

## ОРИГИНАЛЬНАЯ СТАТЬЯ



DOI: 10.26794/2304-022X-2026-16-2-124-135  
УДК 004.9(045)  
JEL M31, C88

## Рекомендательные системы: методологические вызовы и проблемы

Р.С. Кульшин, А.А. Сидоров

Томский государственный университет систем управления и радиоэлектроники, Томск, Российская Федерация

### АННОТАЦИЯ

**Целью** исследования явился анализ эволюции методов построения рекомендательных систем и подходов к оценке их качества в условиях многокритериальности и изменяющихся пользовательских предпочтений. В ходе работы рассмотрены основные этапы развития рекомендательных технологий, систематизированы современные алгоритмы и архитектуры, включая гибридные, нейросетевые и графовые модели. Особое внимание уделено методам оценки качества рекомендаций и классификации метрик, отражающих точность, уровень ранжирования, разнообразие и ресурсную эффективность по группам. В статье также показано, что ориентации исключительно на метрики точности для адекватной оценки результативности рекомендательных систем недостаточно. Авторы исследования рассмотрели ключевые методологические проблемы, включая холодный старт, разреженность данных и необходимость баланса между точностью и разнообразием рекомендаций, и сделали вывод о целесообразности комплексного многокритериального подхода к оценке рекомендательных систем. Результаты исследования могут быть использованы научными работниками, разработчиками рекомендательных систем, специалистами в области машинного обучения и анализа данных, а также представителями бизнеса при выборе и оценке эффективности рекомендательных алгоритмов.

**Ключевые слова:** рекомендательные системы; оценка качества рекомендаций; тенденции развития; персонализация; модели; алгоритмы; метрики

**Для цитирования:** Кульшин Р.С., Сидоров А.А. Рекомендательные системы: методологические вызовы и проблемы. *Управленческие науки = Management Sciences*. 2026;16(2):124-135. DOI: 10.26794/2304-022X-2026-16-2-124-135

## ORIGINAL PAPER

## Recommender Systems: Methodological Challenges and Issues

R.S. Kulshin, A.A. Sidorov

Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation

### ABSTRACT

The **aim** of this study is to analyse the evolution of methods for building recommender systems and approaches to evaluating their performance under conditions of multi-criteria decision-making and dynamically changing user preferences. The paper examines the key stages in the development of recommender technologies and systematises modern algorithms and architectures, including hybrid, neural network-based, and graph-based models. Particular attention is paid to methods for evaluating recommendation quality and to the classification of metrics reflecting accuracy, ranking quality, diversity, and computational efficiency. The study demonstrates that reliance solely on accuracy metrics is insufficient for an adequate assessment of recommender system performance. The authors address key methodological challenges, including the cold-start problem, data sparsity, and the need to balance accuracy and diversity in recommendations. The paper concludes that a comprehensive multi-criteria approach is essential for the effective evaluation of recommender systems. The findings may be useful for researchers, recommender system developers, and specialists in machine learning and data analysis, as well as for business practitioners involved in selecting and evaluating the effectiveness of recommendation algorithms.

**Keywords:** recommender systems; recommendation quality evaluation; development trends; personalization; models; algorithms; metrics

**For citation:** Kulshin R.S., Sidorov A.A. Recommender systems: Methodological challenges and issues. *Upravlencheskie nauki = Management Sciences*. 2026;16(2):124-135. DOI: 10.26794/2304-022X-2026-16-2-124-135

© Кульшин Р.С., Сидоров А.А., 2026

## ВВЕДЕНИЕ

Современное информационное общество характеризуется не только стремительным ростом объемов доступных данных, но и необходимостью их осмысленной фильтрации и персонализации. В условиях перенасыщения коммуникационно-информационного пространства потребители сталкиваются с трудностями выбора среди множества альтернатив, будь то товары, мультимедийный контент, новости или образовательные курсы. Эти обстоятельства способствовали становлению рекомендательных систем — интеллектуальных технологий, предназначенных для формирования кастомизированных предложений на основе анализа пользовательских данных.

Рекомендательные системы представляют собой совокупность алгоритмов и моделей, способных анализировать поведение и предпочтения потребителей с целью прогнозирования их потенциальных интересов. На сегодняшний день подобные технологии, трансформирующие способы взаимодействия между пользователями и контентом, стали неотъемлемой частью цифровых платформ. Их функционирование основывается на различных подходах, ориентированных в том числе на работу с контентом и алгоритмами сходства пользователей, а также гибридных моделях, сочетающих в себе сильные стороны обоих направлений.

Область применения рекомендательных систем чрезвычайно широка. Они активно используются в электронной коммерции для персонализации предложений; в потоковых сервисах — для формирования релевантных подборок контента; в социальных сетях — для оптимизации взаимодействий между потребителями. Благодаря внедрению таких систем появляется возможность адаптации цифровых продуктов к нуждам конкретного клиента, что в значительной степени повышает вовлеченность и удовлетворенность аудитории.

На макроуровне рекомендательные технологии оказывают заметное влияние на развитие цифровой экономики. Они способствуют росту эффективности платформенных решений, улучшая пользовательский опыт и, как следствие, становятся важным фактором конкурентоспособности в digital-пространстве. В условиях глобальной цифровизации и персонифицированного потребления роль рекомендательных систем выходит за рамки вспомогательной и становится стратегической как с точки зрения бизнеса, так и с позиции формирования новых моделей пользовательского поведения.

В связи с растущей значимостью таких систем актуальной становится задача их всестороннего изучения. Целью настоящей работы является анализ эволюции принципов построения рекомендательных систем, а также методов их оценки в условиях многокритериальности и динамического изменения пользовательских предпочтений. Для ее достижения авторами исследования рассмотрены основные этапы развития рекомендательных технологий, а также систематизированы современные алгоритмы и архитектурные подходы, а также выделены основные группы метрик, применяемых для оценки качества рекомендаций. Особое внимание уделено обсуждению методологических вызовов и проблем, стоящих перед разработчиками и исследователями в данной области.

## РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ: ЗАРОЖДЕНИЕ ТЕХНОЛОГИИ

Совершенствование рекомендательных систем тесно переплетено с эволюцией информационных технологий и ростом значимости персонализированного взаимодействия в цифровой среде. По мере стремительного увеличения объемов доступных данных и диверсификации пользовательских потребностей стало очевидно, что традиционные подходы к поиску и фильтрации информации, доминировавшие в ранние периоды развития сетевых технологий, уже не в полной мере удовлетворяют задачам эффективного предоставления релевантного контента.

На первых этапах персонализация рассматривалась как частный случай общей проблемы поиска данных и обработки запросов. Уже в конце 1980-х — начале 1990-х гг. начали формироваться теоретические и практические предпосылки для выделения рекомендательных систем в самостоятельную область исследований, характеризующуюся собственным понятийным аппаратом, методами и задачами.

Одной из первых известных разработок, ориентированных на персональные интересы клиентов, стала система Tapestry [1], представляющая собой электронную доску объявлений, где пользователи могли вручную создавать фильтры для поиска интересных материалов, основываясь на отзывах других участников. Несмотря на примитивность методов, Tapestry заложила фундамент для будущих алгоритмов.

К середине 1990-х гг. стало очевидно, что наилучшие рекомендации можно формировать, опи-

раясь не только на индивидуальные запросы потребителей, но и на существующие поведенческие паттерны. Именно в этот период возникает метод коллаборативной фильтрации, ставший одним из столпов в истории рекомендательных систем. Так, в 1994 г. появился проект GroupLens — исследовательская инициатива, направленная на создание технологий автоматической фильтрации информации на основе пользовательских оценок. Принцип, разработанный GroupLens [2], применялся в новостном сервисе Usenet: клиенты могли рейтинговать прочитанные статьи, а алгоритм предлагал новости, оцененные положительно другими пользователями со схожими вкусами. GroupLens стала одной из первых компаний, реализовавших автоматическую коллаборативную фильтрацию, где использовалось сходство между потребителями для прогнозирования их интересов. Этот проект оказал наиболее существенное влияние на все последующие разработки в области рекомендательных систем: он не только подтвердил практическую применимость данного подхода, но и заложил основы современных архитектурных решений. Именно в рамках этой инициативы началась систематизация методов оценки качества рекомендаций, что сыграло важную роль в формировании методологии последующих исследований.

С начала 2000-х гг. рекомендательные системы стали активно внедряться в инфраструктуру интернет-коммерции [3]. Так, крупнейший онлайн-ритейлер Amazon перешел к использованию их алгоритмов для увеличения продаж за счет персонализированных предложений.

Одним из важнейших событий, способствовавших популяризации и увеличению числа научных работ в этой области, стал международный конкурс 2006 г. Netflix Prize [4], когда компания Netflix предложила исследовательскому сообществу задачу по улучшению точности ее системы рекомендаций. Конкурс продлился до 2009 г. и привел к появлению целого ряда инновационных решений, в том числе алгоритмов, основанных на матричном разложении [5]. Netflix Prize оказал значительное влияние на развитие академического и прикладного интереса к данной сфере, продемонстрировав, что комплексные математические модели могут обеспечивать качественный прирост даже с учетом высоких требований к точности и масштабируемости. После завершения конкурса внимание ученых и практиков сместилось в сто-

рону построения гибридных моделей [6], сочетающих в себе как коллаборативные, так и контентные фильтры, основанные на схожести данных объектов рекомендаций. Параллельно прогрессировали методы, базирующиеся на машинном обучении (включая деревья решений, градиентный бустинг, факторизационные машины и позже — нейросетевые архитектуры).

Историческая траектория развития рекомендательных систем представляет собой переход от эвристических к высокоуровневым моделям, способным успешно обрабатывать большие объемы данных и адаптироваться к разнообразным пользовательским сценариям. Этот эволюционный путь подводит к ключевому вопросу — как оценивать эффективность столь разнородных и сложных систем в условиях многокритериальности и изменчивости пользовательских предпочтений.

## ЛАНДШАФТНОЕ ПРОСТРАНСТВО РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ

Современные рекомендательные системы представлены широким спектром методов и архитектур, различающихся используемыми данными, алгоритмами, способами представления знаний и степенью персонализации. При этом общепринятой основой классификации, на которой основаны рекомендации, служит природа информации.

**Коллаборативная фильтрация** (Collaborative Filtering, CF) [7] базируется на предположении, что пользователи, проявлявшие схожие интересы в прошлом, с высокой вероятностью сохранят их в будущем. Такие системы оперируют исключительно историей взаимодействия — оценками, кликами, покупками — и не требуют сведений о самих объектах.

Существует два основных класса коллаборативной фильтрации: в случае User-based CF: рекомендации строятся на основе сходства между пользователями, а при Item-based CF — между объектами, с которыми взаимодействовали потребители.

Для определения сходства применяются метрики в виде коэффициента корреляции Пирсона и значения косинусного сходства. Коллаборативные методы обладают способностью выявлять скрытые закономерности и учитывать коллективное поведение, однако страдают от проблемы «холодного старта» и масштабируемости в условиях больших данных.

Классическим примером Item-based подхода является Amazon, внедривший в 2003 г. алгоритм рекомендаций, основанный на схожести товаров [8].

Другим знаковым кейсом стала система Cinematch компании Netflix, применявшая User-based технологию. Во время конкурса Netflix Prize было показано, что матричная факторизация и ансамбли моделей (заложившие основу современных CF-алгоритмов) существенно превосходят методы ближайших соседей по точности рекомендаций [5].

**Контентные рекомендательные системы** (Content-Based Filtering, CBF) [9] делают акцент на характеристиках объектов и предпочтениях пользователя, выраженных через взаимодействие с теми или иными типами контента. Рекомендации формируются путем сопоставления профилей потребителя и объекта, которое осуществляется посредством различных методов машинного обучения (включая классификацию, кластеризацию, логистическую регрессию и деревья решений). Достоинство контентных подходов — независимость от данных других пользователей — делает их устойчивыми к проблеме разреженности. Однако они подвержены риску ограничения разнообразия и не способны предлагать что-то новое, выходящее за пределы уже известного профиля интересов.

Контентно-ориентированная модель легла в основу рекомендаций Pandora, формируемых на основе сходства музыкальных треков, описываемых через сотни вручную размеченных признаков. Поведение других потребителей не учитывается, что особенно эффективно при разреженных данных [10].

NewsWeeder — одна из первых рекомендательных систем, использующих контентную фильтрацию для подбора новостных статей. Пользователь помечал как интересующие, так и не привлекающие его внимание материалы, а система автоматически анализировала эту информацию (ключевые слова, тематику) и создавала профиль интересов. Рекомендации строились через сравнение контентных признаков новых статей с этим профилем [11].

**Гибридные рекомендательные системы** сочетают элементы коллаборативных и контентных подходов, компенсируя недостатки каждого из них [12]. Интеграция может осуществляться на различных уровнях: на этапе предварительной обработки данных, в ходе объединения моделей или при постобработке результатов.

Наиболее известные стратегии гибридизации включают:

- Взвешенное объединение — система складывает результаты разных моделей с определенными весами. Например, 70% рекомендаций от коллабо-

ративной модели, 30% — от контентной. Вес можно настраивать вручную или автоматически.

- Каскадирование — сначала одна модель (например, коллаборативная) отбирает кандидатов, затем вторая (например, контентная) их пересматривает и уточняет. Это экономит ресурсы: тяжелые модели применяются только к небольшому списку.

- Метамоделирование — одна модель передает свои выходные данные (или скрытые признаки) второй: например, коллаборативная генерирует векторы, а контентная нейросеть использует их как вход.

Netflix активно задействует гибридную архитектуру, в основе которой лежит двухуровневая система ранжирования, позволяющая учитывать поведение, контекст и метаданные объектов. Такой подход обеспечивает высокую вовлеченность пользователей на платформе [13].

Microsoft Academic разработала масштабируемую гибридную рекомендательную систему для предложения научных статей [14], применяя данные Microsoft Academic Graph (включая названия и аннотации публикаций). Это дает возможность эффективно рекомендовать релевантные публикации и решать проблему «холодного старта».

В последнее время на смену традиционным алгоритмам пришли методы, способные обрабатывать сложные зависимости, учитывать контекст взаимодействия и адаптироваться к изменениям в реальном времени.

Гибридные архитектуры, объединившие коллаборативные и контентные технологии, продемонстрировали эффективность, но по мере роста объемов и сложности информации стали очевидны их ограничения: жесткость эвристик, трудоемкость настройки весов и неспособность явно моделировать нелинейные зависимости. Именно эти вызовы и открыли дорогу новым подходам, в рамках которых появилась возможность извлечения сложных признаков, учета слабых зависимостей, гибкого представления взаимосвязей в пользовательско-товарных экосистемах.

## Нейросети

Развитие глубокого обучения<sup>1</sup> стало переломным этапом в истории рекомендательных систем, от-

<sup>1</sup> Глубокое обучение (Deep Learning) — подраздел машинного обучения, основанный на использовании искусственных нейронных сетей с множеством слоёв.

крыв ранее недоступные возможности [15]. Принципиальное отличие нейросетевых моделей заключается в их способности захватывать и обучаться на сложных, нелинейных зависимостях между потребителями и объектами, что особенно важно в условиях высокой разреженности, характерной для большинства практических кейсов.

Одной из первых нейросетевых технологий, примененных в рекомендательных системах, стали автоэнкодеры [16], которые учатся восстанавливать исходные данные, проходя через сжатое внутреннее представление. Это помогает выявить скрытые предпочтения пользователей и характеристики объектов.

С развитием компьютерного зрения особое внимание разработчиков привлекли сверточные нейронные сети, способные извлекать содержательные признаки из изображений и видеоданных [17]. В контексте рекомендаций они используются, например, для анализа визуальных свойств товаров — цвета, формы, стилистики — что особенно актуально в визуально ориентированных областях.

Не менее важной вехой стало применение рекуррентных нейронных сетей в тех случаях, когда поведение пользователя можно описать как последовательность действий. Такие модели продемонстрировали высокую эффективность в сессионных рекомендательных системах [18], позволяя учитывать временной порядок взаимодействий и предпочтения в зависимости от контекста и быстро изменяющихся интересов (в задачах краткосрочного моделирования последних эти модели незаменимы).

Наиболее ярким проявлением прогресса последних лет стали модели на основе трансформеров. Благодаря механизму внимания [19] они способны отслеживать долгосрочные зависимости в массивах данных, масштабироваться на большие выборки и работать с разнородными источниками информации. Их универсальность сделала возможным построение рекомендаций, в которых одновременно анализируется история взаимодействий, семантика объектов, поведенческие сигналы.

Например, YouTube использует двухэтапную нейросетевую архитектуру: одна модель отбирает объекты-кандидаты для рекомендации, а вторая ранжирует их. Обучение производится на миллиардах пользовательских взаимодействий с учетом сложных поведенческих зависимостей [17]. Современные архитектуры на основе трансформеров, такие как BST (Behavior Sequence Transformer) от

Alibaba, позволяют рассматривать как краткосрочные, так и долгосрочные поведенческие зависимости за счет механизма самовнимания [20].

### Графовые модели

С увеличением объемов данных все большее внимание исследователей и разработчиков привлекают модели, которые могут учитывать взаимосвязи между элементами внутри графов. В этом контексте важную роль начали играть графовые нейронные сети (Graph Neural Networks, GNN) [21], объединившие в себе преимущества нейросетевого моделирования и гибкость представления данных в графовой форме.

В отличие от классических моделей (опирающихся преимущественно на явные взаимодействия между пользователями и объектами), GNN позволяют учитывать различные косвенные связи — например, схожие предпочтения потребителей или общие атрибуты, что значительно расширило горизонты персонализации. Применение GNN дает возможность формировать содержательные рекомендации, отражающие не только индивидуальную историю пользователя, но и его положение в глобальной структуре связей. Такой подход способствует более глубокому пониманию структурных отношений, что положительно сказывается на разнообразии и новизне выдаваемых рекомендаций — двух важных аспектах пользовательского опыта. Благодаря использованию информации об атрибутах узлов и их связи с другими элементами графа, GNN способны извлекать полезные данные даже при минимуме взаимодействий. Особенно это проявляется в сложных средах, таких как маркетплейсы, где каждый товар имеет множество признаков и категорий, или в интеллектуальных системах поиска знаний, где важна семантическая связанность объектов.

Pinterest задействует масштабируемую нейросеть с применением эмбединга, отличающуюся высокой точностью рекомендаций на графе с миллиардами узлов. Alibaba использует графовые эмбединги, которые помогли решить проблему «холодного старта» [22].

### КЛЮЧЕВЫЕ ПРОБЛЕМЫ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ

Рекомендательные системы играют важную роль в цифровой экономике, позволяя пользователям ориентироваться во внушительных объемах информации и персонализировать опыт взаимодей-

ствия с контентом, однако их развитие сопровождается рядом серьезных проблем.

Одной из наиболее распространенных является «холодный старт», проявляющийся в сложности формирования рекомендаций для новых пользователей или объектов, сведения о которых отсутствуют. Это обстоятельство тесно связано с разреженностью данных, когда матрицы взаимоотношений потребителей с объектами характеризуются большим количеством пропущенных значений, что затрудняет выявление закономерностей предпочтений, снижает точность рекомендаций [23] и ведет к снижению вовлеченности пользователей на начальной фазе.

Важным вызовом также становится необходимость поиска баланса между точностью рекомендаций и их разнообразием. Системы, ориентированные исключительно на максимизацию точности, часто приводят к ограничению кругозора пользователей и способствуют возникновению «информационного пузыря». Поэтому в последние годы особое внимание уделяется подходам, которые обеспечивают баланс, стимулируя потребителей к изучению нового и интересного контента при сохранении приемлемого уровня релевантности рекомендаций [24]. Если последние излишне предсказуемы или повторяются, пользователь может быстро утратить интерес, ограничить взаимодействие с сервисом (что уменьшит возможности перекрестных продаж и открытия новых категорий продуктов) или переключиться на конкурентов.

Проблема справедливости рекомендаций приобретает особую актуальность в контексте распространения искусственного интеллекта и автоматизированного принятия решений. Систематические искажения и дискриминация отдельных категорий клиентов могут усиливаться алгоритмами, обученными на исторических данных с уже существующими предубеждениями. Исследователи подчеркивают необходимость создания подходов, обеспечивающих как равномерное удовлетворение интересов различных групп потребителей, так и выявление и устранение предвзятости [25]. Дополнительным значимым аспектом выступают низкие прозрачность и интерпретируемость многих современных рекомендательных моделей, особенно основанных на глубоком обучении. Их «черные ящики» усложняют понимание принципов, по которым формируются рекомендации, снижая доверие пользователей и ограничивая возможности

разработчиков по отладке и совершенствованию алгоритмов. Искажения в последних могут мешать равномерному продвижению товаров/контента на платформе, уменьшая мотивацию партнеров работать с сервисом. В этой связи все чаще ставится задача создания объяснимых систем, способных представить пользователям логичные и понятные обоснования рекомендаций [26]. Игнорирование интересов отдельных групп потребителей может привести к репутационным рискам, росту жалоб, уходу целевых аудиторий и даже стать причиной юридических последствий.

Также важно учитывать уязвимость рекомендательных алгоритмов к манипуляциям и атакам со стороны злоумышленников, которые способны существенно исказить требуемые результаты путем внесения фиктивных данных или проведения адверсарийных атак.

Масштабируемость и ресурсные ограничения представляют собой еще одну серьезную проблему. При расширении аудитории сервиса возрастает нагрузка на рекомендательные алгоритмы, которым требуется оперативно обрабатывать большие объемы данных и выполнять вычисления в реальном времени. Для этого используются методы распределенных вычислений и оптимизации производительности алгоритмов, однако неэффективная система может не справляться с нагрузкой, снижая скорость отклика или выдавая устаревшие рекомендации. Это критично для таких сценариев, где время — ключевой фактор удержания внимания пользователя (стриминговые платформы, маркетплейсы).

Для обеспечения качественной работы рекомендательных систем при разработке и внедрении новых алгоритмов требуется учет множества факторов — от точности и разнообразия рекомендаций до вопросов справедливости и интерпретируемости.

## ОЦЕНКА КАЧЕСТВА РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ

Оценка рекомендательных систем представляет собой сложную многокритериальную задачу, выходящую далеко за рамки традиционного понимания качества как способности точно предсказать интересы пользователя. Для более четкого отражения различных аспектов их работы целесообразно выделить четыре основные группы метрик: ресурсы, точность, ранжирование и разнообразие.

**Ресурсоориентированные метрики** связаны с производительностью системы в условиях

ограниченных вычислительных ресурсов: к ним относятся показатели времени отклика, скорости генерации рекомендаций, потребления памяти, необходимости частого переобучения. Эти характеристики приобретают особую значимость в контексте систем реального времени и масштабируемых платформ с миллионами пользователей, где задержка в рекомендациях может напрямую влиять на клиентский опыт и бизнес-результаты. Кроме того, увеличение сложности моделей усиливает напряженность между качеством рекомендаций и затратами на их получение.

**Метрики точности** традиционно составляют основу офлайн-оценки рекомендательных систем, поскольку позволяют количественно определить способность модели предсказывать интересы пользователя на основе уже известных взаимодействий. При классических постановках задач целью рекомендатора является максимально точное восстановление либо фактических оценок (например, рейтингов фильмов), либо вероятности того, что потребитель заинтересуется предложенным объектом. Наиболее очевидными инструментами измерения точности служат регрессионные метрики ошибки, такие как средняя абсолютная MAE (Mean Absolute Error) и среднеквадратичная RMSE (Root Mean Square Error) ошибки. Эти показатели характеризуют среднее отклонение между предсказанными и фактическими значениями рейтингов. Низкие величины MAE или RMSE свидетельствуют о том, что модель (в среднем) допускает небольшие ошибки при прогнозировании оценок, что является важным индикатором в задачах восстановления предпочтений пользователей в количественной форме.

Кроме регрессионных широко применяются дискретные и вероятностные меры, такие как Precision, Recall. Они, в отличие от MAE и RMSE, предназначены для оценки качества бинарных предсказаний: например, вероятности того, что потребитель заинтересуется данным предложением (объектом) или кликнет на него. Precision отражает долю релевантных объектов среди рекомендованных, Recall — долю найденных релевантных среди всех релевантных в системе. Эти метрики актуальны в приложениях, где пользователю предлагается ограниченное количество объектов и требуется максимизировать вероятность релевантных попаданий.

Метрики точности обладают целым рядом достоинств: они интуитивно понятны, легко вычи-

слимы, хорошо интерпретируемы, обеспечивают устойчивую основу для сравнения моделей на единых наборах данных. При этом исключительно узкая ориентация на точность предсказаний приводит к ряду методологических и практических ограничений, наиболее заметным проявлением которых является парадокс точности. Система, настроенная только на минимизацию ошибок предсказания, может стабильно предлагать пользователю уже знакомые ему объекты или же те, с которыми он ранее взаимодействовал, поскольку именно в подобных случаях вероятность правильного предсказания максимальна; но такое поведение не способствует расширению интересов клиента, открытию нового контента или увеличению вовлеченности. Более того, оно может усиливать эффект «информационного пузыря», ограничивая потребителя рамками уже сложившихся предпочтений.

**Метрики ранжирования** играют важную роль в оценке рекомендательных систем, так как позволяют учитывать не только факт наличия релевантных объектов в списке рекомендаций, но и порядок их расположения. В отличие от метрик точности (фиксирующих бинарный факт совпадения или несовпадения рекомендаций с релевантностью), метрики ранжирования моделируют реальный сценарий потребления информации, предполагая, что объекты, размещенные выше в списке, имеют больше шансов быть замеченными и выбранными пользователем. Ключевыми представителями метрик ранжирования являются NDCG (Normalized Discounted Cumulative Gain), MAP (Mean Average Precision), MRR (Mean Reciprocal Rank) и HitRate, каждая из которых по-своему интерпретирует важность порядка выдачи.

NDCG оценивает полезность списка рекомендаций посредством ее убывающего дисконтирования в зависимости от ранга: чем выше релевантный элемент находится в списке, тем больший вклад он вносит в итоговую оценку качества системы.

MAP агрегирует точность по позициям всех подходящих объектов, учитывая не только наличие первых успешных попаданий, но и то, насколько полно охвачены релевантные результаты в верхней части списка.

MRR определяет обратную величину ранга первого релевантного объекта, акцентируя внимание на скорости достижения первого положительного результата. Это делает данную метрику примени-

мой в системах, где важен именно первый успешный клик (первое успешное взаимодействие).

HitRate фиксирует факт появления хотя бы одного релевантного объекта среди первых рекомендаций, что удобно для грубой оценки покрываемости интересов пользователя на ограниченной глубине просмотра.

Метрики ранжирования приобретают важность в прикладных сценариях, когда потребитель ограничен в количестве просматриваемой информации — например, при отображении рекомендательных блоков на экране мобильных приложений, в потоковых сервисах, системах поиска и онлайн-ритейле. Поскольку пользователь чаще всего просматривает лишь несколько первых позиций списка, корректное упорядочивание последнего становится критически важным фактором качества рекомендаций.

Метрики ранжирования обладают повышенной чувствительностью к приоритетам потребителя: не все релевантные объекты для него равнозначны, и даже небольшие ошибки в ранжировании могут существенно повлиять на пользовательское восприятие.

**Метрики разнообразия** становятся все более важным элементом оценки рекомендательных систем, особенно на фоне стремительного роста объемов доступного контента и повышения конкуренции за внимание клиентов. В условиях избытка информации рекомендации должны не только соответствовать текущим интересам пользователя, но и стимулировать его к открытию новых областей интереса, поддерживать долгосрочную вовлеченность и предотвращать перенасыщение уже знакомым контентом. Разнообразие в списке рекомендаций, характеризуя степень различия между содержащимися в нем объектами, их принадлежность к разным категориям, жанрам, тематикам или даже стилям, отражает способность системы выходить за пределы уже известных предпочтений потребителя и предлагать контент, который потенциально может расширить его интересы. Среди метрик разнообразия можно выделить Diversity, Coverage, Novelty и Serendipity.

Diversity определяет среднюю степень различия между всеми парами объектов в списке рекомендаций. Высокое значение этой метрики указывает, что предложенные объекты существенно различаются между собой, — это снижает вероятность однообразного восприятия контента пользователем.

Coverage оценивает, насколько большая часть всего доступного каталога объектов представлена в рекомендациях. Это важно для систем, стремящихся избежать чрезмерной концентрации на ограниченном наборе популярных элементов.

Novelty фокусируется на измерении «новизны» рекомендаций, то есть на том, насколько предлагаемые объекты отличаются от уже знакомых клиенту. Высокая степень новизны позволяет стимулировать любопытство и открытие новых направлений.

Serendipity характеризует способность системы предлагать неожиданные, но релевантные рекомендации — они оказывались интереснее и ценнее тех, которые пользователь нашел бы самостоятельно.

Включение метрик разнообразия в процесс оценки рекомендательных систем связано с необходимостью противодействовать негативному эффекту «информационного пузыря», то есть предложению одних и тех же объектов, ограничивающему спектр интересов потребителя и снижающему удовлетворенность от его взаимодействия с платформой.

Однако увеличение разнообразия сопровождается определенными методологическими трудностями. С одной стороны, менее предсказуемые рекомендации труднее обосновать на основе имеющихся данных о пользователе, что может приводить к снижению метрик точности. С другой стороны, чрезмерный уход в сторону разнообразия без должного контроля за релевантностью контента способен ухудшить пользовательский опыт, поскольку система начнет предлагать объекты, которые (хотя и отличаются от привычного круга интересов) не вызывают у клиента реального отклика.

Важнейшей особенностью всех четырех групп метрик является то, что их использование требует компромисса: повышение одного показателя может приводить к ухудшению другого. Например, при попытке максимизировать точность наблюдается снижение разнообразия, а в случае улучшения ресурсной эффективности — потеря персонализации. Кроме того, многие метрики обладают высокой чувствительностью к характеристикам данных, что усложняет прямое сравнение моделей.

На этом фоне становится очевидной ограниченность практики применения изолированных показателей качества и встает вопрос о необходимости выработки комплексной оценки, способной объединить множество разнонаправленных критериев в единую систему принятия решений. Подобная оценка должна учитывать не только локальные

свойства алгоритма, но и его поведение в контексте бизнес-процесса, пользовательского взаимодействия и эксплуатационных ограничений.

## ДИСКУССИЯ

Согласно существующим исследованиям, фокусирование лишь на повышении точности ведет к предсказуемым и «скучным» рекомендациям, которые не повышают удовлетворенность и не побуждают пользователей к взаимодействию [27]. Данный тезис был поддержан на Workshop RecSys 2012 [28]: улучшение RMSE не говорит о пользе рекомендаций в реальных условиях. В этой связи перспективным направлением становится разработка метрик, ориентированных не только на точность, но и на субъективную ценность рекомендаций с точки зрения потребителя — например, их новизну, серендипность, доверие, эмоциональное вовлечение и др.

Тем не менее даже самые «продвинутое» метрики остаются лишь косвенными индикаторами реального пользовательского опыта: алгоритмы, которым соответствуют лучшие значения RMSE и NDCG на офлайн-выборках, далеко не всегда обеспечивают рост ключевых бизнес-показателей, таких как вовлеченность и удержание [29]. Ярким примером служит конкурс Netflix Prize: несмотря на десятипроцентное сокращение RMSE, победившее решение не было внедрено — оно не дало заметного прироста удержания пользователей и потребовало чрезмерных инженерных затрат. Таким образом, при оценке рекомендательных систем важно дополнять офлайн-метрики реальными A/B-экспериментами, чтобы учесть баланс между точностью предсказания и практической эффективностью.

В этой связи возникает необходимость интегрированного подхода к оценке, сочетающего различные офлайн-метрики, поведенческий анализ и обратную связь клиентов. Поведенческие KPI (CTR, конверсия, время сессии, удержание потребителя [retention]) дают объективную картину реального взаимодействия, но требуют аккуратного выбора и долгосрочного измерения, чтобы избежать оптимизаций, разрушающих доверие. Актуальной задачей становится формирование устойчивых и интерпретируемых поведенческих индикаторов, способных отражать долгосрочную ценность рекомендаций и минимизировать эффект переобучения на локальные максимумы в поведении пользователя.

Не менее важной остается проблема интерпретируемости результатов работы современных рекомендательных моделей [30]. Архитектуры, основанные на глубоком обучении и графовых нейросетях (несмотря на высокую точность) зачастую действуют как «черные ящики», не давая объяснений, почему система предлагает тот или иной объект. Это не только снижает прозрачность и доверие со стороны потребителя, но и ограничивает возможности разработчиков.

Одной из ключевых методологических дилемм остается баланс между точностью и разнообразием. Умеренная жертва первой в пользу новизны и серендипности зачастую повышает удовлетворенность пользователей, однако требует введения дополнительных контролирующих метрик.

Только комплексный подход позволит не только улучшать алгоритмические показатели, но и гарантировать их восприятие как релевантных. Обсуждение проблем рекомендательных систем выходит за пределы выбора конкретных алгоритмов — их будущее связано не только с повышением эффективности, но и с формированием принципов устойчивости, прозрачности и доверия.

## ВЫВОДЫ

Рекомендательные системы за последние десятилетия превратились из экспериментальных технологий в ключевые компоненты цифровой экономики, обеспечивающие персонализированное взаимодействие пользователей с информацией, товарами и услугами. Их роль не ограничивается узкой задачей поиска релевантного контента — они влияют на формирование потребительских предпочтений, моделируют цифровое потребление и выступают драйверами роста для широкого спектра бизнес-платформ.

Исторический путь развития рекомендательных систем демонстрирует последовательный переход от простых эвристик к сложным математическим моделям. Внедрение методов машинного обучения, глубоких нейронных сетей, графовых подходов и языковых моделей позволяет существенно расширить функциональные возможности рекомендаций, повысить их адаптивность, точность и контекстную осведомленность. Вместе с тем рост сложности моделей сопровождался проблематичностью их объективной оценки.

Развитие рекомендательных систем поставило множество методологических вопросов, в числе которых — «холодный старт» и разреженность

данных, приводящие к потере релевантности для новых пользователей и объектов; необходимость баланса между точностью и разнообразием, без которого система рискует оказаться в «информационном пузыре»; справедливость, прозрачность и уязвимость для манипуляций.

Выше говорилось о значительном разрыве между офлайн-оценками и реальными бизнес-результатами. Дальнейшее совершенствование области требует интеграции офлайн- и онлайн-метрик, постоянного проведения А/В-тестов и разработки новых гиперметрик, отражающих субъективную ценность рекомендаций; это позволит выстраивать системы, отвечающие одновременно инженерным, бизнес- и пользовательским требованиям.

Как показал проведенный в ходе исследования анализ, качество рекомендательных систем невозможно адекватно описать одной метрикой. Множество пользовательских сценариев и целей систем, а также потребность в учете эксплуатационных ограничений привели к формированию четырех ключевых направлений оценки, а именно: ресурсная эффективность, точность предсказаний, качество ранжирования и разнообразие рекомендаций. Множественность метрик в свою очередь породила необходимость комплексной оценки качества рекомендательных систем. Развитие методов решения этой проблемы является неотъемлемым условием построения систем нового поколения, ориентированных не только на эффективность, но и на устойчивое развитие цифровой среды в целом.

## БЛАГОДАРНОСТИ

Работа выполнена в рамках государственного задания Минобрнауки России; проект FEWM-2026-0011.

## ACKNOWLEDGEMENTS

The work was carried out within the framework of the state assignment of the Ministry of Education and Science of the Russian Federation; project FEWM-2026-0011.

## REFERENCES

1. Goldberg D., Nichols D., Oki B. M., Terry D. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*. 1992;35(12):61-70. DOI: 10.1145/138859.138867
2. Resnick P., Iacovou N., Suchak M., Bergstrom P., Riedl J. Grouplens: An open architecture for collaborative filtering of netnews. In: Proc. 1994 ACM conf. on computer supported cooperative work. New York, NY: Association for Computing Machinery; 1994:175-186. DOI: 10.1145/192844.192905
3. Schafer J.B., Konstan J., Riedl J. Recommender systems in e-commerce. In: Proc. 1<sup>st</sup> ACM conf. on electronic commerce. New York, NY: Association for Computing Machinery; 1999:158-166. DOI: 10.1145/336992.337035
4. Bennett J., Elkan C., Liu B., Smyth P., Tikk D. Kdd Cup and workshop 2007. *ACM SIGKDD Explorations Newsletter*. 2007;9(2):51-52. DOI: 10.1145/1345448.1345459
5. Koren Y., Bell R., Volinsky C. Matrix factorization techniques for recommender systems. *Computer*. 2009;42(8):30-37. DOI: 10.1109/MC.2009.263
6. Burke R. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*. 2002;12(4):331-370. DOI: 10.1023/A:1021240730564
7. Herlocker J.L., Konstan J.A., Terveen L.G., Riedl J.T. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*. 2004;22(1):5-53. DOI: 10.1145/963770.963772
8. Linden G., Smith B., York J. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*. 2003;7(1):76-80. DOI: 10.1109/MIC.2003.1167344
9. Lops P., De Gemmis M., Semeraro G. Content-based recommender systems: State of the art and trends. In: Ricci F., Rokach L., Shapira B., Kantor P., eds. Recommender systems handbook. Boston, MA: Springer; 2011:73-105. DOI: 10.1007/978-0-387-85820-3\_3
10. Celma Ò. Music recommendation. In: Music recommendation and discovery: The long tail, long fail, and long play in the digital music space. Berlin, Heidelberg: Springer; 2010:43-85. DOI: 10.1007/978-3-642-13287-2\_3
11. Lang K. Newsweeder: Learning to filter netnews. In: Prieditis A., Russell S.J., eds. Proc. 12<sup>th</sup> Int. conf. on machine learning (ICML'95). (Tahoe City, CA, July 9-12, 1995). San Francisco, CA; Morgan Kaufmann Publishers Inc.; 1995:331-339. DOI: 10.1016/B978-1-55860-377-6.50048-7

12. Burke R. Hybrid web recommender systems. In: Brusilovsky P., Kobsa A., Nejdl W., eds. The adaptive web: Methods and strategies of web personalization. Berlin, Heidelberg: Springer; 2007:377-408. (Lecture notes in computer science. Vol. 4321). DOI: 10.1007/978-3-540-72079-9\_12
13. Gomez-Uribe C.A., Hunt N. The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*. 2015;6(4):1-19. DOI: 10.1145/2843948
14. Kanakia A., Shen Z., Eide D., Wang K. A scalable hybrid research paper recommender system for Microsoft Academic. In: The World Wide Web conf. (WWW'19). New York, NY: Association for Computing Machinery; 2019:2893-2899. DOI: 10.1145/3308558.3313700
15. He X., Liao L., Zhang H., et al. Neural collaborative filtering. In: Proc. 26<sup>th</sup> Int. conf. on World Wide Web (WWW'17). New York, NY: Association for Computing Machinery; 2017:173-182. DOI: 10.1145/3038912.3052569
16. Sedhain S., Menon A.K., Sanner S., Xie L. AutoRec: Autoencoders meet collaborative filtering. In: Proc. 24<sup>th</sup> Int. conf. on World Wide Web (WWW'15). New York, NY: Association for Computing Machinery; 2015:111-112. DOI: 10.1145/2740908.2742726
17. Covington P., Adams J., Sargin E. Deep neural networks for YouTube recommendations. In: Proc. 10<sup>th</sup> ACM conf. on recommender systems (RecSys'16). New York, NY: Association for Computing Machinery; 2016:191-198. DOI: 10.1145/2959100.2959190
18. Hidasi B., Karatzoglou A., Baltrunas L., Tikk D. Session-based recommendations with recurrent neural networks. In: 4<sup>th</sup> Int. conf. on learning representations (ICLR-2016). (San Juan, May 2-4, 2016). Appleton, WI: ICLR; 2016. URL: [https://hidasi.eu/assets/pdf/gru4rec\\_iclr16.pdf](https://hidasi.eu/assets/pdf/gru4rec_iclr16.pdf)
19. Kang W.-C., McAuley J. Self-attentive sequential recommendation. In: 2018 IEEE int. conf. on data mining (ICDM 2018). (Singapore, November 17-20, 2018). New York, NY: IEEE; 2018:197-206. DOI: 10.1109/ICDM.2018.00035
20. Chen Q., Zhao H., Li W., Huang P., Ou W. Behavior sequence transformer for e-commerce recommendation in Alibaba. In: Proc. 1<sup>st</sup> Int. workshop on deep learning practice for high-dimensional sparse data (DLP-KDD'19). New York, NY: Association for Computing Machinery; 2019:1-4. DOI: 10.1145/3326937.3341261
21. Ying R., He R., Chen K., et al. Graph convolutional neural networks for web-scale recommender systems. In: Proc. 24<sup>th</sup> ACM SIGKDD Int. conf. on knowledge discovery & data mining (KDD'18). New York, NY: Association for Computing Machinery; 2018:974-983. DOI: 10.1145/3219819.3219890
22. Wang H., Zhang F., Wang J., et al. Ripplenet: Propagating user preferences on the knowledge graph for recommender systems. In: Proc. 27<sup>th</sup> ACM Int. conf. on information and knowledge management (CIKM'18). New York, NY: Association for Computing Machinery; 2018:417-426. DOI: 10.1145/3269206.3271739
23. Ricci F., Rokach L., Shapira B., eds. Recommender systems handbook. New York, NY: Springer; 2023. 1060 p. DOI: 10.1007/978-1-0716-2197-4
24. Ge M., Delgado-Battenfeld C., Jannach D. Beyond accuracy: Evaluating recommender systems by coverage and serendipity. In: Proc. 4<sup>th</sup> ACM conf. on recommender systems (RecSys'10). New York, NY: Association for Computing Machinery; 2010:257-260. DOI: 10.1145/1864708.1864761
25. Abdollahpouri H., Adomavicius G., Burke R., et al. Beyond personalization: Research directions in multistakeholder recommendation. 2019. URL: [https://www.researchgate.net/publication/332896126\\_Beyond\\_Personalization\\_Research\\_Directions\\_in\\_Multistakeholder\\_Recommendation](https://www.researchgate.net/publication/332896126_Beyond_Personalization_Research_Directions_in_Multistakeholder_Recommendation)
26. Zhang S., Yao L., Sun A., Tay Y. Deep learning based recommender system: A survey and new perspectives. *ACM computing surveys (CSUR)*. 2019;52(1):1-38. DOI: 10.1145/3285029
27. McNee S.M., Riedl J., Konstan J.A. Being accurate is not enough: How accuracy metrics have hurt recommender systems. In: CHI'06 Extended abstracts on human factors in computing systems (CHI EA'06). New York, NY: Association for Computing Machinery; 2006:1097-1101. DOI: 10.1145/1125451.1125659
28. Amatriain X., Castells P., de Vries A., Posse C., Steck H. Workshop on recommendation utility evaluation: Beyond RMSE — RUE 2012. In: Proc. 6<sup>th</sup> ACM conf. on recommender systems (RecSys'12). New York, NY: Association for Computing Machinery; 2012:351-352. DOI: 10.1145/2365952.2366042
29. Jannach D., Jugovac M. Measuring the business value of recommender systems. *ACM Transactions on Management Information Systems (TMIS)*. 2019;10(4):1-23. DOI: 10.1145/3370082
30. Pu P., Chen L., Hu R. A user-centric evaluation framework for recommender systems. In: Proc. 5<sup>th</sup> ACM conf. on recommender systems (RecSys'11). New York, NY: Association for Computing Machinery; 2011:157-164. DOI: 10.1145/2043932.2043962

## ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS



**Роман Сергеевич Кульшин** — старший преподаватель кафедры автоматизации обработки информации, аспирант кафедры автоматизации обработки информации, Томский государственный университет систем управления и радиоэлектроники, Томск, Российская Федерация

**Roman S. Kulshin** — Senior lecturer of the Department of Data Processing Automation, Postgraduate student of the Department of Data Processing Automation, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation

<https://orcid.org/0000-0002-6891-1869>

*Автор для корреспонденции / Corresponding author:*

[roman.s.kulshin@tusur.ru](mailto:roman.s.kulshin@tusur.ru)



**Анатолий Анатольевич Сидоров** — кандидат экономических наук, доцент, заведующий кафедрой автоматизации обработки информации, Томский государственный университет систем управления и радиоэлектроники, Томск, Российская Федерация

**Anatoly A. Sidorov** — Cand. Sci. (Econ.), Assoc. Prof.; Head of the Department of Data Processing Automation, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation

<https://orcid.org/0000-0002-9236-3639>

[anatolii.a.sidorov@tusur.ru](mailto:anatolii.a.sidorov@tusur.ru)

*Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.*

*Conflicts of Interest Statement: The authors have no conflicts of interest to declare.*

*Статья поступила в редакцию 18.11.2025; после рецензирования 11.02.2026; принята к публикации 15.04.2026.*

*Авторы прочитали и одобрили окончательный вариант рукописи.*

*The article was submitted on 18.11.2025; revised on 11.02.2026 and accepted for publication on 15.04.2026.*

*The authors read and approved the final version of the manuscript.*