

УДК 519.87:332.1

Л. П. Вовк, д-р техн. наук, М. В. Волин

**Автомобильно-дорожный институт (филиал)
федерального государственного бюджетного образовательного учреждения
высшего образования «Донецкий национальный технический университет»
в г. Горловка**

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ПОСТРОЕНИЯ СИСТЕМЫ ПЕРСОНАЛЬНЫХ РЕКОМЕНДАЦИЙ НА ОСНОВЕ ПРИЗНАКОВ ОБЪЕКТОВ

В статье рассматривается модель рекомендательной системы в области торговых отношений, основанная на контентной фильтрации. Обоснована актуальность темы в условиях роста объемов данных, формулируются цель и задачи исследования. Представлен простой синтетический эксперимент, иллюстрирующий работу контентно-ориентированного рекомендателя. Описаны метрики оценки качества рекомендаций Precision@N и MAP, а также даны примеры их расчета. Полученные результаты и их обсуждение демонстрируют эффективность предложенного подхода и дают основу для дальнейших исследований.

Ключевые слова: рекомендательная система, контентная фильтрация, профиль пользователя, метрика Precision@N, средняя точность, синтетический эксперимент

Введение

На текущий момент рекомендательные системы (РС) (Recommender Systems, RecSys) являются одними из наиболее значимых и повсеместно используемых компонентов современных информационных систем. Их актуальность и влияние продолжают расти с экспоненциальным увеличением объемов данных и разнообразия контента и товаров, доступных пользователям.

Суть рекомендательных систем заключается в предоставлении персонализированных предложений пользователям, будь то товары, услуги, контент (фильмы, музыка, новости), информация или даже люди. Главная цель РС – помочь пользователям справляться с информационной перегрузкой, повысить их удовлетворенность, увеличить вовлеченность, конверсию (покупки, просмотры) и лояльность.

В современных информационных системах рекомендательные системы стали повседневным инструментом, упрощающим поиск релевантного контента для пользователей. Они позволяют сервисам анализировать интересы пользователей и предлагать им подходящие товары или услуги, что повышает пользовательскую удовлетворенность и экономический эффект [1]. Рекомендательные системы представляют собой сложные алгоритмы, созданные для формирования релевантных рекомендаций на основе анализа прошлых действий пользователей [2]. Таким образом, рекомендательные системы – это не просто дополнительная функция, а фундаментальная часть архитектуры современных информационных систем, которая лежит в основе их персонализации, эффективности и конкурентоспособности.

Основные подходы к построению рекомендаций включают колаборативную фильтрацию, контентную фильтрацию и гибридные методы. В отличие от колаборативной фильтрации, ориентированной на взаимоотношения «пользователь – пользователь» или «товар – товар», контентная фильтрация использует информацию о характеристиках товаров и профилях пользователей [3].

Контентная фильтрация (Content-Based Filtering) – это алгоритм рекомендаций, который основывается на характеристиках (атрибуатах) товаров и предпочтениях пользователей. Главная задача алгоритма контентной фильтрации – сопоставить интересы пользователя с атрибутами товаров, чтобы предложить те, которые лучше всего соответствуют его запросам и вкусам [1]. Контентная фильтрация – это один из основных подходов в РС, который форми-

рует рекомендации, основываясь на сходстве характеристик (содержания) самих объектов и предпочтений пользователя, выраженных через его взаимодействие с этими характеристиками.

Для каждого объекта (статьи, товара) создается подробный «профиль», который описывает его ключевые характеристики (атрибуты, метаданные). Для статьи – это ключевые слова, рубрики, авторы, темы, длина. Для товара – категория, бренд, цвет, размер, функции, технические характеристики. Эти характеристики могут быть представлены в виде векторов признаков. Может быть использован, например, TF-IDF для текста (TF-IDF – это числовая статистика, отражающая важность слова для документа в коллекции или корпусе. Это широко используемый метод в поиске информации и машинном обучении для взвешивания терминов на основе их частоты в документе и их редкости во всей коллекции). Также возможно применение однократного кодирования (one-hot encoding) для категориальных признаков. Однократное кодирование – это метод, используемый в машинном обучении для представления категориальных данных в виде числовых данных. Он преобразует каждую категорию в двоичный вектор, где только один элемент является «горячим» (1), а остальные – «холодными» (0), указывая на наличие или отсутствие этой конкретной категории.

Для каждого пользователя создается «профиль предпочтений», который строится на основе его предыдущих взаимодействий с объектами (просмотры, покупки, лайки, оценки). Этот профиль отражает, какие характеристики объектов пользователь предпочитает. Профиль пользователя также может быть представлен в виде вектора, часто как усредненный или агрегированный вектор характеристик объектов, с которыми пользователь взаимодействовал положительно.

Когда нужно сделать рекомендацию, система сравнивает профиль пользователя с профилями объектов, которые пользователь еще не видел (или не взаимодействовал с ними). Объекты, чьи профили максимально похожи на профиль пользователя, считаются наиболее релевантными и рекомендуются ему. Для измерения сходства часто используются метрики, такие как косинусное сходство (cosine similarity) между векторами признаков.

Для контентной фильтрации не нужны данные о других пользователях. Рекомендации генерируются на основе индивидуальной истории пользователя и свойств объектов. Это позволяет рекомендовать новые объекты (товары, статьи), у которых нет истории взаимодействий от других пользователей. Достаточно, чтобы у нового объекта был сформирован его профиль.

При правильной реализации контентная фильтрация может предлагать более разнообразные рекомендации, если профиль пользователя не слишком узкий, и система способна обнаруживать более тонкие связи между признаками.

Фильтрация контента также широко используется в Интернете для ограничения и исключения из доступа веб-страниц в Интернете или сообщений из электронной почты, которые считаются нежелательными. Он также известен как фильтрация информации. Метод работает путем указания строк символов, которые, если совпадают, указывают на нежелательный контент, который должен быть отсечен. Фильтрация веб-контента – это брандмауэр, блокирующий доступ к определенным сайтам. Фильтрация контента и продукты, которые предлагают эту услугу, можно разделить на веб-фильтрацию – проверку веб-сайтов или страниц, и фильтрацию электронной почты – проверку электронной почты на спам или другой нежелательный контент. В статье [4] представлен полный обзор основных типов, задач, инструментов, процесса и алгоритма, задействованного в фильтрации веб-контента, а также предложена новая методология для проверки текстового контента на веб-страницах и принятия алгоритма решения о том, разрешен или запрещен доступ к веб-странице.

В работах [5, 6] представлено подробное описание фильтрации веб-контента товаров и услуг, ее методов, типов, задач, процесса, пяти типов инструментов и их использования, а также кратко обсуждается алгоритм раннего принятия решения для ускорения фильтрации предлагаемого товара.

Анализ научных публикаций позволяет утверждать, что контентная фильтрация – это мощный подход в рекомендательных системах, но он, как и любой другой метод, имеет свои

существенные недостатки, которые ограничивают его эффективность в различных сценариях.

Самый значительный недостаток предлагаемых моделей – проблема переспециализации (Over-specialization / Filter Bubble). Поскольку контентная фильтрация рекомендует объекты, максимально похожие на те, которые пользователь уже оценил или с которыми взаимодействовал, система имеет тенденцию к чрезмерной специализации. Она постоянно предлагает пользователю похожий контент, не выходя за рамки его «известных» интересов. В результате пользователь попадает в «пузырь фильтров», где ему показывают только то, что соответствует его текущему профилю. Он не может открыть для себя что-то новое, неожиданное, что, возможно, ему тоже понравилось бы, но имеет иные характеристики.

Следует сказать, что эффективность контентной фильтрации напрямую зависит от того, насколько полно, точно и подробно описаны характеристики каждого объекта. Если атрибуты скучны, неполны, неточно извлечены или не способны уловить все нюансы вкуса пользователя, то рекомендации будут низкого качества. В результате модель может упустить важные, но неочевидные факторы, которые влияют на предпочтения пользователя.

В публикациях не проанализирована проблема «холодного старта» для новых пользователей (Cold Start for New Users). Ее суть состоит в том, что для построения точного профиля предпочтений пользователя, системе необходимо достаточное количество его предыдущих взаимодействий (просмотров, оценок, покупок). Если пользователь новый и у него нет истории взаимодействия, системе не на чем основываться рекомендации. Таким образом, новый пользователь не получает персонализированных рекомендаций или получает очень общие, нерелевантные рекомендации (например, по популярности), пока не накопит достаточную активность. Это может привести к его быстрому оттоку.

Из-за этих недостатков контентная фильтрация редко используется в чистом виде в крупных современных рекомендательных системах. Вместо этого ее часто комбинируют с другими подходами, в первую очередь с коллаборативной фильтрацией, создавая гибридные рекомендательные системы. Это позволяет компенсировать слабости одного метода сильными сторонами другого.

Цель настоящего исследования – разработать и проанализировать смешанную модель рекомендательной системы в области торговых отношений на основе контентной фильтрации и провести ее оценку с помощью стандартных метрик качества.

Задачи исследования включают:

- анализ принципов построения контентно-ориентированных рекомендателей и существующих подходов к формированию профилей пользователей;
- разработку методики синтетического эксперимента для оценки модели: генерацию искусственных данных о товарах и предпочтениях пользователей;
- реализацию алгоритма контентной фильтрации на основе вычисления сходства между профилем пользователя и векторами признаков товаров;
- оценку качества рекомендаций с использованием метрик Precision@N и MAP и анализ полученных результатов.

Основной материал исследования

В математической модели контентной фильтрации рассмотрим более детально этапы формирования профиля пользователя, определения схожести и построения рекомендаций. Также уделим внимание каждому этапу модели и инструментам, используемым для оценки качества рекомендаций.

1. Определение множества признаков и векторного представления товаров.

В контентной фильтрации каждый товар описывается набором характеристик (атрибутов), которые могут быть:

- бинарными признаками, указывающими на наличие или отсутствие определенной характеристики, например: «новинка» (1 – да, 0 – нет);

- категориальными признаками, например, брендами, которые часто кодируют с использованием метода one-hot encoding;
- числовыми признаками, такими как цена, рейтинг, объем.

Обозначим товар i_j в виде вектора признаков:

$$x_j = [x_{j1}, x_{j2}, \dots, x_{jk}], \quad (1)$$

где k – число признаков (атрибутов).

2. Формирование профиля пользователя на основе взаимодействий с товарами.

Контентная фильтрация строит профиль предпочтений пользователя p_i на основе товаров, с которыми пользователь u_i взаимодействовал. Взаимодействие может включать просмотр товаров, добавление в корзину, покупки и т. д. [1, 4].

1) Средневзвешенное значение признаков.

Для формирования профиля p_i мы можем усреднить признаки всех товаров, с которыми взаимодействовал пользователь. Обозначим множество таких товаров как I_{u_i} :

$$p_i = \frac{1}{|I_{u_i}|} \sum_{i_j \in I_{u_i}} x_i, \quad (2)$$

где x_i – вектор признаков товара i_j ;

$|I_{u_i}|$ – число товаров, с которыми взаимодействовал пользователь.

2) Взвешенное усреднение на основе уровня взаимодействия.

Если для разных типов взаимодействий (например, «просмотр» или «покупка») мы хотим назначить разный вес, профиль формируется как взвешенное среднее:

$$p_i = \frac{1}{\sum_{l=1}^L w_l |I_{u_i}^l|} \sum_{l=1}^L \sum_{i_j \in I_{u_i}^l} w_l x_i, \quad (3)$$

где L – количество типов взаимодействий;

$I_{u_i}^l$ – подмножество товаров, с которыми у пользователя был тип взаимодействия l ;

w_l – вес данного типа взаимодействия.

3. Применение метода вычисления сходства.

Для рекомендации товаров пользователю необходимо найти меру сходства между профилем пользователя p_i и характеристиками товаров x_i [7, 8].

В контексте рекомендательных систем на основе контентной фильтрации мера сходства между профилем пользователя и характеристиками товаров является ключевым элементом. Она позволяет определить, насколько новый (еще не просмотренный или не оцененный) товар соответствует интересам пользователя, основываясь на его предыдущих предпочтениях. Как профиль пользователя, так и характеристики товаров обычно представляются в виде векторов признаков (feature vectors). Эти векторы могут содержать различные типы данных. Для товара – это, например, категория, бренд, цвет, функции. Вся эта информация может быть закодирована в числовой вектор. Для пользователя это может быть агрегированный вектор из характеристик товаров, которые он ранее положительно оценил, или явные предпочтения, заданные пользователем.

Для измерения сходства между этими векторами используются различные математические метрики [9].

Выбор конкретной меры сходства зависит от типа данных, используемых для описания профилей, и от специфики задачи. В современных системах часто применяют эмбеддинги

(векторные представления, полученные с помощью нейронных сетей) для пользователей и товаров, а затем используют косинусное сходство для сравнения этих эмбеддингов.

Наиболее часто используемые меры сходства включают:

1) Косинусное сходство:

$$\text{sim}(p_i, x_i) = \frac{p_i \cdot x_i}{\|p_i\| \|x_i\|}, \quad (4)$$

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. Значение косинуса варьируется от -1 (абсолютно противоположные векторы) до 1 (идентичные векторы). Значение 0 означает отсутствие линейной зависимости. Мера очень популярна в РС, особенно для текстовых данных (например, когда профили пользователя и товаров представлены как векторы TF-IDF или эмбеддинги). Она хорошо работает, когда важна направленность векторов, а не их абсолютная величина. Нечувствительна к различиям в масштабе векторов. Данная мера очень эффективна для разреженных данных.

2) Евклидово расстояние (Euclidean Distance):

$$\text{sim}(p_i, x_i) = -\|p_i - x_i\| = -\sqrt{\sum_{k=1}^K (p_{ik} - x_{jk})^2}, \quad (5)$$

где $\|p_i - x_i\|$ – евклидово расстояние между вектором пользователя и вектором товара.

Суть этой меры состоит в измерении «прямого» расстояния между двумя точками (векторами) в многомерном пространстве. Меньшее расстояние соответствует большей схожести, поэтому знак может быть изменен для интерпретации результата как меры близости. Лучше всего подходит, когда абсолютные значения признаков имеют значение. После вычисления расстояния, его часто инвертируют [например, $1/(1+\text{distance})$] или используют функцию затухания, чтобы получить меру сходства, где большее значение означает большее сходство. Недостаток этой меры состоит в чувствительности к масштабу признаков и к «проклятию размерности» (curse of dimensionality) в высокоразмерных пространствах.

3) Манхэттенское расстояние (Manhattan Distance / City Block Distance). Также называется L_1 -нормой. Измеряет сумму абсолютных разностей между координатами векторов. Представляет собой расстояние, которое пришлось бы пройти по сетке (как по улицам города). Как и Евклидово расстояние, меньшее значение означает большее сходство. Но оно менее чувствительно к выбросам, чем Евклидово. Манхэттенское расстояние используется, если мы считаем, что каждая характеристика вносит равный вклад:

$$\text{sim}(p_i, x_i) = -\sum_{k=1}^K |p_{ik} - x_{jk}|. \quad (6)$$

4. Построение рекомендаций на основе сходства.

Для каждого пользователя u_i :

- вычисляем $\text{sim}(p_i, x_j)$ для всех товаров $i_j \in I$;
- сортируем товары по значению $\text{sim}(p_i, x_j)$ в порядке убывания;
- рекомендуем пользователю u_i товары с наибольшими значениями $\text{sim}(p_i, x_j)$, выбрав топ- N из них.

Приведем краткий пример. Дано: товары описаны 3 признаками: [Категория: электроника (1/0), Цена (в тыс. руб.), Рейтинг (из 5)], пользователь взаимодействовал с двумя товарами (таблица 1).

Таблица 1 – Данные о товарах пользователя

Товар ID	Электроник	Цена	Рейтинг	Взаимодействие (вес)
A	1	30	4,5	Покупка (вес 2)
B	0	15	3,0	Просмотр (вес 1)

Формирование профиля пользователя на основе представленных данных таблицы 1:

$$\vec{p} = \frac{2 \cdot [1; 30; 4,5] + 1 \cdot [0; 15; 3,0]}{2+1} = [2; 60; 9] + \frac{[0; 15; 3]}{3} = \frac{[2; 75; 12]}{3} = [0,67; 25; 4].$$

Добавим новый товар для рекомендации (таблица 2).

Таблица 2 – Новый рекомендованный товар С

Товар ID	Электроника	Цена	Рейтинг
C	1	20	4,5

Тогда вектором признаков товара С исходя из данных таблицы 2 будет $\vec{x}_C = [0,67; 25; 4]$.

Вычислим косинусное сходство пользователя и нового товара:

$$\begin{aligned} sim(\vec{p}, \vec{x}_C) &= \frac{(0,67 \cdot 1 + 25 \cdot 20 + 4 \cdot 4,5)}{\|\vec{p}\| \cdot \|\vec{x}_C\|} = \frac{0,67 + 500 + 18}{\|\vec{p}\| \cdot \|\vec{x}_C\|} = \\ &= \frac{518,67}{\sqrt{0,67^2 + 25^2 + 4^2} \cdot \sqrt{1^2 + 20^2 + 4,5^2}} = \frac{518,67}{\sqrt{642,45} \cdot \sqrt{421,25}} \approx 0,998. \end{aligned}$$

Как видно, косинусное сходство между вектором профиля пользователя и вектором товара С очень близко к 1, что говорит о высокой степени соответствия. Это означает, что товар С релевантен для пользователя и может быть уверенно рекомендован.

5. Оценка качества модели рекомендаций.

Метрики точности являются полезными метриками для оценки общей производительности машинного обучения, они представляют собой долю правильно классифицированных экземпляров в наборе данных. Метрики оценки в сочетании с такими, как точность и полнота, используются для получения более полного понимания производительности модели. В общем случае точность и полнота сравнивают предсказанный класс с фактическим классом тестового набора и вычисляют отношение правильных предсказаний к общему числу сделанных предсказаний.

Чтобы понять, насколько хорошо работает система рекомендаций, применяем следующие метрики:

а) Precision@N и Recall@N.

Precision@N показывает долю релевантных товаров среди рекомендованных топ- N товаров:

$$\text{Precision @ N} = \frac{\text{количество релевантных товаров в топ} - N}{N}, \quad (7)$$

где Recall@N – доля рекомендованных релевантных товаров от общего количества релевантных товаров:

$$\text{Recall@N} = \frac{\text{количество релевантных товаров в топ-}N}{\text{всего релевантных товаров}}. \quad (8)$$

б) Mean Reciprocal Rank (MRR).

Средний обратный ранг (MRR) измеряет, насколько рано в списке рекомендаций появляется релевантный товар:

$$\text{MRR} = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{1}{\text{ранг первого релевантного товара для } u_i}, \quad (9)$$

в) Mean Average Precision (MAP).

MAP вычисляет точность на различных уровнях N :

$$\text{MAP} = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{1}{N} \sum_{k=1}^N \text{Precision@k}. \quad (10)$$

В качестве иллюстрации рассмотрим упрощенный пример вычисления Precision@N и MAP.

Предположим, что для пользователя сгенерирован список рекомендованных объектов из пяти позиций, из которых релевантными оказались только объекты на 2-й и 4-й позициях. Тогда расчет метрик на примере приведен в таблице 3.

Таблица 3 – Пример вычисления Precision@k для одного пользователя

Позиция k	Рекомендованный товар	Релевантность ($\text{rel}(k)$)	Precision@k
1	Item 1	0	0/1=0,00
2	Item 2	1	1/2=0,50
3	Item 3	0	1/3≈0,33
4	Item 4	1	2/4=0,50
5	Item 5	0	2/5=0,40

Из таблицы 3 видно, что релевантные объекты расположены на позициях 2 и 4. Значения Precision@k на этих позициях равны 0,50 и 0,50 соответственно. Средняя точность (AP) для данного пользователя вычисляется как среднее арифметическое этих двух значений:

$$\text{AP} = \frac{0,50 + 0,50}{2} = 0,50.$$

Поскольку в данном примере один пользователь, то $\text{MAP} = \text{AP} = 0,50$.

Обсуждение

Практический эксперимент на синтетических данных показал, что предложенная модель может давать неплохие рекомендации. Например, при $N = 5$ полученный средний Precision@5 составил около 0,60, а метрика MAP – около 0,56. Данные показатели говорят о том, что более половины первых пяти рекомендаций были релевантны пользователю, что подтверждает работоспособность контентной модели на выбранном наборе данных.

Отмечается, однако, что точность модели зависит от правильного выбора признаков объектов и качества формирования пользовательских профилей. В реальных системах вероятна проблема узкой специализации (overspecialization), когда система повторно рекомендует очень похожие товары. Этого можно избежать, комбинируя контентный подход с колаборативным или включая в модель дополнительные факторы (гибридный подход) [8].

В рамках обсуждения важно подчеркнуть, что оптимизация только одной метрики не

всегда приводит к лучшей пользовательской удовлетворенности. Поэтому анализ качества модели должен дополняться бизнес-метриками (например, CTR, конверсия в покупку) и оценкой разнообразия рекомендаций. Система не может предоставить персонализированные рекомендации до тех пор, пока пользователь не проявит какую-либо активность (просмотрит несколько объектов, поставит оценки, укажет явные предпочтения). Это может привести к потере новых пользователей, если они сразу не получают релевантных предложений.

Тем не менее представленные результаты демонстрируют, что простая контентная модель способна обеспечивать приемлемые рекомендации даже на ограниченных данных.

Выводы

В работе представлено исследование разработанной модели рекомендательной системы в торговой сфере на основе контентной фильтрации. Сформулированы актуальность проблемы, цель и задачи. Описаны методы формирования профилей пользователей и вычисления сходства на основе признаков объектов. Выполнен синтетический эксперимент, подтвердивший, что контентная фильтрация позволяет получать релевантные рекомендации (например, Precision @ 5≈0,60, MAP≈0,56). В качестве дальнейших направлений работы рекомендуется тестирование модели на реальных данных и ее расширение гибридными методами для повышения разнообразия и точности рекомендаций.

Список литературы

1. Федоренко, В. И. Использование методов векторизации текстов на естественном языке для повышения качества контентных рекомендаций фильмов / В. И. Федоренко, В. С. Киреев // Современные научноемкие технологии. – 2018. – № 3. – С. 102–106.
2. Сейдаметова, З. С. Системы рекомендаций в электронной коммерции / З. С. Сейдаметова // Ученые записки Крымского инженерно-педагогического университета. – 2018. – № 3(61). – С. 121–127.
3. Горелов, М. А. О подходах к построению моделей рекомендательных систем / М. А. Горелов, Ю. В. Бруттан // Математическое моделирование систем и процессов. – Псков : Псковский государственный университет, 2024. – С. 82–87.
4. Folorunso, Ojo. An Overview of Web Content Filtering Techniques / Ojo Folorunso. – Morgan State University Maryland USA. – 2024. – 9 p.
5. Karthikeyan, V. K. T. Web Content Filtering Techniques : A Survey / V. K. T. Karthikeyan. – Текст : электронный // A Survey International Journal of Computer Science & Engineering Technology (IJCSET). – 2014. – Vol. 5. – № 03. – P. 203–208. – URL: <https://ijcset.com/docs/IJCSET14-05-03-038.pdf> (дата обращения: 12.06.2025).
6. Черников, С. Ю. Использование системного анализа при управлении организациями / С. Ю. Черников, Р. В. Корольков // Моделирование, оптимизация и информационные технологии. – 2014. – № 2(5). – С. 1–6.
7. Преображенский, Ю. П. О методах создания рекомендательных систем / Ю. П. Преображенский, В. М. Коновалов // Вестник воронежского института высоких технологий. – 2019. – № 4(31). – С. 75–79.
8. Hybrid quality-based recommender systems: A systematic literature review / B. Sabiri, A. Khtira, B. El Asri, M. Rhanoui. – Текст : электронный // Journal of Imaging. – 2025. – Vol. 11. – № 1. – DOI: 10.3390/jimaging11010012. – URL: https://www.researchgate.net/publication/387804221_Hybrid_Quality-Based_Recommender_Systems_A_Systematic_Literature_Review (дата обращения: 12.06.2025).
9. Скворцов, В. А. Примеры метрических пространств / В. А. Скворцов. – Москва : МЦНМО. – 2002. – 24 с. – ISBN: 5-94057-002-X.

Л. П. Вовк, М. В. Волин

**Автомобильно-дорожный институт (филиал)
федерального государственного бюджетного образовательного учреждения
высшего образования «Донецкий национальный технический университет» в г. Горловка
Математическая модель построения системы персональных рекомендаций
на основе признаков объектов**

Рассмотренная в статье модель рекомендательной системы на основе контентной фильтрации демонстрирует эффективность и пригодность для использования в задачах персонализации. Предложенный подход базируется на использовании признаков объектов и профилей пользователей, что позволяет учитывать индивидуальные предпочтения при генерации рекомендаций. Синтетический эксперимент подтвердил работоспособ-

ность модели: показатели Precision@5 и MAP показали, что более половины предложенных рекомендаций были релевантны.

Предлагаемая модель рекомендательной системы на основе контентной фильтрации, несмотря на свои ограничения, обладает рядом уникальных особенностей, которые делают ее особенно ценной для задач персонализации. Эти особенности определяют ее преимущества и недостатки, а также сценарии наиболее эффективного применения. Она не зависит от поведения других пользователей. Профиль каждого пользователя строится исключительно на его личной истории взаимодействий и явных предпочтениях. Это позволяет системе давать очень точные и релевантные рекомендации для пользователей с уникальными, нишевыми или специфическими вкусами, которые могут не совпадать с общими трендами или предпочтениями большинства.

Тем не менее модель чувствительна к выбору признаков и качеству представления пользовательского профиля, а также склонна к эффекту «узкой специализации». Это подчеркивает целесообразность дальнейших исследований в направлении гибридных подходов, сочетающих контентную и колаборативную фильтрацию, а также применение модели к реальным данным.

Развитие предложенной методики может способствовать созданию более точных и разнообразных рекомендательных систем, обладающих высокой пользовательской ценностью и потенциальным коммерческим эффектом.

РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА, КОНТЕНТНАЯ ФИЛЬТРАЦИЯ, ПРОФИЛЬ ПОЛЬЗОВАТЕЛЯ, МЕТРИКА PRECISION@N, СРЕДНЯЯ ТОЧНОСТЬ, СИНТЕТИЧЕСКИЙ ЭКСПЕРИМЕНТ

L. P. Vovk, M. V. Volin

Automobile and Road Institute (Branch) of the Federal State Budget Educational Institution

of Higher Education «Donetsk National Technical University» in Gorlovka

Mathematical Model for Constructing a Personalized Recommendation System

Based on Object Features

The model of a content-based filtering recommendation system considered in the article demonstrates its efficiency and suitability for use in personalization tasks. The proposed approach relies on the use of object attributes and user profiles, which makes it possible to consider individual preferences when generating recommendations. A synthetic experiment confirmed the model's effectiveness: the Precision@5 and MAP metrics indicated that more than half of the recommended items were relevant.

The proposed model of a content-based recommendation system, despite its limitations, has a number of unique features that make it especially valuable for personalization tasks. These features determine its advantages and disadvantages, as well as the scenarios for its most effective use. It does not depend on the behavior of other users. Each user's profile is built exclusively on his personal interaction history and explicit preferences. This allows the system to provide very accurate and relevant recommendations for users with unique, niche, or specific tastes that may not coincide with general trends or preferences of the majority.

However, the model is sensitive to the choice of features and the quality of the user profile representation, and is prone to the «narrow specialization» effect. This underlines the importance of further research in the direction of hybrid approaches that combine content-based and collaborative filtering, as well as testing the model on real-world data.

The development of the proposed methodology can contribute to the creation of more accurate and diverse recommendation systems with high user value and potential commercial impact.

RECOMMENDATION SYSTEM, CONTENT FILTERING, USER PROFILE, PRECISION@N METRIC, AVERAGE ACCURACY, SYNTHETIC EXPERIMENT

Сведения об авторах:

Л. П. Вовк

SPIN-код РИНЦ: 9860-6682

Телефон: +7 949 301-98-55

Эл. почта: leonidvovk166@gmail.com

М. В. Волин

Телефон: +7 949 446-91-00

Эл. почта: michael.ayy@yandex.ru

Статья поступила 18.06.2025

© Л. П. Вовк, М. В. Волин, 2025

Рецензент: М. М. Гуменюк, канд. экон. наук, доц.,

Автомобильно-дорожный институт

(филиал) ДонНТУ в г. Горловка