підвищити швидкість надання рекомендацій, часто використовуються комбіновані (гібридні) методи формування рекомендацій. Гібридні методи створюються шляхом додавання характеристик для КФ-моделей, додавання КФ-характеристик до моделей на основі контенту, поєднання КФ з системами на основі контенту або інших систем, або комбінування різних КФ-алгоритмів.

Існує декілька способів, якими можна поєднати роботу кількох методів формування рекомендацій у гібридних методах, а саме:

- монолітна організація, коли один з методів обирається основним, а інші лише посилюють та підтримують його роботу;
- паралельна організація, при якій кожен з методів працює окремо, а потім результати їх роботи об'єднуються за певними принципами;
- конвеєрна організація, коли всі методи працюють послідовно, при цьому вхідними даними для кожного наступного метода є вихідні дані попереднього.

У своїй роботі автор обмежується лише двома рекомендаційними методами, які є найбільш популярними та результативними на сьогоднішній день, а саме колаборативною фільтрацією та фільтрацією на основі контенту для отримання потенційно найбільш якісних рекомендацій.

Для того, щоб результати роботи методів були найбільш оптимальними, вводиться поняття принципу подвійної організації та розробляється алгоритм,

який поєднуватиме у собі принципи паралельної та конвеєрної організації і використовуватиме відомі методи наступним чином:

- фільтрація відбувається в двох незалежних робочих блоках;
- у одному з блоків першим працюватиме метод колаборативної фільтрації, а по закінченню його роботи отримані результати опрацює метод на основі контенту;
- у іншому блоці, навпаки, спочатку відпрацює фільтрація на основі контенту, а потім колаборативна фільтрація;
- після отримання результатів з двох блоків для кожного об'єкту розраховується сума рейтингів, отриманих паралельними блоками та виводяться об'єкти з найбільшим рейтингом як рекомендації.

Таким чином, гібридний метод, що працює за принципом подвійної організації, за рахунок подвійного опрацювання даних та відсіву може вирішити наступні проблеми існуючих рекомендаційних методів, такі як надмірна схожість рекомендацій, потенційно мала кількість або відсутність рекомендацій взагалі за рахунок компенсації даних з одного блоку даними з іншого. Але у такого методу є і певні недоліки, такі як високе споживання обчислювальних ресурсів та зниження швидкості роботи через трудоємність процесів, що дає гарні приводи для його подальшого розвитку та оптимізації.

Ці недоліки можуть бути легко скомпенсовані встановленням даної програмної системи на більш потужному сервері.

#### **Список використаних джерел:**

1. Goldberg D., Nichols D., Oki B. M., and Terry D. «Using collaborative filtering to weave an information tapestry», Communications of ACM, vol. 35, no. 12, pp. 61–70, 1992.
2. Resnick P. and Varian H. R. «Recommender systems», Communications of the ACM, vol. 40, no. 3, pp. 56–58, 1997.
3. Goldberg K., Roeder T., Gupta D., and Perkins C. «Eigentaste: a constant time collaborative filtering algorithm», Information Retrieval, vol. 4, no. 2, pp. 133–151, 2001.
4. Miller B. N., Konstan J. A., and Riedl J. «PocketLens: toward a personal recommender system», ACM Transactions on Information Systems, vol. 22, no. 3, pp. 437–476, 2004.

**Шевченко С.О.**

*студент,*

*Науковий керівник: Лєсна Н.С.*

*доктор технічних наук, професор,*

*Харківський національний університет радіоелектроніки*

### **РОЗРОБКА РОЗПОДІЛЕНОЇ СИСТЕМИ УПРАВЛІННЯ ЕНЕРГОСПОЖИВАННЯ ОБЧИСЛЮВАЛЬНИХ СЕРВЕРІВ**

У сучасному світі залишається все менше підприємств, котрі не використовують у своїй діяльності можливості центрів обробки даних (ЦОД) [1], серверів чи мережевих сховищ (NAS) [2]. Сервери використовують для