МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

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

Кафедра дифференциальных уравнений и математической экономики

Исследование и разработка методов построения гибридной рекомендательной системы интернет-магазина

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 441 группы

направление 09.03.03 — Прикладная информатика

механико-математического факультета

Белова Владимира Олеговича

Научный руководитель
профессор, д.э.н. профессор _____ В.А. Балаш

Заведующий кафедрой
зав. кафедрой., д.ф.-м.н., доцент _____ В.С. Рыхлов

Саратов 2026

Введение. Актуальность темы обусловлена ростом ассортимента интернет-магазинов, который существенно усложняет выбор товаров для пользователя. Крупные торговые платформы содержат десятки и сотни тысяч товарных позиций, поэтому даже при наличии поиска и фильтров найти действительно подходящий вариант становится все труднее. В этих условиях рекомендательные системы выступают важным инструментом персонализации, позволяющим анализировать поведение пользователей и формировать релевантные товарные предложения.

В интернет-магазинах обычно отсутствуют явные оценки товаров, поэтому основным источником информации служит неявная обратная связь: просмотры карточек, добавления в корзину и покупки. Такие сигналы информативны, но неоднозначны, что требует применения специальных методов их обработки. Дополнительная сложность состоит в том, что на практике необходимо учитывать не только пользовательскую релевантность, но и прикладные ограничения платформы, прежде всего доступность товаров и их коммерческую значимость. Это придает задаче рекомендаций многокритериальный характер и обосновывает использование гибридных подходов.

Целью работы является исследование и программная реализация технологий обработки данных и машинного обучения для построения гибридной рекомендательной системы интернет-магазина на основе неявной обратной связи пользователей с учетом бизнес-ограничений.

Для достижения поставленной цели были поставлены следующие задачи:

1. проанализировать подходы к построению рекомендательных систем и особенности неявной обратной связи;
2. разработать гибридную модель товарных рекомендаций с учетом бизнес-ограничений;
3. описать алгоритм программной обработки данных, обучения моделей и формирования рекомендаций;
4. провести эксперимент на реальных данных и представить результаты обработки в табличной форме.

Объект исследования. Процесс формирования персонализированных товарных рекомендаций в интернет-магазинах.

Предмет исследования. Методы построения гибридных рекомендательных систем, использующих неявную обратную связь пользователей и учитывающих бизнес-ограничения при ранжировании товаров.

Материалы исследования. Эмпирическую основу исследования составили открытый набор данных *RetailRocket*, содержащий журнал пользовательских взаимодействий с товарным каталогом интернет-магазина, и файл свойств товаров. Исходный набор включает 2 756 101 событие, 1 407 580 пользователей и 235 061 товар. В работе анализировались события трех типов: просмотр карточки товара (*view*), добавление товара в корзину (*addtocart*) и покупка (*transaction*).

Программные технологии. Экспериментальная часть реализована на языке *Python* с использованием библиотек *pandas*, *NumPy*, *SciPy* и *implicit*. Программный конвейер выполняет загрузку данных, фильтрацию, построение разреженных матриц, обучение моделей, применение бизнес-правил, расчет метрик и формирование итогового отчета.

Структура работы. Работа включает введение, четыре раздела, заключение, список использованных источников и приложения. Первый раздел посвящен теоретическим основам построения рекомендательных систем, второй — математической модели гибридной рекомендательной системы, третий — описанию данных и организации эксперимента, четвертый — анализу результатов тестирования базовых и гибридных моделей.

Основное содержание работы. В первом разделе рассматриваются теоретические основы построения рекомендательных систем и обосновывается выбор направления исследования. Показано, что рекомендательная система является механизмом персонализации, предназначенным для автоматизированного отбора объектов, наиболее релевантных конкретному пользователю. Для интернет-магазина она выполняет не только информационную, но и прикладную функцию, влияя на качество поиска товаров и общую полезность взаимодействия пользователя с каталогом.

В разделе анализируются контентно-ориентированные методы, коллаборативная фильтрация и гибридные модели. Сделан вывод, что для задач электронной коммерции наиболее перспективен гибридный подход, поскольку он позво-

ляет сочетать устойчивость к разреженности данных, качество ранжирования и учет прикладных ограничений платформы.

Показано, что ни один из базовых подходов в отдельности не решает задачу интернет-магазина полностью. Контентно-ориентированные методы зависят от качества описания товаров, коллаборативная фильтрация чувствительна к разреженности взаимодействий, а простые эвристические правила не обеспечивают достаточной персонализации. Поэтому в работе делается акцент на гибридных моделях, способных объединять несколько источников информации о предпочтениях пользователя и условиях показа товара.

Отдельное внимание уделено специфике электронной коммерции и неявной обратной связи. Показано, что в интернет-магазине рекомендации должны учитывать не только пользовательский интерес, но и такие факторы, как наличие товара и его коммерческая значимость. Рассмотрены метрики *Precision@K*, *Recall@K*, *NDCG@K*, а также бизнес-ориентированные показатели. По итогам раздела сделан вывод о необходимости построения гибридной модели, объединяющей пользовательскую релевантность и бизнес-правила.

Дополнительно подчеркивается, что для электронной коммерции особенно важен компромисс между интересами пользователя и интересами платформы. Чрезмерное смещение выдачи в сторону коммерчески приоритетных товаров может ухудшать пользовательский опыт, тогда как полное игнорирование бизнес-факторов снижает практическую ценность рекомендаций. Этот вывод задает общую исследовательскую логику последующих разделов.

Во втором разделе формализуется задача построения рекомендаций при неявной обратной связи и разрабатывается математическая модель гибридной рекомендательной системы интернет-магазина. Пользовательские взаимодействия интерпретируются как неявные сигналы интереса, для которых вводятся матрицы предпочтений и доверия.

В качестве базовых алгоритмов рассматриваются *Implicit ALS* и *BPR*. Показано, что *Implicit ALS* хорошо работает с взвешенными неявными взаимодействиями и разреженными данными, тогда как *BPR* ориентирована на корректное упорядочивание объектов в верхней части рекомендательного списка.

В работе специально подчеркивается различие ролей этих моделей. *Implicit ALS* обеспечивает устойчивую базу для восстановления скрытой структуры

предпочтений на основе матричной факторизации, тогда как *BPR* позволяет учитывать попарный характер пользовательского выбора и лучше отражает задачу ранжирования. Такое разделение функций делает их совместное использование методологически обоснованным.

На этой основе предлагается гибридная схема, в которой оценки *Implicit ALS* и *BPR* предварительно нормируются и объединяются линейно с коэффициентами λ_1 и λ_2. Дополнительно в модель включаются бизнес-ограничения: жесткий фильтр недопустимых товаров и мягкая бизнес-коррекция, регулируемая коэффициентом μ. Тем самым модель позволяет управлять компромиссом между пользовательской релевантностью и прикладной полезностью выдачи.

Отдельно анализируется интерпретация параметров модели. Коэффициенты λ_1 и λ_2 задают баланс между устойчивостью факторизационного подхода и чувствительностью попарного ранжирования к верхней части списка, а параметр μ управляет силой бизнес-влияния. Благодаря этому математическая постановка позволяет исследовать не единственную фиксированную конфигурацию, а семейство режимов ранжирования с разным соотношением пользовательской и прикладной направленности.

С практической точки зрения разработанная модель может использоваться в нескольких режимах. При минимальном влиянии бизнес-компоненты система работает как персонализированный рекомендатель с ограничением по допустимости товаров. При умеренных значениях μ модель сохраняет ориентацию на интересы пользователя, но позволяет аккуратно усиливать коммерчески значимые позиции. При больших значениях коэффициента бизнес-влияния система становится инструментом управляемого переупорядочивания выдачи в соответствии с операционными целями платформы.

По итогам раздела делается вывод о том, что предложенная математическая схема пригодна для анализа вклада отдельных алгоритмических компонент и для последующей экспериментальной проверки на реальных данных.

В третьем разделе описываются исходные данные, процедура предобработки и организация вычислительного эксперимента. В качестве экспериментальной базы используется открытый набор данных *RetailRocket*, содержащий журнал пользовательских взаимодействий с товарным каталогом интернет-

магазина. Набор охватывает период с 3 мая 2015 года по 18 сентября 2015 года и включает 2 756 101 событие, 1 407 580 пользователей и 235 061 товар.

В работе рассматриваются события трех типов: *view*, *addtocart* и *transaction*. Для них задаются веса 1, 3 и 10 соответственно, после чего агрегированная оценка взаимодействия сглаживается логарифмическим преобразованием. На этапе предобработки исключаются пользователи и товары с редкими взаимодействиями; после фильтрации в эксперименте остаются 23 138 пользователей, 39 653 товара и 264 721 ненулевое взаимодействие.

Такое преобразование данных позволяет привести разнородные пользовательские действия к единой шкале силы сигнала и уменьшить влияние случайных или повторяющихся просмотров. Фильтрация по минимальному числу взаимодействий направлена на удаление наименее информативных наблюдений и на частичное снижение разреженности матрицы «пользователь — товар», что особенно важно для алгоритмов коллаборативной фильтрации.

Данные разделяются по времени в пропорции 80/20 на обучающую и тестовую части. Для построения бизнес-составляющей используются признак доступности товара и косвенный показатель коммерческой значимости, основанный на частоте транзакций в обучающей выборке. Все этапы вычислительного конвейера реализованы программно, что обеспечивает воспроизводимость результатов.

Постановка программной задачи заключается в автоматизированном преобразовании исходного журнала событий в обучающие и тестовые структуры, пригодные для построения рекомендаций. Алгоритм программных действий включает загрузку файлов, фильтрацию редких пользователей и товаров, агрегирование неявных событий, построение разреженных матриц, обучение базовых и гибридной моделей, применение бизнес-правил и расчет метрик. В автореферате результаты обработки представлены в табличной и графической форме.

На рисунке значения метрик нормированы: значение каждой метрики при $\mu = 0$ принято за 100 %.

Выбор временного разбиения обусловлен прикладной постановкой задачи: модель должна строиться по историческим данным и затем использоваться для предсказания будущих предпочтений. В разделе также подчеркивается, что от-

Таблица 1 — Результаты программной обработки данных и тестирования моделей

Показатель	Значение
Пользователей после фильтрации	23 138
Товаров после фильтрации	39 653
Ненулевых взаимодействий user–item	264 721
Лучшая базовая модель	*Implicit ALS*: *NDCG@10* = 0,01081
Лучшая гибридная модель без бизнес-коррекции	*NDCG@10* = 0,01121
Гибридная модель с бизнес-ограничениями	*Availability@10* = 1,00000; *BusinessValue@10* = 0,43180

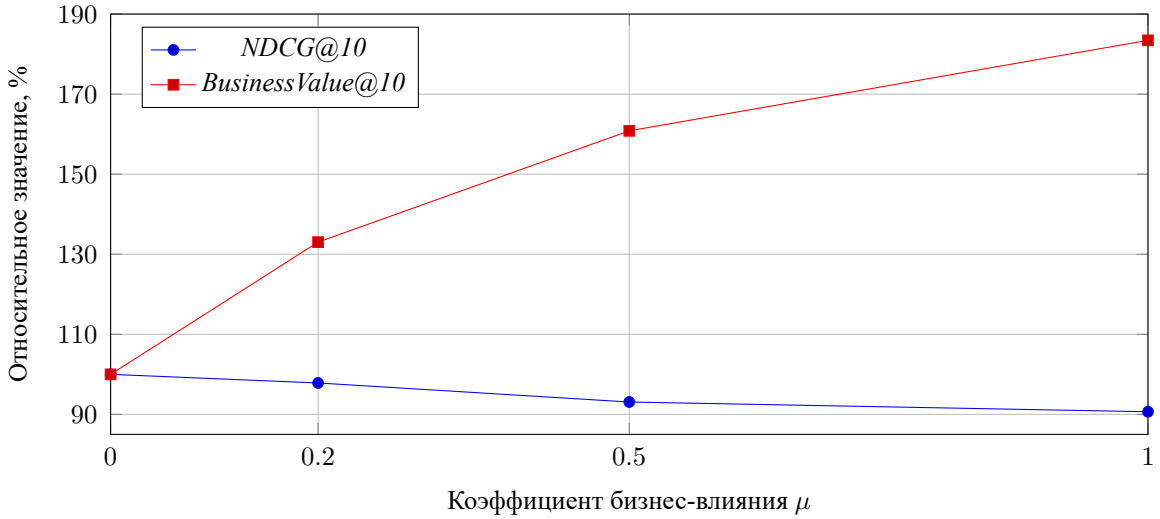

Рисунок 1 — Динамика пользовательской и бизнес-ориентированной метрик при изменении коэффициента μ

крытый набор данных не содержит прямых финансовых показателей, поэтому бизнес-составляющая ранжирования формируется на основе доступных признаков и не претендует на прямое измерение денежного эффекта.

В разделе фиксируются сравниваемые модели и метрики оценки. Рассматриваются популярностная стратегия, *Implicit ALS*, *BPR*, гибридная модель без бизнес-коррекции и гибридная модель с бизнес-правилами. Для анализа качества используются *Precision@10*, *Recall@10*, *NDCG@10*, а также *Availability@10*, *BusinessValue@10* и *Coverage@10*.

Особое внимание уделяется метрике *NDCG@10*, поскольку именно она наиболее чувствительна к порядку расположения релевантных товаров в верхней части списка. Для прикладного сценария интернет-магазина это принципиально важно, так как пользователь, как правило, взаимодействует прежде всего с

первыми позициями рекомендательной выдачи. Включение бизнес-метрик позволяет дополнить оценку персонализации анализом практической пригодности сформированного списка.

В четвертом разделе проводится анализ результатов экспериментального исследования. На первом этапе сопоставляются базовые модели: популярностная стратегия, *Implicit ALS* и *BPR*. Лучшей базовой моделью оказывается *Implicit ALS*, для которой получены значения *Precision@10* = 0,00620, *Recall@10* = 0,01181 и *NDCG@10* = 0,01081. Тем самым подтверждается эффективность факторизационного подхода для работы с неявной обратной связью.

Гибридная модель без бизнес-коррекции показывает *Precision@10* = 0,00620, *Recall@10* = 0,01115 и *NDCG@10* = 0,01121. По сравнению с *Implicit ALS* она почти не меняет *Precision@10*, немного уступает по *Recall@10*, но улучшает *NDCG@10* примерно на 3,7%. Это означает, что гибридизация повышает качество упорядочивания релевантных объектов в верхней части списка.

Отдельно отмечается, что самостоятельные результаты *BPR* в проведенном эксперименте оказались слабее результатов *Implicit ALS*. Однако включение *BPR* в состав гибридной схемы все же оказалось полезным, поскольку добавило информацию о порядке предпочтений и улучшило качество ранжирования верхних позиций. Тем самым подтверждается, что ценность алгоритма в составе гибрида может отличаться от его самостоятельных показателей.

Далее исследуется влияние бизнес-ограничений. Для конфигурации с $\mu = 0,5$ гибридная модель с бизнес-правилами демонстрирует *NDCG@10* = 0,00997, *Availability@10* = 1,0, *BusinessValue@10* = 0,43180 и *Coverage@10* = 0,02764. По сравнению с гибридной моделью без бизнес-коррекции это означает полное исключение явно недоступных товаров из выдачи и рост средней коммерческой значимости списка ценой снижения пользовательского качества и покрытия каталога.

Тем самым в работе количественно подтверждается наличие компромисса между пользовательской релевантностью и прикладной полезностью рекомендаций. При усилении бизнес-компоненты рекомендательный список становится более пригодным для практического использования, но одновременно теряет часть персонализационного качества. Такой результат интерпретируется как

ожидаемое следствие многокритериального характера задачи, а не как недостаток самой модели.

Для оценки чувствительности анализируется коэффициент μ. Показано, что среди конфигураций с обязательной фильтрацией по доступности товаров наилучшее значение *NDCG@10* достигается при $\mu = 0,0$, а конфигурация $\mu = 0,2$ может рассматриваться как разумный компромисс, поскольку уменьшает *NDCG@10* примерно на 2,1% по сравнению с $\mu = 0,0$, но увеличивает *BusinessValue@10* примерно на 33,0%.

В разделе также отмечаются ограничения исследования: отсутствие прямых финансовых характеристик товаров в наборе *RetailRocket*, зависимость результатов от выбранных гиперпараметров и офлайн-характер оценки. Несмотря на это, эксперимент обладает практической ценностью, поскольку демонстрирует работоспособность предложенной схемы и позволяет осмысленно выбирать режим ранжирования в зависимости от целей платформы.

Полученные результаты интерпретируются не как поиск единственной универсально лучшей конфигурации, а как построение набора управляемых режимов работы рекомендательной системы. Если приоритетом является максимальное качество ранжирования верхних позиций, предпочтительной оказывается гибридная модель без бизнес-коррекции. Если же необходимо одновременно исключать недоступные товары и умеренно учитывать коммерческую значимость, наиболее сбалансированной представляется конфигурация с $\mu = 0,2$.

По итогам раздела сделан вывод о том, что гибридная рекомендательная система, основанная на неявной обратной связи и дополненная бизнес-правилами, является практически значимым подходом для задач электронной коммерции и позволяет управляемо регулировать баланс между пользовательской релевантностью и прикладной полезностью рекомендаций.

Заключение. Выпускная квалификационная работа посвящена исследованию и программной реализации технологий построения гибридной рекомендательной системы интернет-магазина на основе неявной обратной связи с учетом бизнес-ограничений. Показано, что для электронной коммерции задача реко-

мендаций не сводится к максимизации пользовательской релевантности и требует учета прикладных ограничений платформы.

В ходе выполнения работы проведен анализ основных подходов к построению рекомендательных систем, формализована математическая модель, объединяющая *Implicit ALS*, *BPR*, фильтрацию по доступности товаров и бизнес-коррекцию ранжирования, а также разработана воспроизводимая схема вычислительного эксперимента на открытом наборе данных *RetailRocket*.

Эксперимент показал, что среди базовых моделей наилучшие результаты демонстрирует *Implicit ALS*, а гибридизация с *BPR* позволяет улучшить качество ранжирования верхней части списка по *NDCG@10*. Учет бизнес-ограничений повышает доступность рекомендованного списка и его среднюю коммерческую значимость, но сопровождается снижением части пользовательских метрик. Следовательно, предложенная модель реализует управляемый компромисс между пользовательской релевантностью и прикладной полезностью выдачи.

Практическая значимость работы заключается в возможности использовать предложенную схему как основу для проектирования рекомендательных модулей интернет-магазинов. Перспективы дальнейшего развития связаны с использованием реальных бизнес-признаков товаров, более полной настройкой гиперпараметров, учетом временной динамики пользовательских предпочтений и проведением онлайн-экспериментов.

Полученные результаты подтверждают, что построение рекомендательной системы для электронной коммерции должно рассматриваться как многокритериальная задача. Предложенный подход показывает, что совместное использование моделей, ориентированных на неявную обратную связь, и явное включение бизнес-правил позволяет не только повысить качество ранжирования, но и сделать выдачу более пригодной для реального применения в интернет-магазине.