

Савчук Т.О., Сакалюк А.В.

Вінницький національний технічний університет, Вінниця, Україна

Модель рекомендаційної системи на основі колаборативної фільтрації

Колаборативна фільтрація використовує схожість думок різних користувачів для видачі рекомендацій щодо об'єктів. Рекомендаційні системи на основі колаборативної фільтрації - зручна альтернатива класичним пошуковим алгоритмам, так як використовують фактори, які неможливо отримати з технічного аналізу інформації. Тому впровадження подібних механізмів прискорить процес пошуку потрібної інформації, збільшить її повноту і достовірність.

З урахуванням особливостей основних моделей колаборативної фільтрації представимо проблему вироблення рекомендацій таким чином: нехай C – множина користувачів, S – множина пропонованих об'єктів. Функція корисності u_{ij} описує корисність об'єкта s_j для користувача c_i , де i - порядковий номер користувача, j - порядковий номер об'єкта. В результаті отримаємо матрицю U значень функції корисності для кожної пари користувач-об'єкт.

Моделі, що базуються на зберіганні вхідних даних (Memory-Based Collaborative Filtering), оперують всією матрицею значень функції корисності, тому потребують додаткових затрат на збереження великих масивів інформації. Подальший аналіз ґрунтується на використанні кореляційної функції для визначення міри подібності:

- подібність клієнтів – це кореляція рядків матриці U ;
- подібність об'єктів - це кореляція стовпців матриці U .

Традиційні підходи не виконують попередній аналіз, і тому кількість обчислень безпосередньо при запиті збільшується з числом користувачів і об'єктів, тобто з ростом потужності матриці U . Такі алгоритми неможливо використовувати на великих наборах даних, якщо вони не понижують розмірності, що призводить до зниження рівня якості рекомендацій.

Тому в потужних системах необхідно використовувати латентні моделі (Latent Models for Collaborative Filtering), які дозволять встановити невідому внутрішню структуру зв'язків між об'єктами. Перевагою таких методів є врахування не лише парної лінійної кореляції, а й взаємодії множини об'єктів між собою. Основними характеристиками таких моделей є:

- оцінювання профілів клієнтів і об'єктів (профіль - це вектор прихованих характеристик);
- зберігання профілів замість зберігання U ;
- подібність клієнтів та об'єктів – це подібність їх профілів.

Як показано на рисунку 1, два головних етапи роботи рекомендаційної системи виявлення знань та рекомендації [2]. На першому етапі вподобання користувача визначаються за допомогою доступу до профілю користувача та log-даних. На наступному етапі отримані дані використовуються для визначення можливих рекомендацій. Дані рекомендації динамічно видаються у відповідь на запит користувача. На етапі виявлення знань, використовуючи попередню інформацію про поведінку кожного користувача, видобувається інформація, що відображає вподобання користувача. Процес отримання даної інформації складається з наступних кроків:

1. Підготовка даних. Log-файли містять корисну інформацію про відвідування та перегляд певних об'єктів користувачами. Її потрібно відфільтрувати, аби не враховувати неважливу інформацію. В якості вхідних даних система отримує інформацію про дії, виконані користувачем, або про оцінки, виставлені об'єктам з набору даних.
2. Кластеризація документів. Інформація про користувача обробляється та розподіляється на кластери у відповідності до схожості. В результаті отримується множина кластерів, яка відображає набір об'єктів, що містять схожий контент.
3. Створення матриці подібності. На цьому кроці будується таблиця розмірністю $m \times n$, де кожна комірка представляє рівень зацікавленості i -го користувача до j -го кластера об'єктів. Отримана матриця нормалізується для подальшої обробки [3].
4. Кластеризація користувачів. Для початку необхідно вибрати властивості, які характеризують користувачів, ними можуть бути як кількісні, так і якісні характеристики. Потім потрібно зменшити розмірність простору характеристичних векторів, тобто виді-

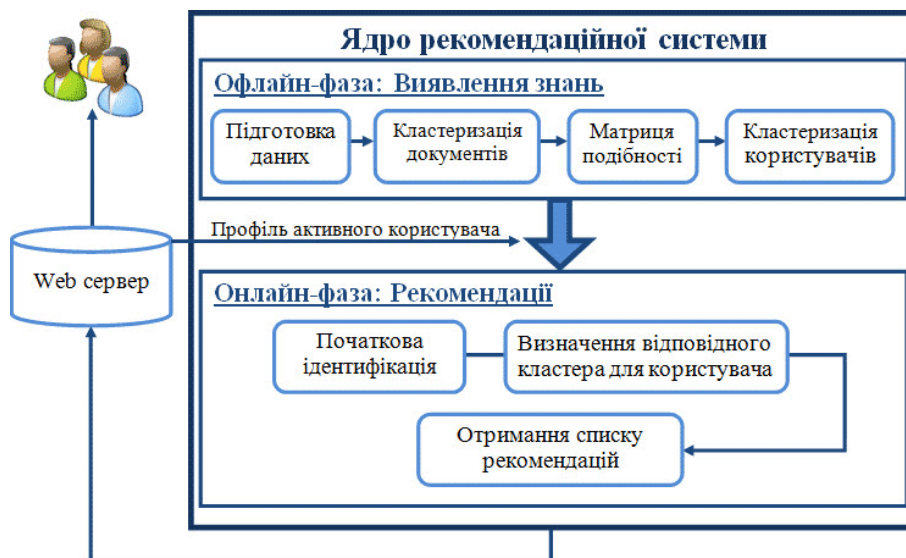


Рис. 1. Модель рекомендаційної системи

лити найбільш важливі властивості об'єктів. Зменшення розмірності прискорює процес кластеризації і в ряді випадків дозволяє візуально оцінювати результати.

На початку роботи алгоритму рекомендаційної системи визначається найближчий за вподобаннями кластер користувачів і розраховуються відповідні рекомендації для нього.

Запит рекомендацій спочатку породжує запит профілю користувача. Рекомендації для користувачів, які виконали недостатню кількість дій (наприклад, менше 10), системою не повертаються - їх профілі поки що побудовані недостатньо добре і рекомендації їм краще будувати без урахування колаборативної фільтрації. Наприклад, для користувача під час холодного старту, можна повертати найпопулярніші об'єкти або об'єкти з найвищим рейтингом. Або ж, з іншого боку, для прискорення навчання є сенс показувати об'єкти, найбільш характерні для конкретних факторів або найбільш суперечливі - з максимальним розкидом оцінок користувачів.

Основними завданнями персоналізації пропозицій є:

- видати оцінку товару s_j для клієнта c_i ;
- видати клієнту c_i список рекомендованих товарів;
- інформувати клієнта про новий товар (up-selling);
- сегментувати клієнтську базу;
- виділити цільові аудиторії за інтересами.

Таким чином, запропонована модель на основі колаборативної фільтрації дасть можливість подолати традиційні проблеми рекомендаційних систем шляхом застосування інтелектуальних підходів до вироблення рекомендацій. Звідси виникає актуальність впровадження засобів інтелектуального аналізу даних для пошуку зв'язків між користувачами, зокрема кластеризації їх вподобань для виявлення прихованих зв'язків та групування користувачів.

Література. 1. Su, Xiaoyuan. A Survey of Collaborative Filtering Techniques / Xiaoyuan Su, Taghi, M. Khoshgoftaar. – Hindawi Publishing Corporation USA, 2009. – 215 p. 2. Benjamin Marlin. Collaborative Filtering a Machine Learning Perspective / National Library of Canada = Bibliothèque nationale du Canada, 2004. 3. Т.О. Савчук, А.В. Сакалюк. Застосування кластерного аналізу для вдосконалення алгоритму колаборативної фільтрації. / Вісник Хмельницького національного університету, 2011. №1. 4. Т. Сегаран. Программируем коллективный разум; пер. с англ. А. Слинкина – СПб: Символ-Плюс, 2008. – 368 с.