

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования

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

Кафедра математической кибернетики и компьютерных наук

**СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ПОСТРОЕНИЯ
РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 411 группы
направления 02.03.02 — Фундаментальная информатика и информационные
технологии
факультета КНиИТ
Бабаева Никиты Павловича

Научный руководитель
доцент, к. ф.-м. н.

Ю. Н. Кондратова

Заведующий кафедрой
доцент, к. ф.-м. н.

С. В. Миронов

Саратов 2025

ВВЕДЕНИЕ

В современном цифровом пространстве персонализированные рекомендательные системы стали неотъемлемой частью множества сервисов – от стриминговых платформ и интернет-магазинов до социальных сетей. С ростом объемов доступного контента пользователи сталкиваются с проблемой информационной перегрузки, когда выбор релевантного контента становится затруднительным. Особенно актуальна эта проблема в индустрии кино и сериалов, где количество доступных произведений исчисляется десятками тысяч наименований.

Эффективные рекомендательные системы решают эту проблему, анализируя предпочтения пользователей и предлагая им наиболее релевантный контент. Согласно исследованиям, около 80% просмотров на платформе Netflix происходит благодаря персонализированным рекомендациям. Это демонстрирует как коммерческую значимость таких систем, так и их полезность для конечных пользователей.

Методы построения рекомендательных систем активно развиваются: от классических алгоритмов коллаборативной фильтрации к более сложным подходам, использующим матричную факторизацию, и далее к современным методам глубокого обучения.

Таким образом, данная работа не только отвечает текущим научным и технологическим трендам, но и имеет непосредственную практическую значимость для стриминговых платформ, стремящихся улучшить пользовательский опыт и увеличить вовлечённость аудитории.

Цель бакалаврской работы: Провести сравнительный анализ методов матричной факторизации, гибридных моделей и подходов на основе трансформеров в задаче построения рекомендательной системы для фильмов.

Поставленная цель определила **следующие задачи:**

1. Провести обзор теоретических основ рекомендательных систем, включая их классификацию, ключевые проблемы и основные метрики качества рекомендаций;
2. Реализовать и настроить модели рекомендаций на данных о пользовательских предпочтениях в просмотре фильмов, используя датасет KION;
3. Провести сравнительный анализ результатов по количественным метрикам, времени обучения и инференса (получения предсказания от модели), анализ вычислительных ресурсов и визуальный анализ выдачи моделей;

4. Сделать выводы о преимуществах и недостатках каждого метода с практическими рекомендациями по их применению в зависимости от характеристик данных и требований к рекомендательной системе;

Структура и объём работы. Бакалаврская работа состоит из введения, 2 разделов, заключения, списка использованных источников и 13 приложений. Общий объём работы – 102 страницы, из них 67 страниц – основное содержание, включая 11 рисунков и 2 таблицы, список использованных источников информации – 20 наименований.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Первый раздел «Теоретические основы по разработке рекомендательных систем» посвящен основным методам и подходам, используемым при разработке рекомендательных систем (РС). РС представляют собой алгоритмы, предназначенные для предсказания предпочтений пользователей относительно различных объектов на основе анализа их поведения или характеристик объектов. Основная цель РС – это помочь пользователю в принятии решения о том, что выбрать из большого массива доступных объектов, будь то фильмы, товары, музыкальные треки или другие типы контента. РС помогают пользователю сориентироваться в мире избыточной информации и предлагают персонализированные варианты выбора, которые наиболее вероятно будут ему интересны.

Основные методы, используемые в РС, включают контентные системы, коллаборативную фильтрацию и гибридные модели, которые сочетают различные подходы для улучшения качества рекомендаций.

Контентные рекомендательные системы создают рекомендации на основе анализа характеристик объектов – в случае фильмов, такими характеристиками могут быть жанр, режиссер, актерский состав, страна производства и другие атрибуты. Принцип работы контентных систем состоит в том, чтобы создать профиль пользователя на основе его предпочтений и взаимодействий, а затем сравнить этот профиль с признаками объектов для формирования рекомендаций. Преимуществом таких систем является их способность рекомендовать новые или малоизвестные объекты, даже если они не были оценены другими пользователями, однако основной недостаток заключается в том, что такие системы не могут эффективно выявлять скрытые предпочтения пользователя, которые не выражаются напрямую через явные характеристики объектов.

Коллаборативная фильтрация, в свою очередь, основывается на анализе поведения и взаимодействий множества пользователей для выявления схожих паттернов и формирования рекомендаций, исходя из предположения, что пользователи со схожими предпочтениями в прошлом будут иметь схожие предпочтения в будущем. Коллаборативная фильтрация делится на два основных типа: user-based, где для целевого пользователя находят схожих пользователей, и рекомендации формируются на основе их предпочтений, и item-based, где для каждого объекта находят похожие объекты, и рекомендации строятся на основе объектов, которые пользователь уже оценил положительно.

Гибридные системы представляют собой комбинацию различных методов, что позволяет преодолеть ограничения отдельных подходов через взвешенные методы (результаты нескольких методов комбинируются с определенными весами), переключаемые методы (выбирают подход в зависимости от ситуации), каскадные подходы (один метод уточняет результаты другого) и комбинированные подходы (используют данные одного метода в качестве входных для другого).

Основной проблемой, с которой сталкиваются рекомендательные системы, является проблема холодного старта, возникающая когда система не имеет достаточно информации для формирования качественных рекомендаций для новых пользователей или объектов. Для решения этой проблемы применяются различные методы, включая использование контентной информации о пользователях и объектах, активное обучение для сбора информации о предпочтениях пользователя, а также использование демографических данных, таких как возраст и пол.

Анализ различных подходов показывает, что контентные системы хороши для обработки новых объектов, но плохо выявляют скрытые предпочтения пользователей; коллаборативная фильтрация хорошо работает при наличии большого количества данных о пользователях, но сталкивается с проблемами холодного старта и разреженности данных; гибридные модели, такие как LightFM, сочетают несколько методов, что позволяет значительно улучшить точность предсказаний и решать проблему холодного старта, однако требуют более сложной настройки и вычислительных ресурсов, при этом в будущем развитие гибридных методов и использование трансформеров могут значительно повысить точность и персонализацию рекомендаций.

Второй раздел «Реализация и сравнительный анализ методов построения рекомендательных систем» посвящен практической реализации и анализу различных методов рекомендательных систем на реальных данных.

В рамках эксперимента использовался датасет KION, содержащий информацию о предпочтениях пользователей в области кино, включая как явные, так и неявные взаимодействия между пользователями и объектами, данные о фильмах, жанрах, рейтингах и просмотрах, что делает его удобным для тестирования различных моделей рекомендательных систем.

Было реализовано четыре модели для сравнительного анализа: базовая

модель Popular (использует популярность объектов), модель на основе матричной факторизации ALS (алгоритм Alternating Least Squares), гибридная модель LightFM (сочетает коллаборативную фильтрацию с контентными признаками) и модель на основе трансформеров BERT4Rec (использует архитектуру BERT для анализа последовательностей взаимодействий), при этом каждая модель была обучена на одинаковых данных для обеспечения справедливого сравнения эффективности.

Таблица 1 – Сравнение метрик качества рекомендаций различных моделей. Жирные оценки являются лучшими в каждой строке, а подчеркнутые - вторые.

Metric	popular	ials	lightfm 1epoch	lightfm 15epoch	bert4rec
Recall@10	0.1642	0.1299	0.1543	<u>0.1695</u>	0.1870
HitRate@10	0.2695	0.2296	0.2573	<u>0.2831</u>	0.3111
MAP@10	0.0789	0.0586	0.0691	<u>0.0821</u>	0.0871
AvgRecPopularity@10	79906.58	44352.21	<u>72316.05</u>	72314.32	40446.33
Serendipity@10	2.07e-06	<u>4.75e-05</u>	5.29e-06	1.53e-05	1.88e-04
Intersection@10 _{popular}	1.0000	0.3478	0.7590	<u>0.7644</u>	0.3154
CoveredUsers@10	1.0000	1.0000	1.0000	1.0000	1.0000

Результаты эксперимента, представленные в таблице 1, показали, что модель BERT4Rec продемонстрировала наилучшие результаты по всем метрикам точности (Recall@10, HitRate@10 и MAP@10) и лидирует по метрике разнообразия Serendipity@10, что объясняется использованием архитектуры трансформеров, эффективно захватывающей долгосрочные зависимости в пользовательских последовательностях взаимодействий.

Гибридная модель LightFM (15 эпох) заняла второе место по показателям точности, благодаря использованию дополнительных признаков пользователей и объектов способна формировать как предсказуемые, так и частично неожиданные рекомендации, что делает её универсальным решением для различных сценариев, особенно эффективным при решении проблемы холодного старта.

Модель iALS показала умеренные результаты по основным метрикам точности, однако заняла второе место по метрике Serendipity@10, демонстрируя способность предлагать менее очевидные объекты, что особенно ценно с точки зрения пользовательского опыта, в то время как модель Popular ожидаемо показала стратегию рекомендации исключительно самых популярных элементов, что обеспечивает простую реализацию, но ограничивает персонализацию и

разнообразие рекомендаций.

Эти результаты подтверждают общую тенденцию: более сложные модели, использующие дополнительную информацию о пользователях и объектах, а также учитывающие последовательный характер взаимодействий, показывают лучшие результаты по метрикам точности рекомендаций.

Таблица 2 – Сравнение времени обучения и инференса моделей

Модель	Время обучения	Время инференса	Размер модели
Popular	< 1 мин	< 1 сек	< 1 МБ
ALS	7 мин	2 сек	3.3 ГБ
LightFM (1 epoch)	12 мин	4 сек	1.5 ГБ
LightFM (15 epochs)	165 мин	4 сек	1.5 ГБ
BERT4Rec	420 мин	12 сек	68 МБ

Помимо метрик качества рекомендаций, важными критериями для выбора модели являются время обучения и скорость инференса. Анализ производительности моделей, представленный в таблице 2, выявил прямую зависимость между сложностью модели и временем обучения (от менее 1 минуты для модели популярности до 7 часов для BERT4Rec), при этом время инференса также увеличивается с ростом сложности модели, но в меньшей степени (от менее 1 секунды для простых моделей до 12 секунд для BERT4Rec), а модель BERT4Rec, несмотря на большое количество параметров, имеет относительно компактный размер (68 МБ) благодаря использованию разделяемых параметров в слоях трансформера.

На основе проведенного анализа можно сформулировать практические рекомендации: базовая модель Popular подходит для быстрого прототипирования и систем с минимальными вычислительными ресурсами; модель ALS эффективна при работе с большими объемами данных и ограниченными вычислительными ресурсами, но имеет ограничения при необходимости учитывать временные зависимости; гибридная модель LightFM предоставляет отличное решение для задач с холодным стартом и подходит для систем, где регулярно появляются новые пользователи и объекты; модель BERT4Rec обеспечивает наивысшую точность рекомендаций и особенно эффективна в задачах с временными зависимостями, подходя для продакшн-систем с высокими требованиями к качеству при наличии достаточных вычислительных ресурсов.

Исследование показало важность выбора подходящей модели в зависи-

мости от целей и условий задачи, таких как доступность данных, требования к вычислительным ресурсам и точности предсказаний. При этом более сложные модели, использующие дополнительную информацию о пользователях и объектах, а также учитывающие последовательный характер взаимодействий, показывают лучшие результаты по метрикам точности рекомендаций, но требуют больших вычислительных затрат.

ЗАКЛЮЧЕНИЕ

В данной работе было проведено сравнительное исследование различных подходов к построению рекомендательных систем для фильмов и сериалов: методов матричной факторизации (ALS), гибридных моделей (LightFM) и моделей на основе трансформеров (BERT4Rec). Для комплексной оценки эффективности этих подходов был использован датасет KION, отражающий реальные взаимодействия пользователей с контентом на одноимённой стриминговой платформе.

На основе проведенного исследования можно сформулировать следующие ключевые выводы:

1. **Превосходство моделей на основе трансформеров.** Модель BERT4Rec продемонстрировала наилучшие результаты по всем метрикам точности ($\text{Recall@10} = 0.1870$, $\text{HitRate@10} = 0.3111$, $\text{MAP@10} = 0.0871$) и разнообразия рекомендаций ($\text{Serendipity@10} = 1.88\text{e-}04$). Это подтверждает высокую эффективность архитектуры трансформеров в задаче рекомендации фильмов благодаря способности улавливать сложные паттерны и долгосрочные зависимости в последовательностях взаимодействий пользователей.
2. **Эффективность гибридных подходов.** Модель LightFM, особенно при увеличении числа эпох обучения, показала хорошие результаты как по метрикам точности, так и по разнообразию рекомендаций. Это подтверждает ценность использования дополнительной информации о пользователях и объектах для повышения качества рекомендаций, особенно в условиях холодного старта.
3. **Неожиданная эффективность модели популярности.** В исследовании модель популярности показала относительно высокие результаты по метрикам точности, превзойдя ALS и приблизившись к показателям LightFM. Это может быть связано с особенностями датасета (короткие последовательности взаимодействий, сильные временные тренды популярности) и указывает на важность учета специфики предметной области при выборе модели рекомендаций.
4. **Компромисс между качеством рекомендаций и вычислительной сложностью.** Наблюдается прямая зависимость между сложностью модели и её эффективностью: более сложные модели (BERT4Rec, LightFM) показыва-

ют лучшие результаты, но требуют значительно больших вычислительных ресурсов для обучения и инференса. Это создает необходимость поиска оптимального баланса в зависимости от конкретных требований и ограничений.

5. **Важность комплексной оценки качества рекомендаций.** Использование различных метрик, оценивающих не только точность, но и разнообразие рекомендаций, их новизну, покрытие пользователей и популярность рекомендуемых объектов, позволило получить более полное представление о преимуществах и недостатках различных подходов. Это подтверждает необходимость многоаспектной оценки рекомендательных систем, выходящей за рамки традиционных метрик точности.

Новизна исследования заключается в комплексном сравнении трех принципиально различных подходов к построению рекомендательных систем: классических методов матричной факторизации, гибридных моделей и современных подходов на основе трансформеров, в то время как большинство существующих работ фокусируются на сравнении моделей в рамках одного подхода либо на сравнении двух смежных подходов. Помимо этого, в работе была предложена методика оценки рекомендательных систем, учитывающая не только точность предсказаний, но и такие важные характеристики, как разнообразие рекомендаций (метрика Serendipity), покрытие пользователей (метрика CoveredUsers) и популярность рекомендуемых объектов (метрика AvgRecPopularity), что даёт более полную картину эффективности различных подходов. Впервые для данного типа систем исследовано влияние структуры данных и особенностей киноиндустрии как предметной области на эффективность различных типов моделей, что расширяет понимание преимуществ и ограничений каждого подхода в конкретном контексте. Существенным вкладом является также проведенное исследование масштабируемости и вычислительной эффективности рассматриваемых методов, что особенно актуально в контексте растущих объемов данных и необходимости обработки рекомендаций в режиме реального времени.

Практическая значимость исследования определяется возможностью использования его результатов разработчиками стриминговых сервисов и других рекомендательных систем при выборе и внедрении алгоритмов персонализированных рекомендаций. Разработанные модели и подходы к их настройке применимы для создания рекомендательных систем в реальных коммерческих

проектах, позволяя оптимизировать как качество рекомендаций, так и вычислительные затраты в зависимости от бизнес-требований и технических ограничений. Предложенная методика комплексной оценки может использоваться для мониторинга и сравнения эффективности различных алгоритмов в промышленных системах, а выявленные особенности работы с данными о пользовательских предпочтениях могут быть учтены при проектировании систем сбора и анализа данных. Анализ вычислительной эффективности и масштабируемости различных подходов дает ценную информацию для оптимизации инфраструктуры рекомендательных систем, что критически важно для сервисов с большим количеством пользователей и обширным каталогом контента. Разработанные программные компоненты и экспериментальный код могут быть использованы как в образовательных целях при изучении методов машинного обучения, так и в качестве основы для дальнейших исследований в этой области.

Исследование также открывает несколько перспективных направлений для дальнейшего развития рекомендательных систем для фильмов и сериалов:

1. **Оптимизация архитектур и гиперпараметров.** Дальнейшее совершенствование архитектур моделей и оптимизация их гиперпараметров с учетом особенностей предметной области может привести к значительному повышению эффективности рекомендаций. Особенно перспективным представляется адаптация архитектуры BERT4Rec для специфики рекомендации фильмов и сериалов.
2. **Исследование новых типов трансформеров.** Применение более современных вариантов трансформеров, таких как XLNet, T5, Longformer, может открыть новые возможности для моделирования сложных зависимостей в данных и повышения качества рекомендаций.
3. **Разработка мультимодальных подходов.** Интеграция текстовой, визуальной и аудиальной информации в рекомендательные системы представляет собой перспективное направление, которое может значительно обогатить представление о контенте и предпочтениях пользователей, особенно в области кино и сериалов, где визуальная и аудиальная составляющие играют важную роль.
4. **Развитие контекстуально-зависимых рекомендаций.** Учет широкого контекста взаимодействия пользователя с системой, включая временные паттерны, ситуационный контекст, настроение и цели просмотра, может

значительно повысить релевантность и своевременность рекомендаций.

5. **Разработка эффективных стратегий ансамблирования.** Комбинирование различных моделей с учетом их сильных сторон и контекста использования может обеспечить оптимальный баланс между различными аспектами качества рекомендаций при разумных вычислительных затратах.

В целом, проведенное исследование подтверждает высокий потенциал современных методов машинного обучения, особенно архитектур на основе трансформеров, для задачи рекомендации фильмов и сериалов. Дальнейшее развитие этих методов и их адаптация к специфике предметной области могут значительно повысить качество персонализированных рекомендаций, улучшая пользовательский опыт и способствуя более эффективному использованию каталога контента на стриминговых платформах.