



РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА НА ОСНОВЕ КОЛЛАБОРАТИВНОЙ ФИЛЬТРАЦИИ

Бескоровайный В.В., Мельничук Ю.В.

Харьковский национальный университет радиоэлектроники

Количество доступной информации в компьютерных сетях (клиентских средах) стремительно увеличивается, многократно превышая возможности человека по ее переработке. Для ее предварительной обработки необходимо создание фильтрующих персонализированных систем. Одним из наиболее распространенных видов таких систем являются рекомендательные системы, представляющие собой приложения и технологии, формирующие советы и предложения для пользователей относительно интересующих их объектов [1].

Разрабатываемая рекомендательная система предназначена для поиска объектов, которые представляют определенный интерес для пользователя. Суть технологии ее работы может быть представлена следующим образом. Имеется список пользователей $U = (u_1, u_2, \dots, u_m)$ и объектов $I = (i_1, i_2, \dots, i_n)$. В ходе взаимодействия с системой пользователи знакомятся с объектами, формируя матрицу рейтингов R (где $r_{u_a, i}$ – рейтинг объекта $i \in I$ у пользователя $u_a \in U$). Задачей рекомендательной системы является вычисление предсказаний и выдача рекомендаций пользователю относительно вновь появляющихся объектов.

Проведенный анализ современных рекомендательных систем позволил установить, что для генерации рекомендаций в них используют ряд подходов, наибольшее распространение среди которых получили: контентный; коллаборативный; основанный на знаниях; основанный на контексте.

Для реализации системы выбран коллаборативный подход. Основная идея алгоритмов коллаборативной (совместной) фильтрации заключается в предложении новых элементов для конкретного пользователя на основе предыдущих предпочтений пользователя и (или) мнениях других «единомышленников» пользователя. Преимуществом коллаборативной фильтрации является то, что при ее использовании вместо опроса нескольких «единомышленников» о нескольких объектах, система автоматически и анонимно учитывает мнения многих (тысяч) клиентов в отношении многих (тысяч) объектов.

Выделяют две категории алгоритмов коллаборативной фильтрации [2]: алгоритмы, основанные на анализе имеющихся оценок и алгоритмы, основанные на анализе модели данных.

Методы, основанные на анализе модели данных, предполагают, что сначала по совокупности оценок формируются описательная модель предпочтений пользователей, объектов и взаимосвязи между ними, а затем вырабатываются рекомендации на основании полученной модели. Такие алгоритмы могут быть основаны на вероятностном подходе [3], кластерном анализе, а также на анализе скрытых факторов [4].



В системе реализованы модели, анализирующие скрытые факторы. Они используются для целостного объяснения наблюдаемых оценок пользователей и формируются на основе факторизации матриц рейтингов (известные также как модели на основе SVD – сингулярного разложения матриц).

Для любой вещественной $n \times m$ матрицы A существуют две вещественные ортогональные $n \times m$ матрицы U и V такие, что: $A = U\Sigma V^T$.

Можно выбрать U и V так, чтобы диагональные элементы Σ имели вид: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \geq \lambda_{r+1} = \dots = \lambda_n = 0$, где r – ранг матрицы A .

В контексте рекомендательных систем используется усеченное разложение матриц, когда из элементов матрицы Σ остаются только первые d , а остальные мы полагаем равными нулю. Это равносильно тому, что у матриц U и V мы оставляем только первые d столбцов, а матрицу Σ обрезаем до квадратной $d \times d$: $A' = U\Sigma'V'^T$.

Полученная матрица A' хорошо приближает исходную матрицу A и, более того, является наилучшим низкоранговым приближением с точки зрения среднеквадратичного отклонения.

Это усечение одновременно достигает двух целей. Во-первых, оно уменьшает размерность векторного пространства, снижает требования хранения и вычислительные требования к модели. Во-вторых, отбрасывая малые сингулярные числа, малые искажения в результате шума в данных удаляются, оставляя только самые сильные эффекты и тенденции в этой модели. Снижение воздействия шума улучшает способность предоставлять высококачественные рекомендации [2].

Для тестирования работы рекомендательного алгоритма на основе коллаборативной фильтрации был использован тестовый набор данных MovieLens. Эксперименты включали в себя сравнение точности работы пользователе- и объектно-базированных алгоритмов в зависимости от выбранной метрики подобия, а также сравнение модельных алгоритмов с различным подходом к факторизации матрицы рейтингов.

Исходя из результатов экспериментов, были сделаны выводы о том, какие из алгоритмов наиболее точны и эффективны для имеющегося набора данных. Была разработана программная реализация прототипа рекомендательной системы кинофильмов на основе алгоритма сингулярного разложения матриц с факторизацией матриц на основе вероятностного градиентного спуска.

1. Recommender Systems Handbook [Текст] / F. Ricci, L. Rokach, B. Shapira, P. B. Kantor. – New York : Springer, 2001. – 845 с.

2. Пятикоп, Е. Е. Использование сингулярного разложения матриц в коллаборативной фильтрации [Текст] / Е. Е. Пятикоп // Проблеми інформатизації та управління. – 2013. – Т. 4, №44. – С. 76–81.

3. Гомзин, А. Г. Системы рекомендаций: обзор современных подходов [Текст] / А. Г. Гомзин, А. В. Коршунов // Труды ИСП РАН. – 2012. – Т. 22, №3. – С. 401–417.

4. Воронцов, К.В. Анализ клиентских сред: выявление скрытых профилей и оценивание сходства клиентов и ресурсов / К.В. Воронцов, В.А. Лексин // Мат. методы распознавания образов: 13-я Всерос. конф. Тезисы докл. – 2007. – С. 488–491.