

способности (ability) системы к персонализации без вырождения в узкий набор популярных позиций.

Отдельное внимание уделяется проектированию механизмов для граничных условий. Для новых пользователей с отсутствующей историей взаимодействий башня пользователя инициализируется на основе контекста и минимальных регистрационных данных, а в выдачу вносится управляемая доля популярных или географически близких событий. Для новых активностей их начальное векторное представление формируется исключительно на основе контентных признаков, что позволяет включать их в рекомендации до накопления поведенческой статистики.

Таким образом, процесс проектирования представляет собой итеративную оптимизацию взаимосвязанных компонентов – от корректного формирования признаков описаний и выбора архитектуры нейросетевых энкодеров до построения отказоустойчивой и масштабируемой инфраструктуры, способной обеспечить работу системы в условиях динамично изменяющегося каталога и разнородного контекста пользовательских запросов.

Библиографический список

1. Петров С.А., И. Ю Филатов // Исследование применения методов искусственного интеллекта в области автоматического планирования свободного времени // Межвузовский сборник научных трудов «Математическое и программное обеспечение вычислительных систем». 2025.
2. Петров С.А., И. Ю Филатов // Разработка системы нечеткого вывода в области автоматического планирования свободного времени // Межвузовский сборник научных трудов «Математическое и программное обеспечение вычислительных систем». 2025.
3. Бишоп Кристофер М., Распознавание образов и машинное обучение // – Издательство "Вильямс", 2020. 960 с. – ISBN 978-5-907144-55-2
4. Сивандам С.Н., Введение в генетические алгоритмы // – Издательство "Springer", 2017. 420 с

УДК 004.85; ГРНТИ 28.23

USER-BASED И ITEM-BASED ПОДХОДЫ КОЛЛАБОРАТИВНОЙ ФИЛЬТРАЦИИ В ЗАДАЧАХ АНАЛИЗА И УПРАВЛЕНИЯ АССОРТИМЕНТОМ

М.Д. Мацуганова, В.Е. Казаков

Витебский государственный технологический университет,
Республика Беларусь, Витебск, macuganovamaria5@gmail.com, wadimchik@mail.ru

Аннотация. В работе рассматриваются методы коллаборативной фильтрации – User-based и Item-based подходы, используемые для построения рекомендательных систем в электронной коммерции. Анализируется их применение не только для персонализации предложений, но и для решения обратной задачи – анализа и оптимизации товарного ассортимента на основе данных о поведении пользователей.

Ключевые слова: коллаборативная фильтрация, рекомендательные системы, управление ассортиментом, машинное обучение, анализ товарных взаимосвязей

USER-BASED AND ITEM-BASED APPROACHES TO COLLABORATIVE FILTERING IN THE PROBLEMS OF ANALYSIS AND ASSORTMENT MANAGEMENT

M.D. Matsuganova, V.E. Kazakov

Vitebsk State Technological University,
Republic of Belarus, Vitebsk, macuganovamaria5@gmail.com, wadimchik@mail.ru

The summary. This article examines collaborative filtering methods – user-based and item-based approaches – used to build recommender systems in e-commerce. It analyzes their application not only to personalizing offers but also to solving the inverse problem of analyzing and optimizing product assortment based on user behavior data.

Keywords: collaborative filtering, recommender systems, inventory management, machine learning, product relationship analysis

С развитием онлайн-торговли объем данных о взаимодействиях пользователей с товарами (просмотры, покупки, добавления в избранное) растет экспоненциально. Эти данные содержат скрытые закономерности о предпочтениях и товарных взаимосвязях, которые невозможно выявить вручную [1]. Именно здесь на первый план выходят методы машинного обучения, способные обрабатывать большие массивы информации и автоматически выявлять сложные паттерны [2].

Благодаря алгоритмам коллаборативной фильтрации компании могут не только рекомендовать товары («клиентам, купившим это, также нравится...»), но и эффективно управлять ассортиментом: выявлять товары-заменители (субституты) и товары-комплементы (дополняющие друг друга), оптимизировать выкладку на виртуальных полках и планировать закупки. Таким образом, переход от простой рекомендации к глубокому анализу товарной матрицы становится ключевым фактором эффективности ритейла.

Рассмотрим методы коллаборативной фильтрации подробнее. Основная идея коллаборативной фильтрации заключается в использовании знаний о предпочтениях группы пользователей для формирования предсказаний в отношении конкретного пользователя или товара. Существует два классических подхода: User-based и Item-based Collaborative Filtering (CF) [3].

User-based Collaborative Filtering (схожесть пользователей) – это исторически первый подход. Его суть заключается в поиске соседей для целевого пользователя – других покупателей со схожими предпочтениями или историей покупок. Рекомендация формируется на основе того, что понравилось «соседям», но еще неизвестно текущему пользователю. Для поиска соседей используются метрики сходства, такие как косинусная мера или коэффициент корреляции Пирсона [4]. Пример концептуальной реализации поиска схожих пользователей на Python представлен на рисунке 1.

```
# Функция для поиска N наиболее похожих пользователей
def find_similar_users(user_id, user_item_matrix, n=5):
    # Расчет косинусного сходства между всеми пользователями
    user_similarity = cosine_similarity(user_item_matrix)
    # Получение сходств для конкретного пользователя
    similar_users = list(enumerate(user_similarity[user_id]))
    # Сортировка по убыванию сходства и исключение самого пользователя
    sorted_users = sorted(similar_users, key=lambda x: x[1], reverse=True)[1:n+1]
    return sorted_users
```

Рис. 1. Концептуальный пример поиска похожих пользователей (User-based CF)*

Главным недостатком User-based подхода является его масштабируемость. Матрица сходства пользователей требует вычисления попарных сравнений для миллионов пользователей, что очень затратно. Кроме того, этот подход плохо объясняет, почему товар был рекомендован, и склонен рекомендовать слишком очевидные или, наоборот, случайные товары из-за "шума" во вкусах соседей.

Item-based Collaborative Filtering (схожесть товаров) было предложено как решение проблем масштабируемости и качества. Вместо поиска похожих людей, алгоритм ищет похожие товары. Основная предпосылка заключается в том, что пользователю, скорее всего, понравятся товары, похожие на те, которые ему уже нравились или которые он купил ранее. Сходство между товарами вычисляется на основе оценок всех пользователей: два товара считаются похожими, если их обычно оценивают или покупают одни и те же люди [5]. Этот подход менее динамичен, так как товарная база меняется медленнее, чем предпочтения пользователей, что позволяет пересчитывать матрицу сходства товаров реже (например, раз в день).

Для эффективного внедрения любого из подходов необходим достоверный массив данных о взаимодействиях. Наиболее распространенным источником информации является матрица «пользователь-товар» (user-item matrix), которая строится на основе транзакционных данных или логов просмотров. Ключевыми параметрами для анализа являются: user_id, item_id, rating (или purchase_flag). Эти данные позволяют количественно оценить связь между сущностями.

На основе исходных транзакционных данных была построена бинарная матрица «пользователь-товар» (user-item matrix), где строки соответствуют уникальным идентификаторам пользователей, столбцы – товарным позициям, а значения (0 или 1) отражают факт приобретения товара конкретным пользователем. Такая матрица является фундаментом для любого вида коллаборативной фильтрации, так как позволяет перейти от списка транзакций к математическому представлению взаимодействий в векторном пространстве.

Ключевым этапом анализа является расчет сходства между векторными представлениями товаров. Для этих целей была применена метрика косинусной близости (cosine similarity), которая измеряет косинус угла между векторами товаров в многомерном пространстве пользователей. Косинусная мера выбрана по следующим причинам: она устойчива к разреженности данных, учитывает только ненулевые значения и эффективно работает с бинарными данными. Значения косинусной близости варьируются от 0 до 1, где 0 означает полное отсутствие общих покупателей, а 1 – идентичные паттерны покупок [6].

Реализация алгоритма на языке Python с использованием библиотек pandas и scikit-learn представлена на рисунке 2.

```
# Функция для рекомендации товаров, похожих на заданный
def recommend_items(item_id, item_similarity_df, n=5):
    similar_scores =
item_similarity_df[item_id].sort_values(ascending=False)
    # Исключаем сам товар и берем топ-N
    top_items = similar_scores.iloc[1:n+1]
    return top_items
```

Рис. 2. Пример реализации расчета сходства товаров

Матрица косинусного сходства между товарами представлена на рисунке 3.

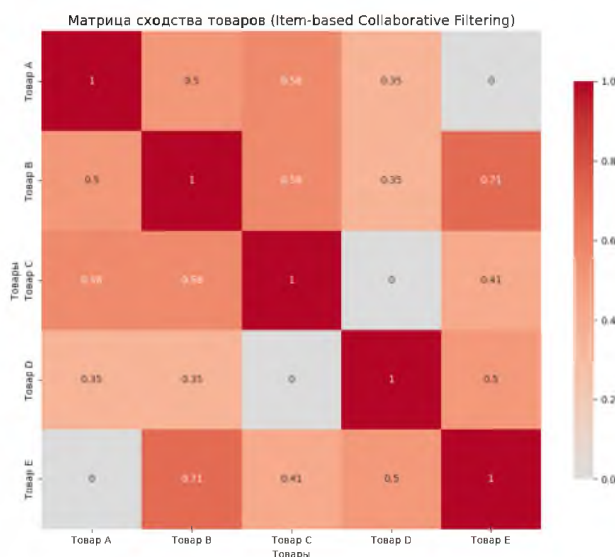


Рис. 3. Визуализация матрицы косинусного сходства между товарами

Таким образом, можно сделать вывод, что Item-based алгоритм автоматически выявляет смысловые и поведенческие связи между товарами, формируя кластеры взаимозаменяемых или взаимодополняющих продуктов.

Полученные в результате работы Item-based CF матрицы сходства открывают широкие возможности для управления ассортиментом, выходящие за рамки простых рекомендаций:

1. *Выявление товаров-комплементов.* Алгоритм показывает, какие товары часто покупают в паре (например, ноутбук и сумка для ноутбука). Это позволяет создавать комплекты («Купи вместе»), увеличивая средний чек;

2. *Анализ товаров-субституттов.* Товары с очень высоким коэффициентом сходства (близким к 1) являются прямыми конкурентами друг для друга в глазах покупателя. Если один из таких товаров продается плохо, а другой хорошо, это сигнал к пересмотру ценообразования, упаковки или целесообразности присутствия обоих товаров в линейке;

3. *Оптимизация каталога.* Анализ разреженных связей позволяет выявить товары, которые не имеют сильных корреляций с другими. Это могут быть нишевые или непопулярные позиции («изгои»), требующие ребрендинга или вывода из ассортимента;

4. *Управление запасами.* Зная, что товар X является сильным комплементом для товара Y, можно прогнозировать рост спроса на X при планировании маркетинговой акции на Y, предотвращая дефицит.

В отличие от User-based подхода, который лучше подходит для сценариев «холодного старта» пользователя (когда о новом пользователе мало данных) и таргетинга в реальном времени, Item-based подход обеспечивает стабильную и интерпретируемую модель товарных взаимосвязей, что делает его незаменимым инструментом для аналитиков и категорийных менеджеров.

Таким образом, User-based и Item-based алгоритмы являются мощным и взаимодополняющим инструментарием. User-based CF позволяет гибко реагировать на сиюминутные предпочтения аудитории, повышая вовлеченность. Item-based CF, в свою очередь, обеспечивает аналитическую базу для стратегического управления ассортиментом: он помогает структурировать товарную матрицу, выявлять скрытые продуктовые кластеры и зависимости, оптимизировать закупки и ценообразование. Используя оба подхода в комбинации, компании могут достичь синергетического эффекта, повышая как лояльность клиентов, так и операционную эффективность бизнеса.

Библиографический список

1. Future of E-commerce: Personalization Statistics [Electronic resource]. – Mode of access: <https://www.example.com/ecommerce-personalization-stats>. – Date of access: 10.01.2026.
2. Гудфеллоу, Я. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль. – М.: ДМК Пресс, 2018. – 652 с.
3. Aggarwal, C. C. Recommender Systems: The Textbook / C. C. Aggarwal. – Springer, 2016. – 498 p.
4. Leskovec, J. Mining of Massive Datasets / J. Leskovec, A. Rajaraman, J. D. Ullman. – Cambridge University Press, 2014. – 513 p.
5. Amazon.com Recommendations: Item-to-Item Collaborative Filtering [Electronic resource] / G. Linden, B. Smith, J. York // IEEE Internet Computing. – Mode of access: <https://example.com/amazon-cf-paper>. – Date of access: 12.01.2026.
6. Системы рекомендаций: коллаборативная фильтрация [Электронный ресурс] // Хабр. – Режим доступа: <https://habr.com/ru/articles/> – Дата доступа: 15.01.2026.