

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.02:517:303.722.4

Мойсюк-Дранько
Павел Алексеевич

АЛГОРИТМЫ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ

АВТОРЕФЕРАТ

на соискание степени магистра технических наук

по специальности 1–40 80 02 «Системный анализ, управление и обработка
информации»

Научный руководитель

М. П. Ревотюк

кандидат технических наук, доцент

Минск 2021

ВВЕДЕНИЕ

Рекомендательные системы возникли и начали развиваться со середины 90-х г. прошлого столетия. Главной задачей рекомендательной системы считается предоставление персонализированных рекомендаций пользователям, способные учитывать их предпочтения в процессе выбора предметов (услуг, товаров или объектов). Потребности пользователей в употреблении определенной информации и их невозможность обработки всей информации привели к бурному развитию технологий, связанных с поиском информации. Поисковые системы могут эффективно реализовывать задачи поиска и делать доступными внушительные объемы информации широкому кругу пользователей. Впрочем, зачастую поисковые системы могут возвращать значительно больше информации, нежели пользователь способен обработать. С увеличением объемов данных, которые сохраняются в сети и предлагаются пользователю, повышается актуальность проблемы, заключающейся в опережении запроса пользователя посредством предложения ему вероятно интересной информации. Данную проблему можно решить при помощи систем предоставления рекомендаций. Ключевое различие алгоритмов систем предоставления рекомендаций от алгоритмов поиска данных состоит в предложении ответа без очевидного запроса от пользователя, одновременно, поисковые алгоритмы отвечают на определенный запрос пользователя. Ведущие компании-разработчики поисковых систем (Yahoo!, Google) отмечают, что будущее всемирной сети именно в системах предоставления рекомендаций. Рекомендательные системы играют важную роль в передовых средствах электронной коммерции. Главная задача электронной коммерции заключается в получении прибыли благодаря максимальному удовлетворению потребностей потребителей. По этой причине субъекты электронной коммерции затрачивают значительные усилия и применяют средства для изучения, развития и усовершенствования средств и методов построения рекомендательных систем. Ведущие компании Amazon, Yandex, Google, Yahoo! проводят исследования, касающиеся развития и усовершенствования средств и методов создания рекомендательных систем для задач электронной коммерции.

К важным задачам развития рекомендательных систем в области электронной коммерции в настоящее время относятся: повышение точности прогнозирования рекомендаций, решение проблемы влияния разреженности и размерности матрицы пользователь-предмет на точность прогнозирования рекомендаций, решение проблемы нового пользователя и нового предмета.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

Актуальной является научная задача разработки методов и средств построения рекомендательных систем для задач электронной коммерции, которые будут включать в себя преимущества существующих методов и средств, и, при этом, повысят конверсию и лояльность пользователей субъекта электронной коммерции, что позволит эффективно решать основную задачу электронной коммерции - получение максимальной прибыли при максимальном удовлетворении потребностей пользователей.

Степень разработанности проблемы

Проблема нового пользователя и нового предмета в рекомендательных системах для задач электронной коммерции играет важное значение. Для любого субъекта электронной коммерции его прибыль зависит от количества пользователей, пользующихся его услугами и количества предметов, которые он предлагает пользователям. При вхождении нового пользователя в систему или введения нового предмета в систему векторы профилей для них отсутствуют. В таком случае рекомендательная система не может провести ни одной рекомендации. Традиционно для создания профиля пользователя или предмета проводится опрос пользователей или анализ поведения пользователя на сайте субъекта электронной коммерции. Однако это требует значительных временных затрат и может снизить конверсию сайта субъекта электронной коммерции. Особенно важным для электронной коммерции является обеспечение лояльности пользователей при первом посещении сайта субъекта электронной коммерции. Поэтому актуальной является задача разработки методов и средств ввода в рекомендательную систему для задач электронной коммерции новых пользователей и новых предметов, которые бы обеспечили максимальную конверсию сайта и лояльность пользователей и, в конечном варианте, обеспечили максимальную прибыль субъекту электронной коммерции.

Цель и задачи исследования

Целью магистерской диссертации является разработка и исследование методов и алгоритмов построения рекомендательных систем, направленных на повышение достоверности, точности и надежности рекомендаций для пользователей услуг электронной коммерции.

Объектом исследования выступают процесс прогноза рекомендаций в рекомендательных системах.

Предметом исследования является методы, средства, модели и алгоритмы прогнозирования рекомендаций в рекомендательных системах.

Для достижения поставленных целей основными задачами являются:

1. Анализ современного состояния методов и средств проектирования рекомендательных систем;
2. Анализ методов расчета коэффициентов сходства векторов профилей пользователей и предметов в моделях прогнозирования рекомендаций;
3. Усовершенствование метода расчета коэффициентов сходства векторов профилей пользователей и предметов для метода взвешенной суммы за счет учета разреженности векторов профилей и евклидового расстояния между векторами;
4. Анализ применения методов интеллектуального анализа данных в методах и алгоритмах построения рекомендательных систем для задач электронной коммерции;
5. Разработка метода ввода нового пользователя в систему, основанный на концепции ассоциативных правил, что не требует регистрации пользователя в системе и учитывает интересы зарегистрированных в системе пользователей;
6. Усовершенствования гибридного метода прогнозирования рекомендаций, включающий метод нечеткой кластеризации *c-means*, что осуществляет контроль деления на кластеры и позволяет использовать в расчетах методом коллаборативной фильтрации только векторы профилей предметов, имеющих наибольшие значения характеристической функции принадлежности к нечетким кластерам;
7. Разработка метода введения в рекомендательную систему профиля нового пользователя и нового предмета с помощью категориальной кластеризации векторов демографических атрибутов пользователей и контентных атрибутов предметов;
8. Разработка математического и алгоритмического обеспечения рекомендательной системы;
9. Исследования разработанных методов и алгоритмов на тестовых данных.

Теоретическая и методологическая основа исследования

При проведении исследования и написании диссертации использованы научные публикации, техническая документация и интернет-источники, посвященные вопросам алгоритмам рекомендательных систем.

Для решения поставленных задач в работе используются: методы линейной алгебры, регрессионного анализа, теория реляционных баз данных, математическая статистика, методы интеллектуального анализа данных, кластерный анализ, теория нечетких множеств.

В работе применялись моделирование и методы компьютерного проектирования.

Научная новизна

Научная новизна заключается в создании алгоритма предоставления рекомендаций новому пользователю в системе, который, в отличие от существующих, учитывает интересы существующих в системе пользователей и не требует регистрации нового пользователя в системе.

Теоретическая значимость работы заключается в исследовании проблемы эффективности рекомендательных алгоритмов.

Практическая значимость работы заключается в анализе и совершенствовании существующих, разработке новых алгоритмов рекомендаций, а также исследования использования комбинаций алгоритмов, которые дают лучшие результаты.

Основные положения, выносимые на защиту

1. Обзор и анализ достоинств и недостатков существующих алгоритмов рекомендательных систем.
2. Анализ использования применения методов интеллектуального анализа данных в методах и алгоритмах построения рекомендательных систем.
3. Реализован метод введения в рекомендательную систему профиля нового пользователя нового предмета на основе категориальной кластеризации.

СОДЕРЖАНИЕ РАБОТЫ

Диссертация состоит из введения, общей характеристики работы, четырех глав, заключения, библиографического списка и приложения.

В первой главе проведен анализ достоинств и недостатков алгоритмов рекомендательных систем, рассмотрены основные моменты рекомендательных систем. Изучены техники коллаборативной фильтрации, рассмотрены основные алгоритмы машинного обучения применяемые с целью оптимизации тестирования.

Во второй главе представлена программные комплексы и средства для извлечения трансформации и преобразования анализа данных.

В третьей главе выполнено проектирование алгоритмов рекомендаций и алгоритмов рекомендательной системы.

В четвертой главе представлены результаты исследования, полученные реализацией программным средством по сравнению с ближайшими аналогами по эффективности использования.

Библиотека БГУИР

ЗАКЛЮЧЕНИЕ

В работе решена научная задача разработки методов и средств построения рекомендательных систем для задач электронной коммерции. При этом получены следующие результаты:

1. Проведен анализ существующего состояния развития методов, средств и алгоритмов построения рекомендательных систем, что позволило определить недостатки методов, алгоритмов и программных систем.

2. Усовершенствован метод расчета коэффициентов сходства векторов профилей пользователей и предметов для метода взвешенной суммы, который, в отличие от существующих, учитывает разреженность и длину векторов профилей объектов, что позволило повысить точность рекомендаций. Проведенные экспериментальные исследования усовершенствованного метода по сравнению с косинусной степени близости и степени близости с корреляцией Пирсона показали увеличение точности прогнозирования на 5%.

3. Получен метод предоставления рекомендаций новому пользователю рекомендательной системы, который, в отличие от существующих, базируется на концепции ассоциативных правил, учитывает интересы существующих в системе пользователей и не требует регистрации нового пользователя в системе. Полученные решения подтверждаются теоретическим обоснованием базовых положений и результатами апробации.

4. Получил дальнейшее развитие гибридный метод прогнозирования рекомендаций, который, в отличие от существующих, использует метод нечеткой кластеризации *c-means*, что осуществляет контроль деления на кластеры и позволяет использовать в расчетах методом коллаборативной фильтрации векторы профилей предметов, имеющих наибольшие значения характеристической функции принадлежности к нечетким кластерам. Проведенные экспериментальные исследования развитого гибридного метода по сравнению с применением классического метода четкой кластеризации показали увеличение точности прогнозирования рекомендаций на 6%.

5. Впервые разработан метод введения в рекомендательную систему профиля нового пользователя и нового предмета на основе категориальной кластеризации, который, в отличие от существующих, использует только демографические атрибуты пользователей и контентные атрибуты предметов, которые вводятся в систему при регистрации объекта и не требует предварительного опроса пользователей или анализа поведения пользователей на сайте. Полученные решения подтверждаются теоретическим обоснованием базовых положений и результатами апробации.

6. Разработаны и исследованы методы прогнозирования рекомендаций, основанные на уравнениях линейной и упрощенной линейной регрессии. Проведенные экспериментальные исследования разработанных методов показали, что при потере в точности прогнозирования 3% метод упрощенной линейной регрессии дает выигрыш во времени расчета 5%.

7. Разработано методика тестирования точности методов и алгоритмов прогнозирования рекомендаций, основанную на использовании тестовой

матрицы рейтингов пользователь-предмет и разделении ее на тестовую и прогнозируемую части.

8. Разработано информационное обеспечение рекомендательной системы, которое включает базу данных оценок пользователь-предмет. Разработана структура программного обеспечения, которое позволяет выбирать в диалоговом режиме метод прогнозирования, метод расчета метрики сходства, метод тестирования точности.

На основании всего изложенного, считаю, что поставленная цель достигнута в полной мере. В процессе работы учитываются все достоинства и исправлены недостатки конкурирующих средств. Выполненная работа принесла мне хороший опыт как молодому специалисту, открыла много возможностей для дальнейшего развития в сфере IT-услуг.

Библиотека БГУИР

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

1-А Мойсюк-Дранько П. А., Методы матричной факторизации для систем рекомендаций / Мойсюк-Дранько П. А. // Материалы международной конференции «Информационные технологии и системы 2020» учреждения образования «Белорусский государственный университет информатики и радиоэлектроники» Минск, БГУИР, 2020.

Библиотека БГУИР