

Министерство образования Республики Беларусь
Учреждение образования
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Сидорович
Антон Олегович

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА АНАЛИЗА И КОНКРЕТИЗАЦИИ
ПОЛЬЗОВАТЕЛЬСКИХ ПОТРЕБНОСТЕЙ

АВТОРЕФЕРАТ

на соискание степени магистра технических наук

по специальности 1-31 80 10 Теоретические основы информатики

Научный руководитель
Качков Владимир Петрович
кандидат технических наук, доцент

Минск 2016

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Обоснование необходимости разработки системы

По большей своей части существующие рекомендательные системы жестко привязаны к конкретному приложению, где они используются. На данный момент существует мало независимых систем, которые можно интегрировать с уже существующими сервисами. Именно задача универсальности и возможности интеграции с уже существующими решениями на данный момент наиболее актуальна.

Пользователи различных систем должны получать предложения от сервисов только тогда, когда они наиболее склонны принять его. При получении оптимальных результатов (максимального количества принятых из предложенных предложений) увеличиваются объемы продаж, удовлетворенность клиентов, а также понижаются объемы затрат и ресурсов. Качественные маркетинговые решения позволяют добиться именно этого.

Большинство существующих систем рекомендаций товаров пользователям основываются только на статических данных, известных о пользователе. Исходя из этого, была поставлена следующая задача – создание системы управления предложениями в реальном времени с ориентированием на централизованное взаимодействие с клиентами по различным каналам. Это означает, что система опирается не только на статические данные о пользователях, но также узнает о предпочтениях пользователей через их прошлые покупки, настраивает маркетинговые сообщения в точках контакта, и предсказывает, какое предложение будет наиболее интересно в текущем контексте.

Цель и задачи исследования

Целью диссертационной работы является разработка алгоритма и программного обеспечения для решения задачи предугадывания предпочтений пользователей, на основании их предыдущего поведения, модели классификации пользовательских запросов и визуализации результатов работы данной системы.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Изучить существующие подходы и методы построения рекомендательных систем.

2. Ознакомиться с существующими аналогичными решениями и программными продуктами и определить основные функциональные требования к системе.
3. Разработать алгоритм классификации данных.
4. На основании разработанного алгоритма классификации разработать программное средство, решающее задачу предугадывания предпочтений пользователей.
5. Изучить существующие способы оценки точности и скорости рекомендационных алгоритмов.
6. Провести экспериментальное исследование разработанного программного продукта.

Объектом исследования являются рекомендательные сервисы.

Предметом исследования является математическое и программное обеспечение компьютерных систем для решения задачи рекомендательной системы, методы и алгоритмы построения классификации данных, оценка скорости и точности существующих и разработанных алгоритмов и визуализация результатов.

Основной *гипотезой*, положенной в основу диссертационной работы, является возможность улучшения точности и скорости работы алгоритмов рекомендации в процессе практического использования системы. Данный подход, основанный на уже существующих алгоритмах классификации, позволяет проанализировать ключевые элементы, обобщить результаты и произвести сравнение производительности различных моделей классификации, существующих на сегодняшний день.

Личный вклад соискателя

Результаты, полученные в диссертации, получены соискателем лично. Вклад научного руководителя В. П. Качкова заключается в формулировке целей и задач исследования.

Структура и объем диссертации

Диссертация состоит из перечня условных обозначений и терминов, введения, общей характеристики работы, четырех глав, заключения, списка использованных источников. В первой главе представлен анализ предметной области, выявлены основные существующие методы и подходы построения рекомендательных систем, проведен их анализ. Вторая глава посвящена разработке алгоритма, решающего задачу рекомендательной системы, в частности алгоритма бинарной классификации. Представлены основные

положения для построения алгоритма классификации. В третьей главе рассмотрены вопросы оценки производительности модели классификации. В рамках решения практической задачи, описанный алгоритм реализован в виде отдельного программного средства, которое рассмотрено в четвертой главе.

Общий объем работы составляет 67 страниц, из которых основного текста – 59 страниц, 15 рисунков на 13 страницах, список использованных источников из 25 наименований на 2 страницах.

Библиотека БГУИР

КРАТКОЕ СОДЕРЖАНИЕ

1.1 Методы и подходы построения рекомендательных систем

На первом этапе работы над диссертацией был выполнен анализ существующих подходов и методов построения рекомендательных систем.

Была рассмотрена одна из самых распространенных и эффективных методик построения рекомендаций – *коллаборативная фильтрация*. В частности изучены два основных подхода: фильтрация по пользователям (user based) и фильтрация по предметам (item based). Детально были изучены метрики для определения сходства пользователей, основанного на расстоянии (Евклидово расстояние, расстояние Хемминга), корреляция как сходство, затронуты алгоритмы кластеризации.

В ходе первой главы магистерской диссертации был проведен обзор существующих решений для задач рекомендательных систем: Apache Mahout, Weka, Solr.

Отдельное внимание было уделено разделу машинного обучения *Data Mining*, были выделены его задачи, а также методы обучения.

1.2 Реализация рекомендательной системы и алгоритма классификации

Вторая часть диссертации посвящена реализации рекомендательной системы как программного продукта, разработке алгоритма классификации пользователей, а также ее экспериментальному использованию.

В этой главе приводится более детальный и конкретный материал, описанный в первой части диссертации. В частности анализируется такая область как *классификация* данных. Детально изучается и анализируется частный случай классификации – бинарная классификация (классификация на множестве двух значений), а также простой классификатор Байеса, который в дальнейшем является основой для построения алгоритма классификации. Дается подробное описание архитектуры системы рекомендаций, ее основных модулей и их взаимодействие. Затрагиваются также ключевые процессы и сообщения внутри программного продукта, описывается разработанный жизненный цикл моделей классификации.

Особое место во второй главе отводится описанию *алгоритма классификации*, его основы, которые затрагивают области теории вероятности

и математической статистики. Больше внимание уделяется вычислению корреляции данных внутри системы, дается обоснование вычислений, подробно описывается пошаговое выполнение алгоритма.

В результате экспериментального использования алгоритма (и системы в целом) был получен ряд зависимостей, которые позволили внести в него некоторые дополнения. Эти изменения повысили эффективность алгоритма классификации, что подтверждается при тестировании на реальных данных.

1.3 Оценка эффективности модели

В третьей главе данной работы рассматриваются способы определения эффективности моделей классификации, основные концепции при построении оценок. В том числе рассмотрены визуальные показатели производительности алгоритма (ROC-кривая, чувствительность и специфичность, Lift и Gain диаграммы).

Рассмотренные методы оценки эффективности модели классификации были внедрены в окончательный программный продукт в панель администрирования, для мониторинга и управления рекомендательной системой, а также анализа работы алгоритма на реальных данных.

1.4 Практическое использование системы

В заключительной главе магистерской диссертации приведены варианты практического использования рекомендательной системы (через панель администрирования), визуализация данных в системе, описание процесса создания предложений.

ЗАКЛЮЧЕНИЕ

Данная магистерская диссертация явилась логическим продолжением дипломного *проекта* магистранта. Результаты использования программного продукта, разработанного в рамках дипломного проекта, были собраны и исследованы.

Результатом данной работы является создание эффективной модели классификации данных для рекомендательной системы и создание специального программного обеспечения, включающего в себя реализацию алгоритма, систему рекомендаций, а также панель мониторинга и администрирования системы предсказания.

В ходе предварительного анализа материалов на тему магистерской диссертации была изучена литература, посвященная методам и подходам построения рекомендательных систем, рассмотрены существующие решения. Были изучены способы машинного обучения с целью дальнейшей реализации. Выделены и обобщены предпосылки, а также приведено обоснование написания магистерской диссертации.

Большая часть работы посвящена архитектуре рекомендательной системы, взаимодействию ее модулей, ключевых положений и наработок. Главной частью магистерской диссертации является анализ существующих алгоритмов классификации и описание разработанного программного продукта. В ходе экспериментального использования данного алгоритма и системы в целом, были выделены закономерности и зависимости, которые в дальнейшей позволили улучшить показания моделей классификации на реальных данных. Работа алгоритма бинарной классификации тесным образом граничит с теорией вероятности и математической статистикой, информация о которых также была изложена в данной работе.

Большое внимание было уделено оценке эффективности работы системы и алгоритма классификации, внедрению данных оценок в окончательный программный продукт, что позволяет производить мониторинг и администрирование системы, с целью дальнейших потенциальных улучшений, а также для анализа поведения пользователей.

Разработанное программное обеспечение было внедрено в производство. За время использования, были собраны пожелания и замечания, которые будут учтены в следующих версиях.

Магистерская диссертация выполнена в соответствии с индивидуальным учебным планом магистерской подготовки. Было создано программное средство, выполняющее задачи, поставленные во время выполнения магистерской работы. У системы имеются широкие перспективы использования в коммерческой сфере как в виде самостоятельного продукта, так и в виде модуля более крупного приложения. В дальнейшем планируется развивать и усовершенствовать алгоритм классификации, а также работать над увеличением производительности рекомендательной системы.

Библиотека БГУИР