

В.Ю. Кучерук, д.т.н., професор, М.В. Глушко, магістр, О. Митковський

## ПОКРАЩЕННЯ АЛГОРИТМУ ІТЕМТОІТЕМ МЕТОДУ КОЛАБОРАТИВНОЇ ФІЛЬТРАЦІЇ ДЛЯ РОЗРОБКИ РЕКОМЕНДАЦІЙНИХ СИСТЕМ ШЛЯХОМ ОЦІНКИ РЕЛЕВАНТНОСТІ

Ключові слова: колаборативна фільтрація, itemtoitem, косинус, вектор, матриця

Рекомендаційна система- підклас системи фільтрації інформації, яка будує рейтинговий перелік об'єктів, яким користувач може надати перевагу. Існують дві основні стратегії створення рекомендаційних систем: фільтрація вмісту і колаборативна фільтрація. В даній науковій роботі використовується колаборативна фільтрація.

Колаборативна фільтрація - метод, який використовується деякими рекомендаційними системами. Колаборативна фільтрація має два значення: вузьке і більш загальне. В цілому, колаборативна фільтрація— процес фільтрації інформації або зразків за допомогою методів за участю співробітництва між декількома агентами, точками зору, джерелами даних і т.д. Застосування колаборативної фільтрації, як правило, пов'язане з дуже великими наборами даних. В нашому випадку колаборативна фільтрація - це один з методів побудови прогнозу рекомендаційних систем, який використовує відомі переваги (оцінки предметів, дії, поведінку) групи користувачів для прогнозування невідомих переваг іншого користувача. Основне припущення колаборативної фільтрації полягає в наступному: ті, хто однаково оцінювали будь-які предмети в минулому, схильні давати схожі оцінки інших предметів і в майбутньому.

Для прогнозування за допомогою алгоритму itemtoitem потрібно використати предмет, на відмінну алгоритму «usertouser», який прогнозує на основі вподобань користувачів. Предмет – це змінна, яка позначається як «0» - предмет не виконаний, або як «1» - предмет виконаний. Змінна предмету може являти певну дію, зроблену з товаром на сайті: найпопулярніші застосування предмету це оцінка товару, або його покупка. Тобто, якщо користувач купує товар, то його значення предмету = 1. Всі користувачі, які купили товар – об'єднують в когорті (групи людей, схожі між собою за певними ознаками, поведінкою, і вже новому користувачу, який буде відвідувати сторінку певного товару – йому запропонується товар, найбільш схожий на той, який він переглядає, на основі вподобань користувачів.

Інтуїтивно зрозуміло, що для рекомендації користувачеві №1 будь-якого продукту, вибрати потрібно з продуктів, які подобаються якимось користувачам 2-3-4-п., Які найбільш схожі за своїми оцінками на користувача №1. Отримати чисельне вираження цієї «схожості» користувачів можемо за допомогою векторів в N-вимірному просторі продуктів, а порівнювати вектора ми можемо наступними відомими методами:

- Косинусна міра
- Коефіцієнт кореляції Пірсона
- Евклідова відстань
- Коефіцієнт Танімото
- Манхеттенський відстань і т.д.

Далеко не завжди у користувачів є можливість виставляти оцінки предметів. Тобто для колаборативної фільтрації можуть бути доступні лише двійкові дані (купував користувач предмет чи ні). Проте даний алгоритм не враховує релевантність (відповідність до пошукового запиту) сторінки для користувача. Тобто якщо людина потрапляє на сторінку товару на веб-ресурсі, даний алгоритм не знає чи задовольняє цей товар користувача, чи ні. Оскільки він реагує тільки на оцінку товару або на його покупку. По статистиці в світі – відсоток конверсії на сайтах становить всього 2-3%. Тобто, рекомендаційна система обробляє тільки 2-3% даних для видачі інформації.

Тому в цій роботі розроблений алгоритм, який враховує релевантність сторінки для 100% користувачів, і на основі цього рекомендує. Суть алгоритму полягає в тому, що замість покупки чи оцінки товару, ми, присвоюючи кожному користувачу (User) – ідентифікатор (ID), запускаємо таймер при перегляді URL-сторінки оголошення (карточка товару). Якщо при цьому, користувач перебуває на сторінці більше 15 секунд, то ця сторінка вважається релевантною, якщо менше, то ні. ( $t > 15 \text{ сек.} = 1$ ,  $t < 15 \text{ сек.} = 0$ )

Спочатку даний алгоритм розраховував подібність предметів як косинус між векторами покупок в матриці користувачів та предметів. (2-3%)

Розглянемо роботу алгоритму на прикладі статистики, де за предмет – приймається час перегляду URL-сторінки оголошення:

| Користувач | Предмет 1 | Предмет 2 | Предмет 3 |
|------------|-----------|-----------|-----------|
| Віктор     | 1         | 0         | 1         |
| Микола     | 0         | 1         | 1         |
| Марія      | 0         | 1         | 0         |

В цьому випадку косинус між «Предмет 1» і «Предмет 2» розраховується так:

$$\frac{(1,0,0) \cdot (0,1,1)}{\|(1,0,0)\| \cdot \|(0,1,1)\|} = 0;$$

між «Предмет 1» і «Предмет 3»:

$$\frac{(1,0,0) \cdot (1,1,0)}{\|(1,0,0)\| \cdot \|(1,1,0)\|} = \frac{1}{\sqrt{2}} \approx 0,71;$$

і між «Предмет 2» і «Предмет 3»:

$$\frac{(0,1,1) \cdot (1,1,0)}{\|(0,1,1)\| \cdot \|(1,1,0)\|} = \frac{1}{2} = 0,5 .$$

Таким чином, користувач, що знаходиться на сторінці опису «Предмету 1», отримає «Предмет 3» в якості рекомендації; на сторінці «Предмету 2» - «Предмет 3» і на сторінці «Предмету 3» - «Предмет 1» (і потім «Предмет 2»). В даному алгоритмі використовується один коефіцієнт на кожен пару предметів (косинус), на підставі якого і створюються рекомендації. Тобто для n предметів буде потрібно розрахувати і зберегти n (n-1) / 2 косинусів.

Отже, робота алгоритму буде проходити в декілька етапів, оскільки спочатку потрібно визначити значення «предмету», яке буде дорівнювати переглянутій сторінці URL оголошення, на якій користувач перебуває не менше 15 секунд, цим самим оцінюючи релевантність сторінки до пошуку користувача. Тільки після визначення таких відбуваються подальші обрахунки із знаходження косинусної міри, при якій береться до уваги 100% інформації про інтереси користувачів, цим самим покращуючи алгоритм ItemtoItem методу колаборативної фільтрації для розробки рекомендаційних систем шляхом оцінки релевантності.

#### Список літературних джерел:

1. Рекомендаційна система [Електронний ресурс] // uk.wikipedia.org. – 2016. – Режим доступу до ресурсу: [https://uk.wikipedia.org/wiki/%D0%A0%D0%B5%D0%BA%D0%BE%D0%BC%D0%B5%D0%BD%D0%B4%D0%B0%D1%86%D1%96%D0%B9%D0%BD%D0%B0\\_%D1%81%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0](https://uk.wikipedia.org/wiki/%D0%A0%D0%B5%D0%BA%D0%BE%D0%BC%D0%B5%D0%BD%D0%B4%D0%B0%D1%86%D1%96%D0%B9%D0%BD%D0%B0_%D1%81%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0).
2. Коллаборативная фильтрация [Електронний ресурс] // habrahabr.ru. – 2012. – Режим доступу до ресурсу: <https://habrahabr.ru/post/150399/>.
3. SlopeOne [Електронний ресурс] // en.wikipedia.org. – 2015. – Режим доступу до ресурсу: [https://en.wikipedia.org/wiki/Slope\\_One](https://en.wikipedia.org/wiki/Slope_One).