УДК 519.85
2.3.1. Системный анализ, управление и обработка информации




А.П. Сергушичева, В.В. Петров
Вологодский государственный университет

РАЗРАБОТКА САМООБУЧАЮЩЕЙСЯ ГРАФОВОЙ НЕЙРОННОЙ СЕТИ ДЛЯ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ ЭЛЕКТРОННОЙ БИБЛИОТЕКИ

В статье рассматривается разработка самообучающейся графовой нейронной сети для рекомендательной системы электронной библиотеки. Представлен анализ современных методов построения рекомендательных систем, включая классические алгоритмы коллаборативной и контентной фильтрации, а также их ограничения. Обоснован выбор графовых нейронных сетей как наиболее перспективного подхода для учета многосвязных зависимостей между пользователями, книгами и метаданными. Описана архитектура предлагаемой модели на основе LightGCN с механизмом адаптивного обновления представлений узлов. Сформулированы направления практической реализации системы.

Рекомендательная система, графовые нейронные сети, самообучение, электронная библиотека, персонализация, LightGCN.

Программные системы, позволяющие на основе информации о пользователе формировать интересные для него предложения по выбору товаров, услуг, контента и иных объектов, получили название рекомендательных. Рекомендательные системы приносят пользу как бизнесу, так и его клиентам. Для первых за счет персонализированных предложений увеличивается объем продаж, повышается эффективность маркетинговых кампаний, анализ спроса и предпочтений клиентов делает возможными оптимизацию управления запасами и прогнозирование перспективных направлений развития бизнеса. А пользователи экономят время на поиск контента, соответствующего их индивидуальным интересам. Рекомендательные технологии позволяют найти релевантные объекты, которые затруднительно обнаружить с помощью традиционного поиска. Кроме того, рекомендательные системы используются для работы со слабоструктурированными и разнородными данными.

Рекомендательные системы применяются в различных отраслях: Spotify и Яндекс.Музыка предлагают плейлисты, учитывая прослушанные композиции; на сайтах интернет-магазинов есть блоки с рекомендациями товаров, Delivery Club предлагает рестораны и блюда, ориентируясь на предыдущие заказы; Booking.com делает выборку отелей в соответствии с запросами туристов и т.д. В условиях обилия информации рекомендательные системы становятся важным инструментом персонализации контента для помощи пользователям в навигации по большим каталогам. Около 35 % выручки крупных онлайн-магазинов и до 75 % просмотров на стриминговых платформах формируется за счет рекомендаций [1]. В электронных библиотеках персонализация повышает удобство, вовлеченность и удовлетворенность пользователей, предлагая релевантные издания вместо простого списка популярных книг.

В работе рекомендательной системы можно выделить три этапа:

1) сбор информации. Чтобы сформировать персонализированные рекомендации и сделать прогноз о том, какой контент может вызвать интерес субъекта, необходимо выявить его особенности: предпочтения, хобби, круг общения и т.п. Можно попросить пользователя оценить рекомендацию (по бальной шкале, с помощью лайка и др.), создать список любимых пользователем объектов, выбрать лучший объект из двух предъявленных, расставить объекты от наилучшего к наихудшему. На практике рейтинги, проставленные пользователями объектам, встречаются сравнительно редко. Больше информации можно получить путем неявного сбора данных, например: фиксируя просматриваемые индивидуумом страницы, клики и добавления в закладки; отслеживая содержимое его компьютера. Однако эти сведения не говорят об отношении пользователя к объекту: выбор заинтересовавшей его книги не дает представления о том, насколько полезной она оказалась;

2) анализ информации. Собранную информацию представляет матрица предпочтений, в которой одна из осей отвечает за пользователей, вторая – за объекты рекомендации. Данная матрица очень разрежена: пользователь может оценить лишь часть объектов. Анализ информации позволяет выявить закономерности и связи между пользователями и объектами и заполнить пустые ячейки матрицы;

3) генерация рекомендаций. Основная задача данного этапа – отобрать и предложить пользователю лучшие варианты по вероятности их соответствия запросу. Традиционные методы построения рекомендаций включают коллаборативную фильтрацию, контентную фильтрацию и гибридный подход [2].

Основная идея коллаборативной фильтрации (CF) [3] – похожим пользователям нравятся похожие объекты, т.е. если в прошлом у двух пользователей были

близкие оценки одинаковых объектов, то вероятно их вкусы совпадут и в будущем. Метод предполагает объединение пользователей в группы (кластеры) по схожести их взаимодействий с объектами. В этом случае не имеет значения, какие типы объектов рассматривались: покупки, книги, фильмы и/или т.п. При этом могут быть отражены неявные характеристики, которые сложно учесть при создании профиля. Предсказание предпочтений пользователя для конкретного объекта осуществляется на основе агрегированных оценок представителей кластера, к которому данный субъект принадлежит. Метод также предполагает определение аналогичности объектов. Существуют Memory-based подходы, напрямую вычисляющие похожесть между пользователями или объектами, и Model-based подходы, использующие факторизационные модели для выявления скрытых факторов (факторный анализ показывает уровень влияния разных факторов на конечный результат).

Преимущества коллаборативной фильтрации заключаются в учете реальных поведенческих паттернов, способности рекомендовать объекты, не похожие по контенту и вне зависимости от признаков объектов. Основным недостатком метода считается проблема «холодного старта» для новых пользователей и объектов. Другие ограничения: разреженность данных в больших системах, квадратичная сложность Memory-based CF, отсутствие учета контекста и специфики пользователя, низкая интерпретируемость результатов.

Контентная фильтрация (фильтрация на основе содержимого**)** [4] предполагает построение рекомендаций по результатам сравнения описательных характеристик объектов с пользовательскими предпочтениями. Для каждого субъекта создается профиль. Профиль пользователя может включать демографическую информацию (возраст, пол, этническую принадлежность и т.п.), ответы на определенный набор вопросов об интересах респондента. Описание объектов содержит атрибутивную информацию в зависимости от их типа, которая будет использоваться для выявления предпочтений пользователя. Например, для музыкального произведения это автор, жанр, характер музыки, ритм, темп и другие. Для книги в подобный список входят жанр, библиографические данные (ISBN, авторы, название, год, издательство), ключевые слова, цена, элементы оформления и т.п. Данный подход устраняет проблему «холодного старта» (не нужны отзывы других пользователей) и обеспечивает прозрачность рекомендаций через общие признаки. Ограничения метода связаны со склонностью «зацикливаться» на одном типе контента, зависимостью от полноты метаданных и сложностью работы с субъективными атрибутами.

Гибридные системы объединяют коллаборативную фильтрацию и контентные методы для компенсации их взаимных недостатков: смягчают «холодный старт», повышают точность и обеспечивают большее разнообразие рекомендаций. Например, гибридная система может использовать контентную фильтрацию для новых пользователей, переходя по мере накопления данных взаимодействий на коллаборативные методы. Однако сложность их разработки и сопровождения, высокие вычислительные требования делают полноценные гибриды нерациональными для небольших команд и академических проектов.

Возрастание требований к качеству рекомендаций (пользователи ожидают рекомендации, точно соответствующие их вкусам, и доверяют системам, способным объяснить свои советы) стимулирует переход от традиционных алгоритмов к более совершенным методам. К современным рекомендательным технологиям относятся: глубокое обучение, графовые алгоритмы, обучение с подкреплением [2]. Перечисленные методы основаны на работе искусственных нейронных сетей.

Искусственные нейронные сети построены на принципах биологических нейронных сетей и состоят из способных обрабатывать сигналы искусственных нейронов, соединенных синапсами [5]. Как правило, нейроны организованы в слои. Выделяют входной, скрытые (вычислительные) и выходной слои. Разные слои могут выполнять различные виды преобразований. Нейроны и синапсы имеют вес, который влияет на силу сигнала и изменяется в процессе обучения. Обучение нейросети ведется на обучающем наборе данных (датасете), содержащем примеры с правильными результатами решений. Нейронная сеть пытается воспроизвести результат самостоятельно на новом наборе данных. Обучение нейросети происходит в два этапа: 1) прямое распространение ошибки – направлен на предсказание ответа; 2) обратное распространение ошибки – минимизируется ошибка между фактическим ответом и предсказанным. Нейронные сети используются для решения различных задач: машинное зрение, распознавание речи, машинный перевод, медицинская диагностика и другие. Развитие теории искусственных нейронных сетей и увеличение вычислительных мощностей позволили создавать сложные технологические архитектуры нейронных сетей.

Глубокая нейронная сеть (DNN – Deep Neural Network) содержит несколько вычислительных слоев и подходит для обработки данных без четкой структуры (например, изображений и текста). При этом эффективность и качество решения задач сопоставимо с возможностями человека. В рекомендательных системах алгоритмы глубокого обучения используют нейронные сети для выявления сложных нелинейных зависимостей между пользователями и объектами. Методы глубокого обучения NCF, AutoRec, DeepFM и рекуррентных архитектур для последовательных рекомендаций обладают высокой предсказательной мощностью и гибкостью, автоматическим выделением признаков. NCF (нейронная коллаборативная фильтрация) предлагает простой слой представления сразу для пользователей и объектов, за которым следует простая нейронная сеть вроде многослойного перцептрона, которая должна усвоить зависимость между представлениями пользователя и объекта. DeepFM (глубокие машины факторизации) – это смешанный подход, включающий факторизацию матриц и глубокую нейронную сеть, причем оба используют один и тот же входной слой представления (embedding). Идея этой комбинации – параллельно изучать зависимости между признаками низких и высоких порядков. DLRM – рекомендационная модель

глубокого обучения. На первой стадии каждый категориальный признак выражается вектором представления, а постоянные признаки обрабатываются многослойным перцептроном. На второй стадии рассчитываются взаимные произведения всех векторов представления и выходных векторов перцептрона. Затем произведения передаются в другой многослойный перцептрон и далее в функцию сигмоиды, выдающую вероятность. Глубокие модели способны обрабатывать разнообразные и большие наборы данных, обеспечивая высокую точность рекомендаций. Однако эти алгоритмы обладают низкой интерпретируемостью и высокой вычислительной сложностью [2].

Обучение с подкреплением (Reinforcement Learning, RL) применяется для оптимизации рекомендаций в режиме реального времени и фактически имитирует процесс самообучения человека [6]. Система (агент) анализирует свое окружение и определяет его состояние. Основываясь на текущей стратегии, агент совершает действие, тем самым меняя состояние окружения, и получает «награду» (сигнал подкрепления): положительную, если состояние среды улучшилось, и отрицательную, если ухудшилось. Агент анализирует результаты своих действий и обновляет стратегию, чтобы улучшить будущие результаты. В рекомендательных системах агент стремится максимизировать долгосрочное удовлетворение пользователя рекомендациями, основываясь на обратной связи от него (клики, оценки). Метод позволяет учитывать предпочтения клиентов в динамическом режиме, однако требует значительного объема данных и сопровождается рисками негативного опыта в процессе обучения.

Графовая нейронная сеть (GNN) применяется для анализа данных, представленных в виде графов (узлов и связей). Графы не обладают регулярной структурой и количество соседей у каждого узла – величина переменная. Для каждого узла и связи составляется векторная модель, учитывающая и атрибуты, и контекст в графе. Процесс и результат преобразования сложных данных в компактные числовые многомерные векторы также называют графовым эмбеддингом. За счет векторного представления схожие по смыслу объекты оказываются близко друг к другу. Эмбеддинг позволяет «понимать» семантику, измерять сходство использовать математические методы для поиска, рекомендаций и классификации. GNN работают с нерегулярными, сложными структурами, передавая информацию между узлами через их соседей. Каждый узел собирает информацию (признаки) от своих соседей и агрегирует ее (например, усредняет берет максимум и т.п.). GNN эффективно обрабатывают данные, учитывая и характеристики объектов, и структуру связей [7].

Сочетание растущей потребности в персонализации и новых графовых методов обусловливает выбор GNN для построения рекомендательной системы электронной библиотеки. Данные рекомендательных систем по своей природе формируют графовую структуру. Для электронной библиотеки граф обычно состоит из двух типов узлов: пользователи и книги (объекты). Ребра отражают их взаимодействия (оценки, покупки, просмотры). Расширение графа разнородными узлами (авторы, жанры, теги) позволяет формировать гетерогенный граф. Каждый узел обновляет свое представление, агрегируя информацию от соседних узлов по нескольким итерациям. Построение рекомендательной системы на базе GNN позволяет учесть структурную информацию и многопутевые связи через явное распространение коллаборативных сигналов, интегрировать гетерогенные данные в едином графовом представлении. GNN в электронных библиотеках устраняет проблему «холодного старта»: новые книги получают возможность быть рекомендованными через связи с жанрами, авторами или схожими изданиями, а новые пользователи – быть представленными через базовый профиль или социальные связи. Также графовые нейронные сети обеспечивают высокую точность по метрикам ранжирования на стандартных датасетах и возможность частичной интерпретации рекомендаций через пути в графе [8].

В рекомендательных системах применяются графовые нейронные сети различной архитектуры. Модель GC-MC – ранняя адаптация графовой сверточной сети к задаче восстановления рейтингов; PinSage – масштабируемое решение для графов с миллиардами связей; архитектура NGCF учитывает нескольких слоев пропагации. Одной из наиболее популярных моделей для исследований является LightGCN. Это упрощенная версия GCN без весовых матриц и активаций, превосходящая классические методы по метрикам качества [7]. Для рекомендательной системы электронной библиотеки авторами предлагается использовать LightGCN-подобную модель. Выбор обоснован ее точностью, простотой и эффективностью.

Структура предлагаемой модели представлена на рисунке. В GCN убраны весовые матрицы и активации, оставлено линейное усреднение соседей. Модель расширена учетом контекстных признаков и метаданных электронных книг. Предложена модификация механизма агрегации соседей для учета весов жанровых и авторских связей, что позволяет формировать более информативные эмбеддинги узлов. Структура модели включает следующие компоненты: преобразователь данных в гетерогенный граф; слои пропагации сообщений GNN; механизм самообучения на основе отрицательной выборки и регуляризации; модуль обновления эмбеддингов при изменении пользовательской активности; подсистему хранения и пересчета представлений узлов.

Методика обучения модели включает: многоступенчатое сэмплирование соседей для снижения вычислительной сложности; адаптивную негативную выборку, увеличивающую устойчивость к перекосу в сторону популярных книг; динамическое обновление эмбеддингов при поступлении новых данных; оптимизацию функции потерь с учетом структурных особенностей графа; регуляризацию многослойной пропагации, предотвращающую эффект oversmoothing (чрезмерное сглаживание). Принципы самообучения модели предусматривают обновление ее представлений на основе новых пользовательских действий без полного переобучения. Это обеспечивает адаптацию системы к изменениям пользовательского поведения и позволяет оперативно добавлять новые узлы и ребра в граф взаимодействий.

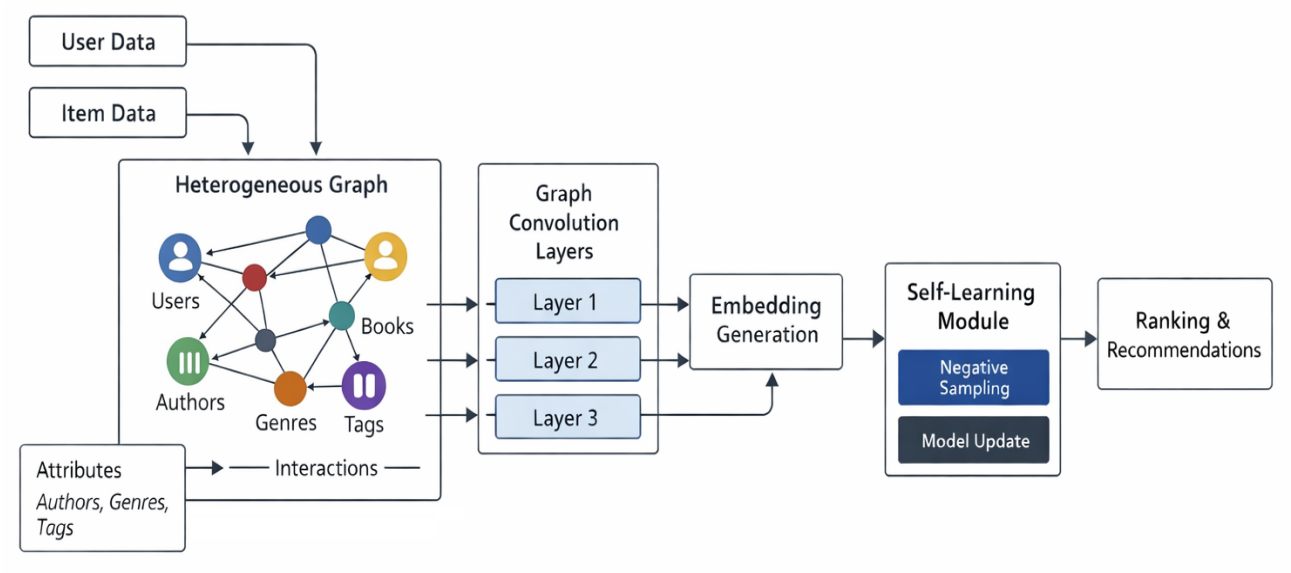


Рис. Архитектура самообучающейся графовой нейронной сети для рекомендательной системы электронной библиотеки

На выходе модели формируется ранжированный список книг для конкретного пользователя. Оценка релевантности может вычисляться по близости эмбеддингов пользователя и книги, после чего выполняется сортировка кандидатов и отбор рекомендаций Top-N. Такой подход обеспечивает совместимость предложенной архитектуры с прикладным веб-сервисом электронной библиотеки.

Предлагаемая модель реализована в первой версии рекомендательной системы электронной библиотеки. Приложение развернуто и показало свою работоспособность. В настоящее время ведется работа над совершенствованием системы и реализацией автоматического дообучения модели.

Таким образом, использование GNN позволяет естественным образом описывать данные в виде графа «пользователь – объект», учитывать многопутевые связи и интегрировать гетерогенную информацию о метаданных книг. Механизм самообучения обеспечивает адаптацию к динамически изменяющимся данным без полного переобучения модели, что особенно важно для электронных библиотек с постоянно пополняемым фондом и меняющейся пользовательской активностью. Полная реализация предложенной модели дает возможность повысить релевантность рекомендаций, снизить влияние проблемы «холодного старта», ускорить адаптацию системы к новым действиям пользователей и обеспечить более полное использование метаданных электронного каталога.

Литература

1. Анатомия рекомендательных систем. Часть первая : Хабр. – URL: https://habr.com/ru/companies/lanit/articles/420499/ (дата обращения: 28.09.2026). – Текст : электронный.

2. Петров, В. В. Современные алгоритмы рекомендательных систем: обзор и сравнительный анализ / В. В. Петров, С. М. Катернюк, М. С. Ухин // Молодые исследователи – регионам : материалы Международной научной конференции (Вологда, 22 апреля 2025 г.). – Вологда : ВоГУ, 2025. – С. 318–320.

3. Рекомендательные системы: как ML-инженеры создают персонализированный опыт? – URL: https://simulative.ru/blog/rekomendatelnie-sistemi-kak-ml-inzheneri-sozdayut (дата обращения: 28.09.2026). – Текст : электронный.

4. Оболенский, Д. М. Обзор современных методов построения рекомендательных систем на основе контента и гибридные системы / Д. М. Оболенский, В. И. Шевченко // Мир компьютерных технологий : сборник статей всероссийской научно-технической конференции студентов, аспирантов и молодых ученых. – Севастополь : Севастопольский государственный университет, 2021. – С. 151–156.

5. Гафаров, Ф. М. Искусственные нейронные сети и приложения : учебное пособие / Ф. М. Гафаров, А. Ф. Галимянов. – Казань : Издательство Казанского университета, 2018. – 121 с.

6. Курзанова, А. И. Особенности обучения рекомендательных систем на примере обучения системы приложения для путешественников / А. И. Курзанова // ИИАСУ'24 – Искусственный интеллект в автоматизированных системах управления и обработки данных: Сборник статей III Всероссийской научной конференции. – Москва : ООО «Издательский дом КДУ», 2025. – С. 223–228.

7. Графовые сети в рекомендательных системах : Хабр. – URL: https://habr.com/ru/companies/wildberries/articles/826422/ (дата обращения: 28.09.2026). – Текст : электронный.

8. Recommender Systems: The Rise of Graph Neural Networks : развитие графовых нейронных сетей. – URL: https://www.shaped.ai/blog/recommender-system-family-tree-gnn (дата обращения: 28.09.2026). – Текст : электронный.

A.P. Sergushicheva, V.V. Petrov
Vologda State University

DEVELOPING A SELF-SUPERVISED GRAPH NEURAL NETWORK FOR A RECOMMENDER SYSTEM OF AN ELECTRONIC LIBRARY

The article discusses the development of a self-learning graph neural network for an electronic library recommendation system. An analysis of modern methods for building recommendation systems is presented, including classical algorithms of collaborative and content-based filtering, as well as their limitations. The choice of graph neural networks as the most promising approach for accounting for multi-connected dependencies between users, books, and metadata is justified. The architecture of the proposed model based on LightGCN with an adaptive node representation update mechanism is described. Directions for practical implementation of the system are formulated.

Recommendation system, graph neural networks, self-learning, electronic library, personalization, LightGCN.