

**Анализ современных подходов в проектировании рекомендательных систем**  
**Analysis of modern approaches of recommendation systems design**



**УДК 004.588**

**Разуваев Кирилл Андреевич,**

*Московский Государственный Технический Университет имени Н.Э. Баумана, г. Москва*

**Гринберг Ханна Эдуардовна,**

*Московский Государственный Технический Университет имени Н.Э. Баумана, г. Москва*

**Маслова Александра Сергеевна,**

*Московский Государственный Технический Университет имени Н.Э. Баумана, г. Москва*

**Веинский Владимир Андреевич,**

*Московский Государственный Технический Университет имени Н.Э. Баумана, г. Москва*

**Милютин Алексей Валерьевич,**

*Московский Государственный Технический Университет имени Н.Э. Баумана, г. Москва*

**Razuvaev Kirill Andreevich,**

*Graduate Student, Bauman Moscow State Technical University, Moscow, e-mail: volkovkirill97@mail.ru*

**Grinberg Khanna Eduardovna,**

*Graduate Student, Bauman Moscow State Technical University, Moscow, e-mail: hanna.grinberg@gmail.com*

**Maslova Alexandra Sergeevna,**

*Graduate Student, Bauman Moscow State Technical University, Moscow, e-mail: m\_sasha1998@mail.ru*

**Veinskiy Vladimir Andreevich,**

*Graduate Student, Bauman Moscow State Technical University, Moscow, e-mail: VeinskiyVA@mail.ru*

**Milyutin Alexey Valerievich,**

*Graduate Student, Bauman Moscow State Technical University, Moscow, e-mail: alexejmilyutin200@gmail.com*

**Аннотация.** В сети интернет существует огромное количество вариантов выбора различных товаров, поэтому необходимо производить фильтрацию для предоставления наиболее

необходимой информации покупателям и сокращения количества вариантов этих товаров у конкретного пользователя. Одним из решений данной проблемы являются рекомендательные системы, которые посредством поиска и обработки информации в большом объеме данных, предлагают результаты, наиболее подходящие конкретным пользователям, и предоставляют персонализированный контент и услуги. В данной статье будут рассмотрены различные характеристики и возможности различных методов прогнозирования в рекомендательных системах.

**Summary.** There are a huge number of options for choosing different products on the Internet. So, it is necessary to filter them order to provide the most necessary information to customers and to reduce the number of these products for a particular user. One of the better solutions to decide this problem is recommendation systems. There are offer results which the most suitable for users by searching and processing information in a large volume of data. More recommendation systems and provide personalized content and services for a particular user. This article will discuss the different characteristics and possibilities of different forecasting methods in recommendation systems.

**Ключевые слова:** рекомендательные системы, коллаборативная фильтрация, фильтрация, основанная на описании товара, гибридная фильтрация, предсказание.

**Keywords:** recommendation systems, collaborative filtering, content-based filtering, hybrid filtering, precision.

### Введение

В связи со стремительным ростом количества пользователей и информации в сети Интернет, пользователи сталкиваются с проблемой быстрого поиска необходимого им контента. Современные поисковые системы частично решили эту проблему, однако до сих пор существует задача персонализированных предложений для клиентов. Рекомендательные системы позволяют предсказывать предпочтения пользователей на основании их данных.

Рекомендательные системы полезны как для потребителей, так и для поставщиков услуг. Они помогают снизить затраты на поиск и фильтрацию информации в среде интернет-магазинов, онлайн-кинотеатров, музыкальных сервисов и т.д., а также улучшают процесс принятия решений. Рекомендательные системы в сфере электронной коммерции увеличивают доходы, поскольку являются эффективным средством для увеличения продаж.

#### 1. Анализ области

Рекомендательная система помогает пользователям принимать решения в сложных информационных системах. Также, система обеспечивает доступ к возможности поиска по данным, основанным на интересах и предпочтениях пользователей. В случаях, если данных о

предпочтениях пользователя недостаточно, рекомендательные системы помогают в принятии решений на основании рекомендаций для других людей.

В настоящее время разработаны различные подходы к проектированию рекомендательных систем, таких как: коллаборативная фильтрация (Collaborative Filtering), модели, основанные на описании товара (Content-based), и гибридная рекомендательная система (Hybrid Filtering)[6, с. 162].

Метод коллаборативной фильтрации является наиболее распространенным на данный момент. Коллаборативная фильтрация позволяет рекомендовать товары на основании выбора других пользователей с похожими вкусами. Рекомендательные системы на основе коллаборативной фильтрации активно используются в различных областях. Одними из примеров таких систем являются Netflix, Ringo и Amazon.

Netflix — это стриминговый сервис, предоставляющий разнообразные фильмы и сериалы на основании предпочтений пользователей. Ringo использует коллаборативную фильтрацию для формирования профилей пользователей на основании их оценок музыкальных альбомов. Amazon применяет коллаборативную фильтрацию и алгоритмы диверсификации для составления матрицы продуктов и последующей их рекомендации при анализе предыдущих покупок пользователей [4, с. 6208].

Content-based системы, наоборот, основываются на информации о конкретном пользователе, а не других пользователях. Данные системы сопоставляют описание товара с интересами пользователя, полученными из его предыдущих оценок, и требуют наличие описание у каждого товара.

Несмотря на то, что перечисленные выше системы хорошо себя зарекомендовали, у них есть несколько минусов, снижающих качество рекомендаций. Рекомендательные системы, основанные на content-based фильтрации ограничиваются отдельными категориями товаров и плохо взаимодействуют с разреженными данными. Методы коллаборативной фильтрации подвержены таким проблемам, как: холодный старт (отсутствие необходимого количества данных для корректной работы системы), разреженность и масштабируемость данных [5, с. 115].

Гибридные рекомендательные системы включают в себя два и более методов фильтрации различными способами для повышения точности и производительности системы. Применение данного типа рекомендательных систем дает возможность использовать только сильные стороны различных методов и сводить к нулю слабые.

## **2. Этапы работы рекомендательных систем**

### **2.1. Сбор информации**

На этапе сбора информации производится накопление информации для создания профилей пользователей и моделей для прогнозирования на основе их предпочтений и поведения. Системе необходимо собрать как можно больше информации о пользователе, иначе она будет работать некорректно.

Информация о пользователе включает в себя его интересы, предпочтения, когнитивные особенности, а также данные о прошлых взаимодействиях с системой.

Рекомендательные системы могут получать информацию о предпочтениях пользователей как напрямую от них, так и на основании их предыдущих действий в сети Интернет (отзывы, оценки, лайки и т.д.).

Модели пользователей необходимы для получения наиболее точных рекомендаций вне зависимости от используемого метода прогнозирования.

Способы получения информации о пользователях в рекомендательных системах подразделяются на:

- прямое взаимодействие;
- косвенное взаимодействие;
- гибридное взаимодействие.

Рассмотрим способы взаимодействия с пользователем для получения информации подробнее.

### **2.1.1. Прямое взаимодействие**

При прямом способе получения информации пользователю предлагается напрямую оценить различные продукты для формирования и улучшения его модели. Точность предсказания в данном случае напрямую зависит от количества оценок, предоставленных пользователем.

Единственным недостатком данного метода является то, что пользователям необходимо самостоятельно предоставить достаточно информации о своих предпочтениях, но что они не всегда готовы тратить время. Несмотря на это, текущий подход считается наиболее надежным, поскольку предоставленная лично пользователем информация является наиболее достоверной. Из этого следует, что рекомендательная система предоставит наиболее точные прогнозы [3, с. 309].

### **2.1.2. Косвенное взаимодействие**

В процессе косвенного взаимодействия система автоматически получает информацию о действиях и предпочтениях пользователей на основании истории их предыдущих покупок, навигации, нажатия на кнопки или ссылки на странице и времени, проведенном на различных страницах продуктов и услуг. Косвенное взаимодействие не требует усилий от

пользователя, однако рекомендации, основанные на информации, собранной данным способом, будут менее точными [3, с. 318].

Однако существует также и мнение, что сбор информации таким способом является более объективным, поскольку не возникает предвзятости в поведении пользователя, то есть действия происходят на подсознательном уровне и лучше отражают его индивидуальность выбора и личные особенности.

### 2.1.3. Гибридное взаимодействие

В гибридном способе используются сильные стороны обоих методов, что позволяет увеличить эффективность работы системы и минимизировать их недостатки. Особенность данного метода заключается в том, что модель, построенная на основании одного из них (прямого, либо косвенного), корректируется в соответствии с информацией, полученной другим методом.

## 2.2. Обучение системы

Процесс обучения строится на основании собранных данных о пользователе и предоставленной им информации. Для обучения рекомендательной системы используются различные алгоритмы машинного обучения. Как и для любого алгоритма машинного обучения, мы должны уметь оценивать производительность наших рекомендательных систем, чтобы решить, какой алгоритм лучше всего подходит для нашей ситуации. Методы оценки для рекомендательных систем в основном можно разделить на два набора: оценка, основанная на четко определенных показателях, и оценка, в основном основанная на человеческом суждении и оценке удовлетворенности.

## 2.3. Процесс рекомендации

На завершающем этапе обученная система предсказывает и выделяет самые подходящие продукты для пользователя. После этого система собирает новые данные, переобучается и уточняет прогнозы.

## 3. Основные подходы в рекомендательных системах

На данный момент в рекомендательных системах используются методы, представленные на рисунке 1.



Рисунок 1 – Виды фильтрации в рекомендательных системах

Наиболее популярные среди них — коллаборативная фильтрация и content-based фильтрация.

### **3.1. Content-based фильтрация**

Данный метод является алгоритмом, зависящим от предметной области, поскольку прогнозирование основано на атрибутах товаров. Рекомендации на основе описания больше всего подходят для веб-страниц, публикаций, новостей и статей. Процесс рекомендации базируется на особенностях предыдущих действий пользователя, а интересы других пользователей никак на него не влияют. Также, в случае изменения профиля пользователя система может скорректировать предсказания за достаточно короткое время.

В content-based фильтрации используются различные типы моделей для поиска сходства между документами и создании точных предсказаний. Наиболее популярными моделями являются: наивный Байесовский классификатор, TF/IDF, деревья решений и нейронные сети [2, с. 355].

Недостатком такого метода является то, что он требует обязательного описания товара и предыдущих оценок пользователя.

### **3.2. Коллаборативная фильтрация**

Коллаборативная фильтрация — это метод, при котором рекомендации даются на основе интересов других похожих пользователей. Он работает путем создания матрицы предпочтений пользователей и товаров. Далее, он вычисляет сходства между профилями пользователей, чтобы подобрать необходимые продукты.

В результате работы коллаборативной фильтрации можно получать как прогнозы, так и рекомендации. Прогноз — это числовое значение  $R_{ij}$ , отражающее прогнозируемую оценку элемента  $j$  для пользователя  $i$ , а рекомендация — это список из  $N$  лучших продуктов, которые понравятся пользователю больше всего [1, с. 110].

В свою очередь у данного метода существует ряд недостатков, таких как холодный старт (неполнота и недостаточность информации в самом начале о новом продукте, услуге, а так же и о пользователе), масштабируемость (невозможность моментального отклика системы из-за сложных расчетов алгоритма коллаборативной фильтрации вмещающем информацию о большом количестве пользователей и предметов в системе), разреженность данных (пользователи не всегда ставят оценки всем приобретенным товарам, из-за чего построенная системой матрица «предмет-пользователь» получается очень большой и разреженной, что приводит к сложности расчета рекомендаций).

### **3.3. Гибридная фильтрация**

Гибридная фильтрация сочетает в себе различные методы и подходы, что позволяет минимизировать ошибки в процессе рекомендаций. Идея подхода заключается в том, что недостатки одного метода нивелируются преимуществами другого [2, с. 362].

Работа гибридной фильтрации может быть построена либо на отдельной реализации алгоритмов и объединении их результата, либо на создании единой рекомендательной системы, совместно использующей эти алгоритмы.

#### 4. Оценка качества рекомендаций

В процессе оценки качества работы рекомендательной системы используются различные метрики в зависимости от используемого метода фильтрации. Эти метрики используются для определения точности рекомендаций. Точность — это доля правильных рекомендаций из общего числа возможных рекомендаций, в то время как охват измеряет долю объектов в области поиска, для которых система может предоставить рекомендации.

Метрики, используемые для оценки качества можно разделить на:

- статистические показатели точности;
- показатели точности принятия решений.

##### 4.1 Статистические метрики

Для оценки статистических метрик проводится сравнение прогнозируемых оценок с фактическими оценками пользователя. В качестве показателей здесь используются: среднее абсолютное отклонение (Mean Absolute Error), среднеквадратичная ошибка (Mean Squared Error) и корень среднеквадратичной ошибки (Root Mean Squared Error).

MAE вычисляется по следующей формуле:

$$MAE = \frac{1}{N} \sum_{u,i} |p_{u,i} - r_{u,i}|$$

где:

$P_{u,i}$  — предсказанный рейтинг продукта  $i$  для пользователя  $u$ ;

$R_{u,i}$  — действительный рейтинг продукта  $i$  для пользователя  $u$ ;

$N$  — размер выборки.

MSE вычисляется по следующей формуле:

$$MSE = \frac{1}{N} \sum_{u,i} (p_{u,i} - r_{u,i})^2$$

RMSE вычисляется по следующей формуле:

$$RMSE = \sqrt{\frac{1}{N} \sum_{u,i} (p_{u,i} - r_{u,i})^2}$$

#### 4.2 Метрики точности принятия решений

Метрики данного класса работают с бинарными данными (например, 0 и 1, да и нет и т.п.). То есть выставляется порог и если оценка меньше, например, 3,5, она считается “плохой”, а если больше — “хорошей”.

Данные метрики включают вычисление трех основных значений. Вначале находится Precision (P, доля рекомендаций, понравившихся пользователю) по формуле:

$$P = \frac{TP}{TP + FP}$$

Затем Recall (R, доля интересных пользователю товаров, которая показана) по формуле:

$$R = \frac{TP}{TP + FN}$$

Среднегармоническое метрик P (Precision) и R (Recall) F1-measure вычисляется по формуле:

$$F1 = \frac{2PR}{P + R}$$

Данные показатели рассчитываются на основании того, сколько раз система приняла верное или неверное решение. TP — истинно-положительное решение (должны быть в результате и предсказаны), TN — истинно-отрицательное (должны быть в результате, но не предсказаны), FP – ложно-положительное (не должны быть в результате но предсказаны), FN — ложно-отрицательное (не должны быть в результате и не предсказаны).

#### Заключение

Рекомендательные системы открывают новые возможности для предоставления клиентам персонализированных рекомендаций товаров и услуг. Они помогают решить проблему информационной перегрузки, при которой пользователю приходится тратить достаточно много времени на поиск необходимых именно ему продуктов.

В данной статье были рассмотрены основные концепции и подходы в реализации рекомендательных систем, выявлены их сильные и слабые стороны и метрики оценки.

#### Список литературы

1. Bobadilla J. et al. Recommender systems survey //Knowledge-based systems. – 2013. – Т. 46. – С. 109-132.
2. Burke R. Hybrid recommender systems: Survey and experiments //User modeling and user-adapted interaction. – 2002. – Т. 12. – №. 4. – С. 331-370.
3. Buder J., Schwind C. Learning with personalized recommender systems: A psychological view //Computers in Human Behavior. – 2012. – Т. 28. – №. 1. – С. 207-216.
4. Jalali M. et al. WebPUM: A Web-based recommendation system to predict user future movements //Expert Systems with Applications. – 2010. – Т. 37. – №. 9. – С. 6201-6212.
5. Konstan J. A., Riedl J. Recommender systems: from algorithms to user experience //User modeling and user-adapted interaction. – 2012. – Т. 22. – №. 1. – С. 101-123.
6. Pathak B. et al. Empirical analysis of the impact of recommender systems on sales //Journal of Management Information Systems. – 2010. – Т. 27. – №. 2. – С. 159-188.