

РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА С КОНТЕНТНОЙ ФИЛЬТРАЦИЕЙ В ЭЛЕКТРОННОЙ КОММЕРЦИИ

Волин М.В.¹, Вовк Л.П.²

¹Волин Михаил Владимирович – магистрант
по направлению «Информационные системы и технологии»,
факультета «Транспортные и информационные технологии»;
²Вовк Леонид Петрович – доктор технических наук, профессор,
кафедра «Математическое моделирование»,
Автомобильно-дорожный институт (филиал) федерального
государственного бюджетного образовательного учреждения высшего
образования «Донецкий национальный технический университет»,
г. Горловка

Аннотация: в статье рассматривается подход к созданию рекомендательной системы, основанной на контентной фильтрации. Описываются основные этапы построения профиля пользователя и оценки сходства объектов на основе их признаков. Система предлагается применять в электронной коммерции, где она обеспечивает персонализацию предложений товаров. Для полноты исследования проводится сравнение с альтернативным подходом – моделью коллаборативной фильтрации, широко используемой в e-commerce. Приведен новый синтетический пример онлайн-магазина продуктов питания, иллюстрирующий работу системы рекомендаций. Описан практический эксперимент при помощи Python с использованием библиотеки pandas и sklearn. В заключение обсуждаются преимущества и недостатки обоих методов, а также перспективы их совмещения.

Ключевые слова: рекомендательная система; контентная фильтрация; коллаборативная фильтрация; электронная коммерция; продукты питания; Python.

RECOMMENDATION SYSTEM WITH CONTENT FILTERING IN E-COMMERCE

Volin M.V.¹, Vovk L.P.²

¹Volin Michael Vladimirovich – Master's student
SPECIALTY INFORMATION SYSTEMS AND TECHNOLOGIES,
FACULTY OF TRANSPORT AND INFORMATION TECHNOLOGIES;
²Vovk Leonid Petrovich – Doctor of Technical Sciences, Professor,
DEPARTMENT OF MATHEMATICAL MODELING,
AUTOMOBILE AND ROAD INSTITUTE (BRANCH) OF THE FEDERAL STATE BUDGETARY EDUCATIONAL
INSTITUTION OF HIGHER EDUCATION
DONETSK NATIONAL TECHNICAL UNIVERSITY,
GORLOVKA

Abstract: the article discusses an approach to creating a recommender system based on content filtering. The main stages of building a user profile and assessing the similarity of objects based on their attributes are described. The system is proposed for use in e-commerce, where it provides personalization of product offers. For the sake of completeness, a comparison is made with an alternative approach - the collaborative filtering model, widely used in e-commerce. A new synthetic example of an online grocery store is given, illustrating the operation of the recommender system. A practical experiment is described using Python with the pandas library and sklearn. In conclusion, the advantages and disadvantages of both methods are discussed, as well as the prospects for combining them.

Keywords: recommender system; content filtering; collaborative filtering; e-commerce; food; Python.

УДК 004.738.5

Введение.

С развитием интернет-торговли и огромного выбора товаров персональные рекомендации становятся критически важными для повышения удовлетворенности пользователей и роста продаж. Рекомендательная система – это программный комплекс, анализирующий явные и неявные параметры интересов пользователя и формирующий на их основе списки персонализированных товаров [1, с. 7; 2, с. 64]. Существует два классических подхода к построению таких систем: контентная фильтрация и коллаборативная фильтрация [1, с. 7]. Контентная фильтрация (content-based filtering) опирается на характеристики самих товаров: система рекомендует пользователю те предметы, атрибуты которых максимально близки тем, которые уже понравились этому пользователю. В противоположность этому, коллаборативная фильтрация (collaborative filtering) оперирует историями покупок или оценками

множества пользователей, выявляя схожие предпочтения и предлагая товары, которые понравились «похожим» людям [3, с. 13].

Оба подхода имеют как преимущества, так и недостатки [2, с. 70]. Контентная фильтрация позволяет рекомендовать новые или редко покупаемые товары (решая проблему «холодного старта» для товара), но может «зацикливаться» на узком круге интересов пользователя. Коллаборативная фильтрация хорошо улавливает сложные межпредметные связи и объединяет людей по схожим вкусам, однако требует больших объемов данных и сталкивается с проблемой недостатка рейтингов новых пользователей. Многие современные системы в электронной коммерции используют гибридные модели, объединяя оба подхода [5, с. 176]. Например, популярные сервисы (Netflix, Amazon) формируют рекомендации одновременно на основе коллаборативных и контентных методов, используя демографические и другие данные для повышения качества предсказаний.

В данной работе подробно рассматривается алгоритм контентной фильтрации и его применение к задаче рекомендаций в интернет-магазине продуктов. Цель – рассмотреть сущность контентной фильтрации, приведя сравнение с коллаборативной фильтрацией и пример в области онлайн-продаж пищевых товаров.

Методы.

Контентная фильтрация предполагает, что каждый товар описывается вектором признаков (категория, описание, бренд, питательные свойства и т.п.), а профиль пользователя строится как совокупность векторов уже потребленных им товаров.

Алгоритм включает следующие основные этапы:

- Построение векторного описания товаров. На основе имеющихся метаданных (название, категория, свойства) или текстового описания каждого продукта формируется вектор признаков. Часто используются методы TF-IDF или векторные представления эмбедингов (word2vec, FastText и пр.) [1, с. 9].

- Формирование профиля пользователя. Профиль представляется суммарным или усредненным вектором тех товаров, которые пользователь уже приобрел или высоко оценил. Например, если пользователь купил обезжиренное молоко и йогурт, то профиль пользователя будет акцентировать признаки «молочные продукты» и «низкая жирность».

- Оценка сходства. Для каждого товара, не приобретенного пользователем, вычисляется мера сходства его векторного представления с вектором профиля пользователя. Обычно используется косинусная близость или другая метрика евклидова/косинусного расстояния.

- Генерация рекомендаций. На основе рассчитанных значений сходства выбираются топ-N товаров с наибольшим показателем. Эти товары представляются в качестве персональных рекомендаций.

В результате контентная модель рекомендует объекты на основе общих признаков с ранее покупаемыми товарами.

Например, если профиль пользователя сфокусирован на низкокалорийной молочной продукции, система предложит схожие продукты (сметана с низким процентом жирности, творожки и т.п.). Алгоритмически такой подход можно описать, например, через косинусную меру:

$$(p_i, x_i) = \frac{p_i \cdot x_i}{||p_i|| ||x_i||}, \quad (1)$$

где $p_i \cdot x_i$ – скалярное произведение профиля пользователя и вектора признаков товара;
 $||p_i|| ||x_i||$ – их нормы. Косинусное сходство измеряет угол между двумя векторами, что позволяет определить степень их схожести.

Для контентной модели характерны следующие особенности:

- Она автоматически учитывает семантику и категории товаров.
- Не требует информации о других пользователях, что хорошо при дефиците данных о поведении пользователей.

Недостаток – слабая способность предлагать что то сильно отличающееся от уже известных предпочтений, а также сложность создания качественных векторов признаков для каждого товара.

Для сравнения в рамках методов приведем ключевые особенности коллаборативной фильтрации. В традиционных системах коллаборативной фильтрации используется объединение оценок (или покупок) множества пользователей. Существует два направления: user-based и item-based. В первом схожие пользователи получают рекомендации по товарам друг друга, во втором – товары, которые часто покупаются совместно, рекомендуются тем, кто приобрел аналогичный набор товаров. Методы памяти (neighbor-based) сравнивают известных соседей (пользователей или товаров), а методы моделей (model-based) применяют машины обучения (например, матричную факторизацию или нейросети) [4, с. 16]. В целом при коллаборативной фильтрации товары рекомендуются «по поведению» группы людей.

Совмещение этих подходов позволяет компенсировать их недостатки. Например, при гибридном подходе система может считать, как косинусное сходство профилей пользователя и товара (контентный метод), так и оценивать частоту совместных покупок (коллаборативный). Это улучшает качество рекомендаций, особенно когда разные методы обнаруживают различные «скрытые» связи между товарами.

Для практической демонстрации работы контентной фильтрации рассмотрим синтетическую задачу в контексте онлайн-торговли продуктами питания. В качестве данных используется таблица признаков восьми товаров, каждый из которых представлен бинарными характеристиками: принадлежность к категории (молочный продукт, фрукт, злаки и др.), содержание белков, жиров, углеводов и т.п. Вектор признаков строится в виде булевой матрицы, приведённой в таблице 1, на основании которой формируется профиль пользователя и рассчитывается сходство. Для иллюстрации методов контентной фильтрации будет использоваться пример на языке Python с использованием pandas и sklearn.

Таблица 1. Матрица признаков товаров.

Продукт	Молочный	Фрукт	Зерновой	Белковый	Жирный	Углеводный
Молоко	1	0	0	1	0	0
Хлеб	0	0	1	0	0	1
Яблоко	0	1	0	0	0	1
Сыр	1	0	0	1	0	0
Банан	0	1	0	0	0	1
Масло сливочное	1		0	0	0	1
Творог	1	0	0	1	0	0
Йогурт	1	0	0	1	0	0

Предположим, что пользователь положительно оценил товары «Молоко» и «Масло сливочное». На их основе строится его профиль путём усреднения соответствующих векторов. Далее производится вычисление косинусного сходства между профилем пользователя и признаками остальных товаров. Этот алгоритм реализуется на Python с помощью библиотек pandas и sklearn:

```

1  import pandas as pd
2  from sklearn.metrics.pairwise import cosine_similarity
3
4  # Матрица признаков товаров
5  df = pd.DataFrame([
6      [1,0,0,1,0,0], # Молоко
7      [0,0,1,0,0,1], # Хлеб
8      [0,1,0,0,0,1], # Яблоко
9      [1,0,0,1,0,0], # Сыр
10     [0,1,0,0,0,1], # Банан
11     [1,0,0,0,1,0], # Масло
12     [1,0,0,1,0,0], # Творог
13     [1,0,0,1,0,0], # Йогурт
14 ], index=[ 'Молоко', 'Хлеб', 'Яблоко', 'Сыр', 'Банан', 'Масло', 'Творог', 'Йогурт' ],
15     columns=[ 'Молочный', 'Фрукт', 'Зерновой', 'Белковый', 'Жирный', 'Углеводный' ])
16
17 # Формирование профиля пользователя
18 user_items = [ 'Молоко', 'Масло' ]
19 user_profile = df.loc[user_items].mean()
20
21 # Вычисление сходства и сортировка
22 similarities = cosine_similarity(df, [user_profile]).flatten()
23 recommendations = pd.Series(similarities, index=df.index).sort_values(ascending=False)

```

Рис. 1. Алгоритм фильтрации на языке Python

Приведя таблицу сходств остальных продуктов, приведём их косинусное сходство в таблице 2.

Таблица 2. Ранжирование товаров по степени сходства с профилем пользователя

Продукт	Косинусное сходство
Сыр	0.87
Йогурт	0.87
Творог	0.87
Хлеб	0.00
Яблоко	0.00
Банан	0.00

Как видно, товары, близкие по структуре (молочные и белковые продукты), получили высокие оценки, тогда как злаки и фрукты — низкие. Такой подход позволяет формировать прозрачные и интерпретируемые рекомендации в системах электронной торговли. В отличие от коллаборативной модели, в данном случае не требуется информация о других пользователях или их действиях, что особенно полезно при запуске нового магазина или при работе с ограниченными пользовательскими данными

Обсуждение.

Полученные результаты подчеркивают сильные и слабые стороны каждого метода. Контентная модель легко добавляет новые товары в каталог — достаточно их верно описать признаками. Например, поступление нового сорта йогурта сразу попадет в рекомендационный пул для тех пользователей, у которых в профилях есть молочные продукты. Эта гибкость важна для e-commerce, где ассортимент часто обновляется. Однако такая модель не способна рекомендовать товары из совершенно других категорий, даже если они могут заинтересовать пользователя [2, с. 74].

Напротив, коллаборативная фильтрация нередко обнаруживает неожиданные связи: например, если данные показывают, что «покупатели молока» часто покупают «грецкие орехи», система об этом узнает и будет предлагать орехи любителям молока. Но коллаборативные алгоритмы не работают с новыми товарами или с новыми пользователями без истории (так называемый «холодный старт»). Кроме того, при малом количестве пользователей качество рекомендаций может быть низким.

В прикладной практике крупные компании комбинируют методы. Как отмечают исследователи, Netflix строит рекомендации одновременно на коллаборативной фильтрации (по интересам пользователей) и на контентной (по характеристикам фильмов). Аналогично, Amazon анализирует совместные покупки и использует рейтинги товаров, что сочетает элементы обоих подходов. Современные системы электронной коммерции могут также учитывать демографию, время покупки, сезонность и другие факторы, расширяя традиционные модели.

Важно подчеркнуть, что выбор модели зависит от конкретных задач и данных. Если ключевым является быстрое развертывание системы с новыми товарами и ясная интерпретируемость — на первый план выходит контентная фильтрация. Если же доступны обширные данные о поведении пользователей, сильной стороной станет коллаборативный подход. Гибридные решения позволяют оптимизировать качество рекомендаций, сочетая сильные черты обоих методов

Вывод.

В работе описана рекомендательная система на основе контентной фильтрации, которая использует метаданные товаров и историю покупок пользователя для генерации персональных рекомендаций. Сохранена общая структура исходного исследования, но текст полностью переформулирован: уточнена постановка задачи, добавлены сравнительный анализ и практический пример. Показано, что при рекомендательной фильтрации интернет-магазина продуктов питания контентный метод предлагает пользователю товары с близкими характеристиками к уже выбранным им, тогда как коллаборативный метод выявляет «корзинные» связи среди покупателей. Выполненное сравнение подчеркивает, что оба подхода дополняют друг друга, что важно учитывать при разработке систем e-commerce. В дальнейшем планируется реализовать синтетическую модель и провести эксперименты с реальными данными для оценки эффективности описанного алгоритма.

Список литературы / References:

1. Тимофеев А.А. Механизмы работы рекомендательных систем и их применение // Материалы Всероссийской научной конференции «Социальный инженер – 2022». М., 2022. С. 7–13.
2. Соколов А.В., Сычев И.А., Соколова О.Л., Волкова Д.Б., Селетков И.П., Яшичев Д.Л., Ясницкий Л.Н. Теоретические и практические аспекты построения рекомендательных моделей: типология, архитектура и направления проектирования // Вестник Пермского ун-та. Математика. Механика. Информатика. 2024. № 3(66). С. 64–77.
3. Пименов В.И., Аникина Т.С., Пименов И.В. Алгоритмическое обеспечение рекомендательных систем в сфере электронной коммерции // Информационно-управляющие системы. 2023. № 1. С. 13–20.

4. *Смирнов А.В., Шилов Н.Г., Пономарёв А.В., Кашевник А.М., Парфенов В.Г.* Групповые контекстно-управляемые рекомендующие системы: архитектура и модели // Интеллектуальные системы и технологии. Интеллект и принятие решений. 2013. № 3. С. 14–25.
5. *Нефёдова Ю.С.* Архитектура гибридной рекомендательной системы GEFEST (Generation–Expansion–Filtering–Sorting–Truncation) // Системы и средства информатики. 2012. Т. 22. № 2. С. 176–196.