



УДК 004  
doi: 10.21685/2587-7704-2024-9-1-3



Open  
Access

RESEARCH  
ARTICLE

## Современные нейросетевые подходы для рекомендательных систем

**Евгений Валентинович Гришин**

Пензенский государственный университет, Россия, г. Пенза, ул. Красная, 40  
grishin\_evgenii@inbox.ru

**Лев Вячеславович Гурьянов**

Пензенский государственный университет, Россия, г. Пенза, ул. Красная, 40  
leo8087@yandex.ru

**Аннотация.** С развитием технологий машинного обучения рекомендательные системы претерпели значительные изменения касательно их архитектуры и способов построения. В работе рассматриваются современные подходы к проектированию рекомендательных систем, в которых применяются нейронные сети.

**Ключевые слова:** рекомендательные системы, нейронные сети, обработка эмбедингов, извлечение признаков

**Для цитирования:** Гришин Е. В., Гурьянов Л. В. Современные нейросетевые подходы для рекомендательных систем // Инжиниринг и технологии. 2024. Т. 9 (1). С. 1–5. doi: 10.21685/2587-7704-2024-9-1-3

## Modern Neural Network applications in Recommendation Systems

**Evgenii V. Grishin**

Penza State University, 40 Krasnaya Street, Penza, Russia  
grishin\_evgenii@inbox.ru

**Lev V. Guryanov**

Penza State University, 40 Krasnaya Street, Penza, Russia  
leo8087@yandex.ru

**Abstract.** With a rapid development of machine learning technologies, recommendation systems have undergone dramatic changes concerning their architecture and the way they are built. The focus of this article is on modern designs of recommendation systems that apply the power of deep neural networks.

**Keywords:** recommendation systems, neural networks, embeddings processing, feature extraction

**For citation:** Grishin E.V., Guryanov L.V. Modern Neural Network applications in Recommendation Systems. *Inzhiniring i tekhnologii = Engineering and Technology*. 2024;9(1):1–5. (In Russ.). doi: 10.21685/2587-7704-2024-9-1-3

Рекомендательные системы являются одним из ключевых компонентов социальных приложений. Используя рекомендательные алгоритмы, пользователи получают возможность быстрее находить персонализированный контент, делая сервис эффективнее. Во время разработки систем приходится преодолевать такие проблемы, как высокая нагрузка, холодный старт, требования к гибкости, время отклика, частота обновлений, выделение критерия качества. Целью исследования является разработка рекомендательной системы, способной функционировать на единственном устройстве и предоставлять рекомендации в режиме реального времени.

Авторское решение основано на использовании графовых представлений среды выполнения рекомендательной системы [1]. Такие системы можно представить в качестве двудольных графов,



в которых одна из долей представлена агрегаторами, а другая – рекомендуемыми сущностями. Этот подход позволяет упростить структуру и свести оптимизацию к сжатию графа и алгоритмам случайных обходов. Несмотря на это, стоит обратить внимание на нейросетевые архитектуры рекомендательных систем, как на конкурентоспособные алгоритмы, расширяющие свое влияние и применяемые в крупномасштабных проектах, в качестве потенциальных алгоритмов для сравнения эффективности.

Среди рассмотренных работ можно выделить несколько основных векторов развития архитектур нейронных сетей:

- сравнение рекомендуемых объектов по содержанию;
- ускорение извлечения признаков при использовании глубокого обучения;
- совершенствование способов представления данных в нейронных сетях;
- разделение процесса предоставления рекомендаций на несколько стадий;
- корректировка целевых метрик и методов оценки качества модели.

Модели рекомендательных систем восходят к методам построения экспертных систем, и одним из первых способов являлось использование ассоциативных правил. Однако этот подход отличается высоким вмешательством эксперта в процесс, низкой динамичностью рекомендаций и медленным циклом обновления системы при добавлении нового контента. Поэтому ему на смену пришли алгоритмы, которые на текущий момент разделяют на два класса: коллаборативная фильтрация и фильтрация по содержанию (рис. 1).

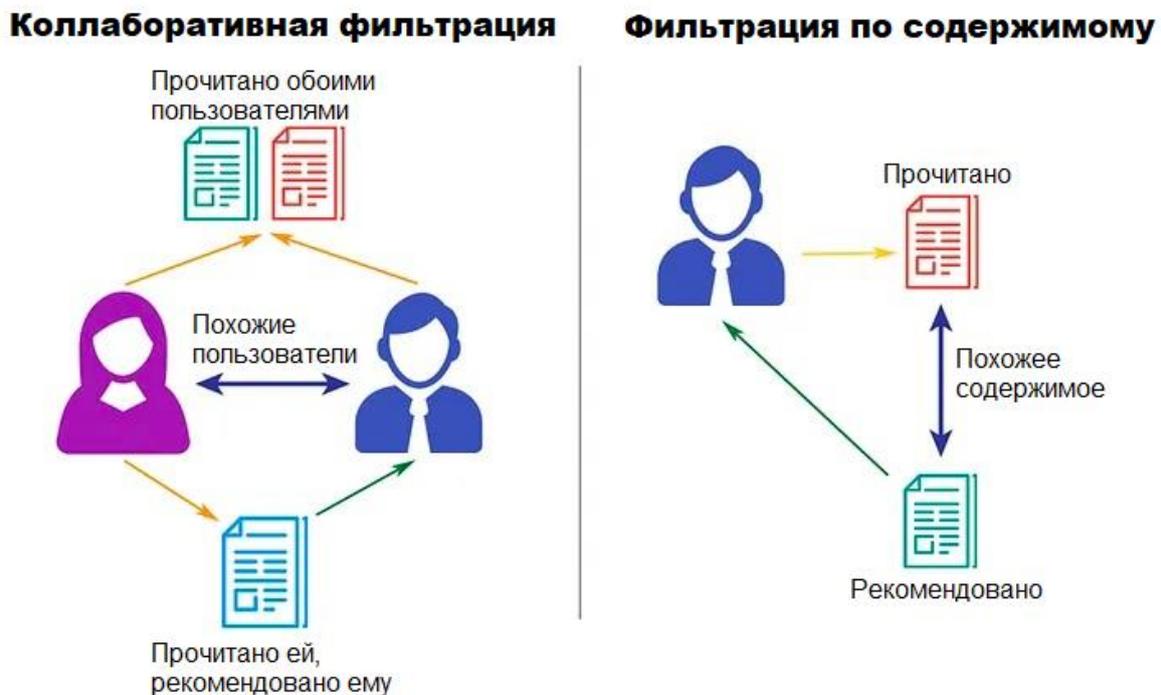


Рис. 1. Принцип работы алгоритмов коллаборативной фильтрации и на основе содержимого

В случае применения коллаборативной фильтрации, алгоритмы направлены на корректное сопоставление пользователей системы по их действиям в системе. В результате, пользователи, чья активность наиболее коррелирует относительно некоторой метрики, получают похожие элементы для просмотра. Противопоставлены им алгоритмы, сравнивающие содержание рекомендуемого контента. Большой проблемой является сложность анализа различных типов данных. Однако с расширением применения нейронных сетей в областях компьютерного зрения и естественного языка появились новые возможности для построения таких систем. В результате одним из векторов развития нейронных сетей является их использование для оценки схожести двух рекомендуемых элементов или составление векторных представлений (эмбеддингов) содержимого по данным их использования [2].

Появление новых способов представления данных с использованием нейронных сетей породило новую волну применения машинного обучения в рекомендательных системах. Глубокие нейронные сети обладают потенциалом для извлечения признаков данных со снижением вмешательства человека в этот процесс. Дальнейшие исследования были направлены на способы ускорения обучения нейронных сетей. Так как объем данных велик, модели должны оставаться легковесными для





Наиболее свежим взглядом на данные рекомендательных систем стала попытка обучения графовых векторных представлений. Так, нейронная система TwHIN [5] используется для обучения зависимостей между объектами гетерогенных графов, объединяя множество наборов графовых данных в одну масштабную модель, шардируемую на нескольких GPU. При использовании таких комплексных моделей, зависимости между объектами позволяют намного точнее описать взаимодействие пользователей с внутренними объектами окружения, в котором работает рекомендательная система. Это позволяет использовать единое векторное представление для различных целей: отбора кандидатов, выделение негативно воспринимаемого контента, продвижение рекламы и предсказание действий пользователя при взаимодействии с элементом (рис. 4).

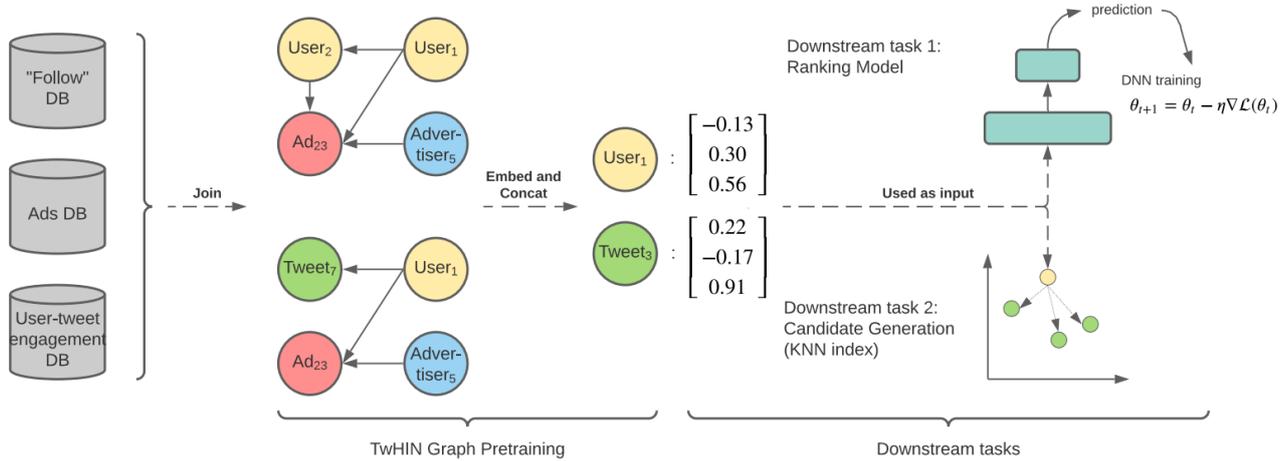


Рис. 4. Векторные представления гетерогенных графовых зависимостей

В результате этих изменений обновились способы организации рекомендательных систем. Так как различные алгоритмы имеют различные сильные и слабые стороны, они всегда комбинировались в итоговом продукте. Однако на текущий момент разработчиками была выработана единообразная схема построения рекомендательных систем, включающая такие компоненты, как генераторы кандидатов, ранкеры, миксеры и фильтры (рис. 5).

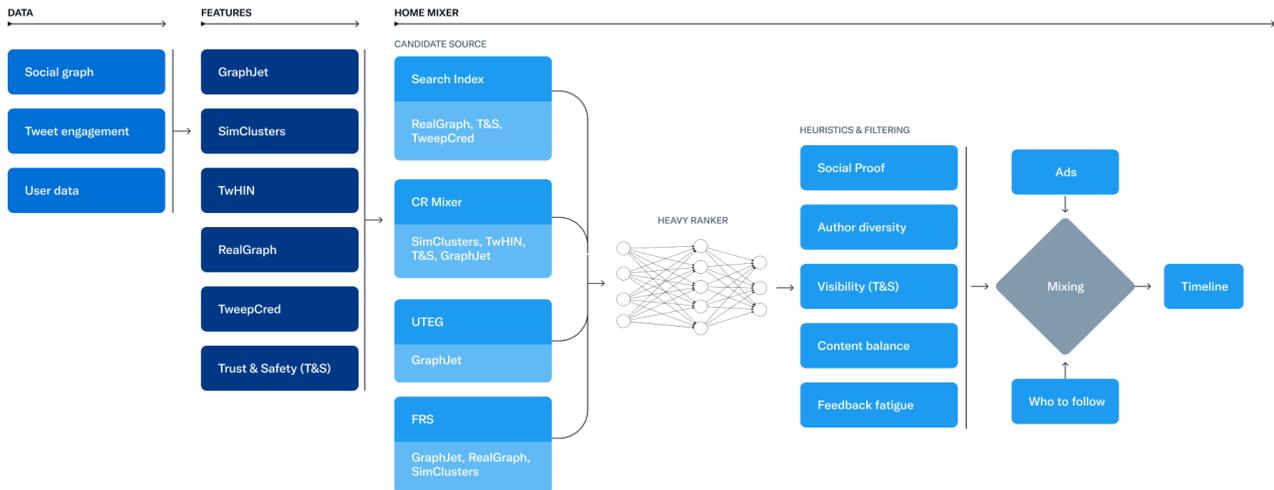


Рис. 5. Архитектура «алгоритма» Twitter, построенного по современной модели рекомендательных систем

В качестве генераторов часто выступают методы кластеризации, обработки эмбедингов, статистики и случайные обходы из-за их более низкой требовательности к ресурсам на больших объемах данных. После первоначального отбора кандидатов они оцениваются ранкерами, которые делают рекомендации персонализированными и отсортированными в порядке прогнозируемого интереса, который проявит пользователь. Миксеры распределяют содержимое из разных источников и увеличивают степень вариативности среди рекомендаций. Фильтры проверяют предоставляемые элементы на соответствие этике, состоянию просмотра этих элементов пользователем.



Несмотря на все новые методы работы с нейронными сетями и их эффективностью, обоснованной маркетинговыми исследованиями, совершенствование технологий породило новые задачи для решения, связанные с адекватной оценкой эффективности системы. Изменение алгоритма на платформе YouTube привело к широкой охоте за «кликбейтами». Когда авторы контента осознали важность Click-Through Rate (CTR) для рекомендации их видео, фокус платформы сместился с набора подписчиков на подготовку громких заголовков и захватывающих превью. Фактически это перевело систему из режима почтового сервиса, где пользователи подписаны на своих блоггеров и просматривают их видео, привлекая новую аудиторию к формату «желтой прессы», где сенсация важнее качества содержимого. Подобный пример напоминает о важности метрик, которые оптимизируются рекомендательной системой, а также о роли оценки качества. Возможно, из-за этого фокус платформы смещается в новом направлении метрики «возвращения пользователя» после просмотра некоторого контента к другому контенту этого автора.

### Список литературы

1. Гришин Е. В., Гурьянов Л. В. Обзор алгоритмов и моделей рекомендательных систем // Информационные технологии в науке и образовании. Проблемы и перспективы : сб. ст. по материалам X Всерос. науч.-практ. конф., Пенза, 15 марта 2023 года. Пенза : Изд-во ПГУ, 2023. С. 115–118.
2. Van den Oord A., Dieleman S., Schrauwen B. Deep content-based music recommendation // *Advances in neural information processing systems*. 2013. Vol. 26.
3. Wang R. Dcn v2: Improved deep & cross network and practical lessons for web-scale learning to rank systems // *Proceedings of the web conference 2021*. 2021. P. 1785–1797.
4. Wang Z., She Q., Zhang J. MaskNet: Introducing feature-wise multiplication to CTR ranking models by instance-guided mask // *arXiv preprint arXiv:2102.07619*. 2021.
5. El-Kishky A. TwHIN: Embedding the twitter heterogeneous information network for personalized recommendation // *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*. 2022. P. 2842–2850.

### References

1. Grishin E.V., Gur'janov L.V. Overview of algorithms and models of recommendation systems. *Informacionnye tehnologii v nauke i obrazovanii. Problemy i perspektivy: sb. st. po materialam X Vseros. nauch.-prakt. konf., Penza, 15 marta 2023 goda = Information technologies in science and education. Problems and prospects : collection of articles based on the materials of the X All-Russian Scientific.-practical conference, Penza, March 15, 2023*. Penza: Izd-vo PGU, 2023:115–118. (In Russ.)
2. Van den Oord A., Dieleman S., Schrauwen B. Deep content-based music recommendation. *Advances in neural information processing systems*. 2013;26.
3. Wang R. Dcn v2: Improved deep & cross network and practical lessons for web-scale learning to rank systems. *Proceedings of the web conference 2021*. 2021:1785–1797.
4. Wang Z., She Q., Zhang J. MaskNet: Introducing feature-wise multiplication to CTR ranking models by instance-guided mask. *arXiv preprint arXiv:2102.07619*. 2021.
5. El-Kishky A. TwHIN: Embedding the twitter heterogeneous information network for personalized recommendation. *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*. 2022:2842–2850.

Поступила в редакцию / Received 10.03.2024

Принята к публикации / Accepted 10.04.2024