

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.
ЧЕРНЫШЕВСКОГО»**

Кафедра Дифференциальных уравнений и математической экономики

**Разработка рекомендательной системы на основе
алгоритмов машинного обучения**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 441 группы

направления 09.03.03 Прикладная информатика

механико-математического факультета

Ковзалова Александра Алексеевича

Научный руководитель
профессор, д.э.н., профессор

В.А. Балаш

Заведующий кафедрой,
зав.кафедрой, д.ф-м.н., профессор

С.И. Дудов

Саратов 2024

Введение В современном информационном обществе огромное количество данных и контента доступно для пользователей через различные онлайн-платформы: от интернет-магазинов и видеосервисов до социальных сетей и новостных порталов. Однако с избытием выбора возникает проблема информационного перегруза, когда пользователи сталкиваются с трудностями в поиске контента, который соответствует их интересам и предпочтениям.

В этом контексте рекомендательные системы играют важную роль, помогая пользователям находить наиболее релевантный контент на основе их прошлого поведения, интересов и предпочтений. Рекомендации могут быть в различных форматах: от рекомендаций товаров и фильмов до музыки, статей новостей и даже людей для социальной сети.

В данной работе будут рассмотрены основные подходы к построению рекомендательных систем, включая методы коллаборативной фильтрации, контентные методы, гибридные подходы и новейшие технологии, такие как глубокое обучение. Основной акцент будет сделан на анализе принципов работы каждого метода, их преимуществах и недостатках, а также на практических аспектах их применения.

Целью данной работы является исследование различных подходов к построению рекомендательных систем, выявление их особенностей и эффективности, а также предоставление практических рекомендаций для проектирования и реализации рекомендательных систем в различных прикладных областях а также разработка предсказательной модели вида

$$\hat{r}(i_1, i_2, \dots, i_M) \approx r_{i_1, \dots, i_M},$$

$$i_1 \in I_1, \dots, i_M \in I_M,$$

где I_1, \dots, I_M — конечные множества объектов некоторой природы, у которых можно брать признаковые описания $x_1(i_1) \in R^{D_1}, \dots, x_M(i_M) \in R^{D_M}$ в качестве дополнительной информации. Требуется обучать модель на известных наборах $\{(r_{i_1 i_M}, i_1, \dots, i_M)\}$, при этом учитывать потенциальную проблему данных без явного отклика. Используя язык программирования Python

Актуальность В настоящее время рекомендательные системы играют важную роль в повседневной жизни людей и в различных сферах бизне-

са. С постоянным ростом объемов данных и контента в онлайн-пространстве пользователи сталкиваются с проблемой информационного перегруза, когда им сложно ориентироваться в огромном количестве предложений и выбрать наиболее подходящий контент.

Рекомендательные системы помогают решить эту проблему, предоставляя персонализированные рекомендации на основе предпочтений, интересов и поведения пользователей. Это позволяет улучшить пользовательский опыт, повысить удовлетворенность клиентов и увеличить конверсию в различных онлайн-платформах, таких как интернет-магазины, стриминговые сервисы, социальные сети и другие.

В контексте бизнеса рекомендательные системы играют ключевую роль в увеличении продаж, улучшении клиентского обслуживания и повышении лояльности клиентов. Они помогают компаниям адаптироваться к индивидуальным потребностям своих клиентов, предлагая им наиболее релевантный контент и товары.

Таким образом, исследование и разработка рекомендательных систем остаются актуальными и востребованными в современном мире, и их роль будет продолжать расти с развитием технологий и изменением потребительских предпочтений. В этом контексте выполнение дипломной работы, посвященной разработке и исследованию рекомендательной системы, представляет собой значимый вклад в область информационных технологий и бизнеса.

Задачи:

- Классифицировать подходы к построению рекомендательных систем
- Исследование различных подходов к построению рекомендательных систем
- Разработка собственной рекомендательной системы

Структура и содержание работы

- Постановка задачи
- Коллаборативная фильтрация
- Обратная связь
- Способы оценки рекомендательных систем
- Практическая часть
- Приложение А Код построения рекомендательной системы

Постановка задачи для рекомендательных систем

Раздел «Постановка задачи» в дипломном проекте является ключевым этапом, определяющим цели и задачи, стоящие перед исследованием и разработкой рекомендательной системы. В данном разделе описывается контекст проблемы, формулируются цели и основные задачи, которые необходимо решить в процессе работы.

В автореферате представлен обзор основных проблем, связанных с построением рекомендательных систем, таких как информационный перегруз, холодный старт, разреженность данных и неявная обратная связь. Далее описывается цель работы, которая может включать разработку нового метода рекомендации, сравнение существующих алгоритмов, улучшение качества рекомендаций и т. д.

Затем в разделе формулируются конкретные задачи, необходимые для достижения поставленной цели. Это может включать в себя выбор подходящего алгоритма рекомендации, сбор и подготовку данных, разработку и тестирование модели, анализ результатов и т. д. Каждая задача описывается в достаточной детализации для понимания ее сути и взаимосвязи с общей целью работы.

Также в разделе могут быть указаны критерии успешного выполнения поставленных задач и методы их оценки. Это позволяет объективно оценить результаты работы и достигнуть поставленной цели.

Раздел «Постановка задачи» является основой всего дипломного проекта, поскольку четкое определение целей и задач позволяет систематизировать и структурировать работу, а также оценивать ее результаты на основе заданных критериев.

Коллаборативная фильтрация

Раздел "Коллаборативная фильтрация" в данном дипломном проекте представляет собой обзор основных методов и алгоритмов этого подхода к построению рекомендательных систем. Коллаборативная фильтрация основывается на идее использования информации о предпочтениях пользователей для генерации персонализированных рекомендаций.

В автореферате рассматриваются два основных типа коллаборативной фильтрации: пользовательский и предметный подходы. Пользовательская

коллаборативная фильтрация основана на схожести между пользователями, тогда как предметная коллаборативная фильтрация использует схожесть между объектами (товарами, фильмами и т. д.).

Для каждого типа коллаборативной фильтрации представлены основные принципы работы, а также описаны популярные алгоритмы, такие как метод ближайших соседей (kNN), сингулярное разложение (SVD) и скрытые факторные модели. Для оценки качества рекомендаций также рассматриваются различные метрики, такие как точность, полнота, среднее значение ошибки и другие.

Особое внимание в разделе уделяется проблемам, связанным с реализацией коллаборативной фильтрации в условиях неявной обратной связи и больших объемов данных. Также обсуждаются методы обработки разреженных матриц оценок пользователей и алгоритмы учета динамических изменений в предпочтениях пользователей.

Раздел "Коллаборативная фильтрация" представляет собой важный этап в исследовании и разработке рекомендательной системы, так как этот подход широко используется в практике и имеет ряд преимуществ, таких как способность к персонализации и эффективность в условиях разреженных данных.

Обратная связь

Раздел «Обратная связь» в данном дипломном проекте посвящен изучению роли и влияния обратной связи на работу рекомендательных систем. Обратная связь играет важную роль в улучшении качества рекомендаций, поскольку предоставляет информацию о реакции пользователей на предложенный контент.

В автореферате представлен обзор различных типов обратной связи, используемых в рекомендательных системах, включая явную и неявную обратную связь. Явная обратная связь включает в себя оценки, отзывы, рейтинги и другие действия, совершаемые пользователями, чтобы выразить свое мнение о контенте. В то время как неявная обратная связь основана на поведении пользователей, таком как просмотр, покупка, клики и т. д., которые могут свидетельствовать о предпочтениях и интересах.

Далее в разделе рассматривается методика сбора и анализа обратной связи, включая техники обработки и интерпретации полученных данных. Обсуждаются методы агрегирования обратной связи, фильтрации шума и использования обратной связи для обучения и адаптации рекомендательных моделей.

Особое внимание в разделе уделяется проблемам, связанным с учетом неоднородности и динамической природы обратной связи, а также методам управления обратной связью для предотвращения искажений в рекомендациях.

Раздел «Обратная связь» является важным компонентом дипломного проекта, поскольку позволяет понять механизмы работы системы и улучшить ее качество на основе обратной связи со стороны пользователей.

Способы оценки рекомендательных систем

В данном разделе дипломного проекта рассматриваются различные методы и метрики, используемые для оценки качества работы рекомендательных систем. Оценка рекомендательных систем является важным этапом в их разработке и оптимизации, поскольку она позволяет оценить эффективность предлагаемых алгоритмов и улучшить качество рекомендаций для пользователей.

В автореферате представлены основные критерии оценки, включая точность, полноту, покрытие, разнообразие и новизну рекомендаций. Каждая метрика имеет свои особенности и цели, например, точность оценивает долю правильных рекомендаций среди всех предложенных, а полнота измеряет способность системы предложить все подходящие элементы.

Для каждой метрики представлены формулы и методы вычисления, а также обсуждаются их преимущества и недостатки. Также обсуждаются сценарии использования каждой метрики и их роль в оценке различных типов рекомендательных систем, таких как коллаборативная фильтрация, контентные методы и гибридные подходы.

Особое внимание в разделе уделяется проблемам, связанным с оценкой рекомендательных систем в условиях неявной обратной связи и холодного старта. Обсуждаются методы кросс-валидации, а также различные страте-

гии формирования тестовой и обучающей выборок для корректной оценки работы системы.

Раздел "Способы оценки рекомендательных систем" представляет собой важный инструмент для исследования и разработки рекомендательных систем, поскольку он позволяет проводить сравнительный анализ различных подходов и выбирать наиболее эффективные методы для конкретных задач.

Заключение В ходе данной дипломной работы были решены ряд важных задач, связанных с исследованием и разработкой рекомендательной системы. Прежде всего, был проведен обширный обзор техник построения рекомендательных систем, что позволило выявить необходимость в разработке алгоритма коллаборативной фильтрации, способного учитывать произвольные признаковые описания признаков и работать с данными, содержащими неявную обратную связь.

Далее был осуществлен обзор различных алгоритмов коллаборативной фильтрации, что позволило оценить их преимущества и недостатки в контексте поставленной задачи. Также были рассмотрены способы оценки рекомендаций, что является важным аспектом при разработке рекомендательных систем.

Одним из ключевых моментов работы был экспериментальный анализ и сравнение алгоритмов, учитывающих неявную обратную связь и признаки пользователей. Этот этап позволил оценить эффективность различных подходов и выбрать наиболее подходящий для реализации рекомендательной системы.

В результате выполнения всех поставленных задач цель работы, а именно разработка рекомендательной системы, была успешно достигнута. Полученные результаты могут быть полезны для дальнейших исследований в области рекомендательных систем, а также для практического применения в различных сферах, таких как электронная коммерция, социальные сети, медиа и другие.