

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

УДК 004.418:659.1

Холодок
Антон Александрович

СИСТЕМА ИНФОРМИРОВАНИЯ И ОПОВЕЩЕНИЯ ПОТРЕБИТЕЛЕЙ НА
ОСНОВАНИИ ФИЛЬТРАЦИИ ТОВАРОВ

Автореферат
на соискание степени магистра технических наук

по специальности 1-40 80 04 Математическое моделирование, численные
методы и комплексы программ

Научный руководитель
Новиков Владимир Иванович
кандидат технических наук, доцент

Минск 2015

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью диссертационной работы является разработка алгоритма и программного обеспечения для решения задачи предсказания поведения пользователей, на основании модели классификации 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 решениями и программными продуктами для построения рекомендательных сервисов.
3. Разработать алгоритм классификации.
4. На основании разработанного алгоритма классификации разработать программное средство, решающее задачу реального сектора экономики.
5. Изучить существующие способы оценки производительности модели рекомендации. Интегрировать данную оценку в разработанный программный продукт.
6. Провести экспериментальное исследование разработанного программного продукта.

Объектом исследования являются рекомендательные сервисы и визуализация данных.

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

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

Личный вклад соискателя

Результаты, полученные в диссертации, получены соискателем лично. Вклад научного руководителя В. И. Новикова 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 62 страницы, из которых основного текста – 54 страницы, 21 рисунок на 17 страницах, список использованных источников из 22 наименований на 2 страницах.

ВВЕДЕНИЕ

В эпоху научно-технического прогресса все большую популярность на рынке клиентского обслуживания стали приобретать услуги информационно-справочных служб. Секрет популярности использования данного инструмента заключается в том, что именно этот метод организации сервиса приносит невероятный эффект при реализации продукции предприятий. Используя услуги информационно-справочных служб, предприятия предоставляют круглосуточный доступ к информации о своей продукции всем потенциальным потребителям, причем независимо от места их расположения, а также поддерживают обратную связь со своими клиентами, закрепляя долгие и крепкие отношения сотрудничества. У предприятий отпадает срочная потребность в использовании дополнительных рекламных кампаний, так как всю информацию клиент может узнать именно благодаря использованию информационно-справочного сервиса.

Профессионально организованный информационно-справочный сервис предоставляет возможность отслеживать поток клиентов и их интересов, тем самым добывая бесценную маркетинговую информацию для усовершенствования продукции предприятий согласно желаниям потребителя. В качестве услуги сервис предоставляет помощь в выборе необходимого продукта из всего предложенного ассортимента. [1]

Таким образом, услуги информационно-справочного сервиса обеспечивают ряд преимуществ для бизнес-процесса предприятий, среди которых:

- круглосуточная информационная поддержка потребителей;
- обратная связь с клиентами вне зависимости от места расположения;
- возможность учета желаний и предпочтений клиентов;
- отсутствие зависимости от успешности рекламных кампаний.

На сегодняшний день стабильно растет доля веб-приложений и количество пользователей таких систем. Такой рост вполне логичен – сейчас все стремятся к удобству и использованию новых возможностей во всех сферах жизни. Веб-приложения уходят из стационарных компьютеров и ноутбуков в мобильные телефоны и смартфоны – адаптивные приложения. [2] Один и тот же ресурс доступен со всех видов устройств, в любом месте, в любое удобное для пользователя время.

В современном мире приложения обеспечивают выполнение необходимых пользователю функций. Люди получают и обмениваются полезной информацией из любого места, экономя свое время и ресурсы. Веб-приложение – это в первую очередь инструмент для решения какой-либо задачи.

В основу данной работы легли результаты полученные магистрантом во время подготовки дипломного проекта. Несмотря на объем работы проделанный в рамках дипломного проекта, некоторые вопросы остались за его рамками. Исследование этих вопросы и явилось основной целью данной работы. Материал изложенный в диссертации расширяет накопленный теоритический материал. Формулирует гипотезы о возможности его применения, для улучшения некоторых показателей эффективности работы алгоритма классификации. Эти гипотезы, в случае возможности, проверяются на практике. По результатам проверки делается заключение о целесообразности применения новых методов и подходов. В результате некоторые из них применяются в реализации алгоритма и используются на практике.

Также уделяется внимание архитектуре разработанного программного средства. Факт того, что алгоритм является универсальным, формирует требования к компонентам программы. Компоненты содержащие общую логику выносятся в отдельный модуль, который может быть использован для решения смежных задач.

Связь исследуемого предмета с реальным сектором экономики дает возможность провести проверку полученных результатов и внедрить разработанное программное на производство. Результатом выполненной работы должен являться алгоритм, который сочетает в себе уже наработанные методы в области классификации данных. Этот алгоритм должен быть оптимизирован с учетом специфики исследуемой задачи. Также должны быть исследованы возможности расширения уже существующих алгоритмов с целью получение максимальной информации о сравниваемых моделях данных.

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. В конце первой главы приведено обоснование необходимости разработки рекомендательной системы и ее предпосылки.

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, статистику и динамику изменений, включающие в себя:

- статистику существующих данных в системе
- динамику прибыли клиента
- популярные предложения, динамика их изменений и соотношений
- динамика эффективности моделей классификации
- корреляции, Gains и Lift диаграммы

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 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 тщательно изучена литература, посвященная методам и подходам построения рекомендательных систем, рассмотрены существующие решения. Были тщательно 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 описание разработанного алгоритма классификации. 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, внедрение данных оценок в окончательный программный продукт, что позволяет производить мониторинг и администрирование системы, с целью дальнейших потенциальных улучшений, а также для анализа поведения пользователей.

Проработанное программное обеспечение было внедрено в производство и получило положительные оценки клиентов. Вместе с тем, за

время использования, были собраны пожелания и замечания, которые будут учтены в следующих версиях.

В итоге получилось раскрыть тему магистерской диссертации и на практике реализовать описанные идеи. В дальнейшем планируется развивать и усовершенствовать алгоритм классификации, а также работать над увеличением производительности рекомендательной системы.

Библиотека БГУИР