

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

УДК \_\_\_\_\_

Асташкевич  
Михаил Геннадьевич

Биллинг-система локальной сети общежития

Автореферат  
на соискание степени магистра технических наук  
по специальности 1-40 80 04 «Математическое моделирование, численные  
методы и комплексы программ»

Научный руководитель  
Волорова Наталья Алексеевна  
кандидат технических наук, доцент

Минск 2016

## КРАТКОЕ ВВЕДЕНИЕ

Развитие информационных технологий, в частности увеличение вычислительной мощности техники и удешевление хранения данных, позволяет собирать и анализировать огромные объемы информации о пользователях, в том числе данные о предпочтениях и поведении пользователей. Эти данные представляют большую ценность для различных организаций. Например, поисковые гиганты, такие как Google, Yandex и другие, используют такие данные о том или ином пользователе сети для ранжирования результатов поисковых запросов, а также для более эффективного подбора показываемых рекламных объявлений. Также в качестве примера можно привести эксперимент исследователей из бизнес-школы Уорика (Warwick Business School), которые предположили возможность прогнозирования поведения рынка ценных бумаг на основе анализа открытых данных сервиса Google Trends. В результате исследователи смогли увеличить портфель акций на 8% за две недели - что является достаточно впечатляющим результатом, учитывая, что исходные данные брались из открытых источников.

Таким образом преследуемая автором цель заключается в исследовании современных подходов к прогнозированию и построению рекомендаций, а также изучение возможности предоставления исходных данных для таких алгоритмов разработанной ранее автором биллинг-системой.

# ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

## Цель и задачи исследования

Основные цели данной работы следующие:

- исследовать возможность прогнозирования на основе данных, обрабатываемых биллинг-системой;
- провести анализ существующих методов анализа и обработки данных для последующего прогнозирования;
- реализовать рекомендательную систему на основе алгоритма Random Walk with Restarts;
- использовать традиционные методы расчёта рекомендации для составления графа социальной сети;
- оценить качество рекомендаций для разных способов построения графа;
- предложить метод ускорения расчёта рекомендаций для применения в крупных социальных сетях;
- оценить время и качество рекомендаций при использовании предложенного метода.

*Объектом* исследования являются алгоритмы прогнозирования.

*Предметом* исследования является математическое и программное обеспечение компьютерных систем для решения задач составлений пользовательских рекомендаций, методы и алгоритмы рекомендационных систем.

Основной *гипотезой*, положенной в основу диссертационной работы, является возможность составления графа рекомендаций для работы в режиме реального времени, а также ускорение данного процесса путем разделения графа рекомендаций на персонализированные подграфы. Разработка тест-системы, позволяющей провести сравнение модификаций алгоритма.

## Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя Н. А. Волоровой, заключается в формулировке целей и задач исследования.

## Апробация результатов диссертации

Составные части диссертационной работы докладывались и обсуждались на 52-й научной конференции аспирантов, магистрантов и студентов (Минск, Беларусь, 2016).

### **Опубликованность результатов исследования**

По теме диссертации опубликована 1 печатная работа в сборнике трудов и материалов локальной конференции.

### **Структура и объем диссертации**

Диссертация состоит из введения, общей характеристики работы, четырех глав, заключения, списка использованных источников, списка публикаций автора. В первой главе произведен анализ существующей биллинг-системы и обзор ее возможностей по сбору интересующих нас данных. Во второй главе проанализирован современный подход к задаче о предоставлении рекомендаций, сделан обзор классических и современных алгоритмов. Проведен обзор методов оценки качества рекомендательных систем, выбран оптимальный метод оценки качества, подходящий под тему исследования. В третьей главе описаны детали реализации описанного алгоритма и разработки системы тестирования алгоритмов для последующего сравнения качества предоставляемых алгоритмами результатов. В четвертой главе проводится разбор результатов исследования. Рассмотрены различные подходы к составлению социального графа и описаны методы, позволяющие существенно улучшить качество рекомендаций по сравнению с оригинальной версией алгоритма. В рамках решения практической задачи, описанные алгоритмы реализованы в виде отдельного программного средства. Общий объем работы составляет 50 страниц, из которых основного текста – 40 страниц, 9 рисунков на 9 страницах, список использованных источников из 13 наименований на 2 страницах.

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

В первой главе данной работы произведен анализ существующей биллинг-системы и обзор ее возможностей по сбору интересующих нас данных.

Во второй главе проанализирован современный подход к задаче о предоставлении рекомендаций, сделан обзор классических и современных алгоритмов. Подробнее разобраны алгоритмы SVD (как наиболее эффективный алгоритм, используемый многими сервисами) и Random Walk with Restarts (как алгоритм, являющийся непосредственным объектом исследования в данной работе). Кроме того проведен обзор методов оценки качества рекомендательных систем, выбран оптимальный метод оценки качества, подходящий под тему исследования.

В третьей главе описаны детали реализации описанного алгоритма и разработки системы тестирования алгоритмов для последующего сравнения качества предоставляемых алгоритмами результатов. Описаны теоретический и практический методы получения тестовых данных. Описана платформа Apache Mahout, предоставляющая свободный доступ к распределенным алгоритмам машинного обучения: кластеризации, сингулярного разложения, классификации и алгоритмам коллаборативной фильтрации и рекомендации.

В четвертой главе проводится разбор результатов исследования. Рассмотрены различные подходы к составлению социального графа и описаны методы, позволяющие существенно улучшить качество рекомендаций по сравнению с оригинальной версией алгоритма. Предложены и исследованы подходы к упрощению алгоритма: метод построения персональных подграфов для каждого пользователя, симуляция описанного алгоритма, призванная заменить точный расчет предпочтений.

## ЗАКЛЮЧЕНИЕ

В работе исследованы подходы к использованию данных, предоставляемых разработанной ранее биллинг-системой для расчета персонализированных рекомендаций пользователям на примере анализа музыкальных предпочтений пользователя, полученных путем синтаксического анализа заголовков HTTP-запросов. Рассмотрены различные подходы к данной задаче, подробно исследован алгоритм Random Walk with Restarts. Разработаны и проанализированы две оптимизации данного алгоритма, один из которых показал лучшие результаты, чем оригинальный алгоритм, а также имеет ряд преимуществ, такие как возможность использовать его в режиме реального времени, а также более низкую ресурсоемкость.

- реализована Java-платформа для конфигурирования и тестирования рекомендательных систем;

- предложена модификация алгоритма Random Walk with Restarts, использующая традиционные способы коллаборативной фильтрации и дающая более качественные рекомендации;

- предложен способ упрощения оригинального алгоритма, подходящий для применения в крупных социальных сетях, проведено исследование качества предложенного метода;

- приведены результаты симуляции алгоритма Random Walk with Restarts;

- описанные методы протестированы на наборе данных социальной сети Last.fm.

## СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

[1 - А.] Асташкевич М.Г. Исследование современных методов построения рекомендательных систем / М.Г. Асташкевич // 52-я научно-техническая конференция аспирантов, магистрантов и студентов БГУИР: Тезисы доклада - Минск, 2016 - С. 34 - 35.

Библиотека БГУИР