**Приложение 2**

**Индекс УДК: 004.896**

**Писарева А.И.**

**АЛГОРИТМЫ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ: ПРОБЛЕМЫ И РЕАЛИЗАЦИЯ**

**Аннотация статьи на русском языке**

Статья посвящена автоматизации бизнес-процессов при помощи внедрения алгоритмов рекомендательных систем. Рассмотрены алгоритмы реализации рекомендательных систем. Описаны основные проблемы при реализации алгоритмов рекомендаций. Рассмотрены подходы к вычислению наилучших рекомендаций.

**Ключевые слова на русском языке**

Рекомендательная система, бизнес-процессы, матрица предпочтений, искусственный интеллект, персонализация пользовательского опыта, неперсонализированные рекомендации, content-based рекомендации, коллаборативная фильтрация, стандартизация данных.

Современный уровень технологий оказывает сильное влияние на организацию процесса управления деятельности предприятия. Разработка сквозных цифровых технологий, таких как искусственный интеллект, большие данные, системы описательной и предиктивной аналитики, виртуальная реальность, рекомендательные системы и др., определяет вектор развития экономики настоящего времени, способствуя активному внедрению в работу предприятий автоматизированных систем.

Внедрение автоматизированных рекомендательных систем – сложный и трудоемкий процесс, требующий особого внимания, так как от работы алгоритмов системы зависит вся процедура взаимодействия с пользователем, определение его предпочтений и вывод соответствующих рекомендаций.

Среди основных характеристик рекомендательных систем, таких как степень персонализации и прозрачность, алгоритмы, используемые в рекомендательных системах, подразделяют на следующие виды:

* Неперсональные (Summary-based);
* Модели, основанные на описании товара (Content-based);
* Коллаборативная фильтрация (Collaborative Filtering);
* Методы, основанные на матричном разложении (Matrix Factorization)

В независимости от вида рекомендательной системы, ядром ее является матрица предпочтений.

Матрица предпочтений состоит из пользователей и объектов. На пересечении (пользователь, объект) матрица заполняется вероятностными значениями в соответствии с заинтересованностью пользователя совершить покупку того или иного товара. Вероятностные значения заинтересованности клиента в свою очередь, выявляется благодаря обобщению данных, которые совершает рекомендательная система, получая из небольшой выборки, в которой пользователи предварительно производили оценку.

Рекомендательные системы содержат несколько типов групп товаров, это связано с тем, что рекомендуемые товары делятся на повторяемые и неповторяемые товары. Однако существуют товары, которые тяжело классифицировать из-за персональных особенностей покупателя, возможной приверженности бренду, или желанием попробовать все, что представлено на полках.

В независимости от типа рекомендаций, лучше подбирать рекомендации под конкретного пользователя - моделировать категории клиента. Моделирование категории клиента осуществляется через работу с рейтингами. Получение таких данных происходит либо явно, когда пользователь назначает рейтинг товару или как-то его комментирует, либо неявно, когда пользователь совершил покупку и значит, выбрал товар наилучшим среди прочих.

Вычисление наилучших рекомендаций осуществляется при помощи различных подходов.

### **Неперсонализированные рекомендации.**

Интерес пользователя определяется средним рейтингом товара. Данный подход в настоящее время использует сервис TripAdvisor, его суть заключается в том, что если нравится большинству, рекомендацию стоит показывать всем.

Такой подход не лишен недостатков, одним из которых является проблема холодного старта. Данная проблема выражается в недостатке данных для отладки работы системы в случае, когда товар покупают не часто или, он вовсе является новым. В таких ситуациях принято искусственно создавать рейтинги товаров для отладки работы системы. Также можно применить сглаженное среднее значение, идея которого заключается в том, что при небольшом количестве проставленных рейтингов, округление происходит в пользу наиболее безопасного среднего показателя, а как только набирается достаточное количество новых оценок, усредняющая корректировка перестает действовать.

### **Content-based рекомендации.**

Идея подхода заключается в сопоставлении описания товара и интересов пользователя, в соответствии с данными прошлых поставленных им рейтингов или комментариев. Неструктурированные признаки описываются для текста векторами в пространстве слов (Vector-Space model). Каждый элемент такого вектора – признак, потенциально характеризующий интерес пользователя. Аналогично, продукт – вектор в том же пространстве. По мере взаимодействия пользователя с системой, векторные описания приобретенных им товаров объединяются (суммируются и нормализуются) в единый вектор и, таким образом, формируется вектор его интересов. Далее достаточно найти товар, описание которого наиболее близко к вектору интересов, т.е. решить задачу поиска n ближайших соседей. Не все элементы одинаково значимы: например, союзные слова, очевидно, не несут никакой полезной нагрузки. Поэтому при определении числа совпадающих элементов в двух векторах все измерения нужно предварительно взвешивать по их значимости. Данную задачу решает известное в Text Mining преобразование TF-IDF, которое назначает больший вес более редким интересам. Такая фильтрация почти полностью повторяет механизм query-document matching, используемый в поисковых системах типа Яндекс и Google.

### **Коллаборативная фильтрация**

Реализация такого подхода основывается на действиях многих пользователей, в основе идея k ближайших соседей. Идея такова, что для каждого пользователя нужно найти k наиболее похожих на него предпочтений и дополнить информацию о пользователе известными данными по его соседям. По причине своей сложности данный алгоритм трудно реализуем на практике. Данная проблема отчасти может быть решена использованием высокопроизводительных устройств. Кроме того, необходимо ввести корректировки в алгоритм:

* обновлять расстояния батчами (например, раз в день),
* обновлять матрицу расстояний инкрементально,
* сделать выбор в пользу итеративных и приближенных алгоритмов (например, ALS).

Эффективность алгоритма требует выполнения следующих допущений:

* Вкусы людей не меняются временем (или меняются, но для всех одинаково).
* Если вкусы людей совпадают, то они совпадают во всем.

Важный этап подготовки данных — нормализация оценок, осуществляемая при помощи стандартизации данных (scaling). Процедура нормализации может быть реализована следующими способами:

* центрирование (mean-centering) — применима исключительно для небинарных матриц. Центрирование происходит путем вычитания из оценок пользователя средней оценки.
* стандартизация (z-score) — вдобавок к центрированию происходит деление оценки на стандартное отклонение.
* двойная стандартизация — первый раз нормируем оценками пользователя, второй раз — оценками товара.   
  Если у фильма «Самый лучший фильм» средняя оценка 2.5, а пользователь ей ставит 5, то это сильный фактор, говорящий о том, что такие фильмы ему явно по вкусу.

Проблему «похожести» или корреляции предпочтений двух пользователей можно решить разными способами, суть которых заключается в сравнении двух векторов.

Корреляция Пирсона — классический коэффициент, который применим и при сравнении векторов. Основной его минус — когда пересечение по оценкам низкое, корреляция может быть высокой случайно.

Корреляция Спирмана. Основное отличие — коэффициент ранговый, т.е. работает не с абсолютными значениями рейтингов, а с их порядковыми номерами. Полученные результаты очень близки к корреляции Пирсона.

Косинусное расстояние - еще один классический коэффициент. Косинус угла между стандартизированными векторами — это и есть корреляция Пирсона.

Важно, чтобы пользователь доверял рекомендательной системе, а для этого она должна быть проста и понятна. При необходимости всегда должно быть доступно понятное объяснение рекомендации. В рамках объяснения можно показывать оценку товара соседями, по какому именно атрибуту было совпадение, а также выводить уверенность системы в оценке (confidence). Чтобы не перегружать интерфейс, допускается перенос информации в кнопку «Tell me more». Внедрение рекомендательных систем эффективно, как для пользователя, так и для компании, только в том случае, когда грамотно и верно обоснованы и рассчитаны все необходимые алгоритмы их работы.

**Информация об авторе (-ах) на русском языке**

Писарева Анастасия Игоревна, Россия, Санкт-Петербург. Студент 1 курса магистратуры,

факультет «Технологического менеджмента и инноваций», Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Кронверкский проспект 49, abit@itmo.ru.

**Библиографический список на русском языке**

1. N. Palmer. iBPMS: Intelligent BPM Systems: Intelligent BPM Systems: Impact and Opportunity [Текст] / N. Palmer. – М.: Future Strategies Inc., 2014. – 220 c.
2. N. Perkin. P. Abraham. Building The Agile Business Through Digital Transformation [Текст] / N. Perkin, P. Abraham – М.: Kogan Page Ltd., 2017. – 288 c.

**Pisareva A.I.**

**ALGORITHMS FOR RECOMMENDATION SYSTEMS: PROBLEMS AND IMPLEMENTATION**

**Аннотация статьи на английском языке**

*The article is devoted to the automation of business processes by introducing algorithms of recommendation systems. Algorithms for the implementation of recommendation systems are considered. The main problems in the implementation of recommendation algorithms are described. Approaches to calculating the best recommendations are considered.*

**Ключевые слова на английском языке**

*Recommendation system, business processes, preference matrix, artificial intelligence, personalization of user experience, non-personalized recommendations, content-based recommendations, collaborative filtering, data standardization.*

**Информация об авторе (-ах) на английском языке**

Pisareva Anastasia Igorevna, Russia, St. Petersburg. 1st year master's student of the Faculty «Technology Management and Innovation», St. Petersburg National Research University of   
Information Technologies, Mechanics and Optics. Kronverksky Avenue 49 , abit@itmo.ru.

**Библиографический список на английском языке**

1. N. Palmer. iBPMS: Intelligent BPM Systems: Intelligent BPM Systems: Impact and Opportunity [Текст] / N. Palmer. – М.: Future Strategies Inc., 2014. – 220 c.
2. N. Perkin. P. Abraham. Building The Agile Business Through Digital Transformation [Текст] / N. Perkin, P. Abraham – М.: Kogan Page Ltd., 2017. – 288 c.