

ЗАСТОСУВАННЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ПРИ КОЛАБОРАТИВНІЙ ФІЛЬТРАЦІЇ

Сакалюк А.В.

Науковий керівник – к.т.н., проф. Савчук Т.О.

Світ сучасної економіки наповнений товарами, властивості яких важко оцінити завчасно, виходячи з характеристик, даних виробником. Також важливим аспектом є те, що зазвичай таких параметрів не один десяток, і просумувати їх цінність важко. Рішенням даної проблеми є використання рекомендацій тих, хто вже знайомий з даним об'єктом. Колаборативна фільтрація надає колективній обробці інформації принципово нову якість. Кожний апробує свої об'єкти, а з допомогою колаборативних алгоритмів індивідуальний досвід враховується для всієї системи, звільняючи користувачів від непотрібних проб і помилок.

Експертні фільтри не справляються з потоком інформації. Колаборативна технологія дозволяє урівноважити цей дисбаланс: який би темп не набрала генерація контенту, їй протидіє не менш продуктивна масова селекція.

Принципи колаборативної фільтрації:

1. Користувачі надають дані про свої уподобання, оцінюючи товари інтернет-магазину або вітрини.
2. Чим більше даних надає про себе користувач, тим більш якісні рекомендації він отримує.
3. Чим більше сумарний обсяг даних, тим краще рекомендації.

Застосування простої кластеризації або статистичних моделей кластеру вважається перспективним засобом для проблеми розрідженості оцінок. Замість використання розріджених матриць рейтингу користувач-об'єкт, матриця спочатку згладжується на основі середнього рейтингу користувача.

Зниження розмірності вхідних даних при кластеризації дозволяє:

1. Зменшити обсяг обчислень при розрахунку відстані між векторами.
2. Підвищити якість кластеризації за рахунок особливих властивостей векторів головних компонент.
3. Зменшити час кластеризації.

Завдяки використанню кластерного аналізу для розбиття користувачів за схожістю вподобань вдосконалений алгоритм роботи рекомендаційної системи забезпечує зменшення часу на обробку даних та підвищує точність рекомендацій.

Колаборативна фільтрація з попереднім згладженням набору даних та кластеризацією користувачів забезпечує високу якість рекомендацій для користувачів.