

*Колебцев Владислав Игоревич, студент-магистр, Калужский филиал ФГБОУ
ВО «Московский государственный технический университет имени Н.Э.*

Баумана (национальный исследовательский университет)»

*Гришунов Степан Сергеевич, аспирант, Калужский филиал ФГБОУ ВО
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»*

*Белов Юрий Сергеевич, к.ф. -м.н., доцент, Калужский филиал ФГБОУ ВО
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»*

МЕТОДЫ И ПОДХОДЫ, ИСПОЛЬЗУЮЩИЕСЯ ПРИ ПОСТРОЕНИИ НОВОСТНЫХ СИСТЕМ РЕКОМЕНДАЦИЙ

Аннотация: Технология персонализированных рекомендаций — это инструмент, помогающий пользователям быстро узнать информацию, которая их, скорее всего, заинтересует. Из-за присущих сетевым новостям характеристик, таких как массивность данных, неоднородность, быстрое обновление и изменение, своевременность и сильная географическая осведомленность и так далее, прогресс технологии персонализированных рекомендаций в области новостей отстает от других систем рекомендаций (музыка, кино и др.). Поэтому основной задачей современной системы рекомендаций новостей является интеграция существующих технологий персонализированных рекомендаций в область рекомендаций новостей, изучение способов обработки массивных разнородных новостных данных, построение оптимальной модели предпочтений пользователей и повышение общей производительности, персонализированной рекомендации новостей. В данной статье представлены современные персонализированные новостные

рекомендательные технологии последних лет, а также проанализированы преимущества и недостатки мейнстрим-технологии.

Ключевые слова: рекомендация новостей, метод сбора данных, классификация, персонализированная рекомендация новостей, контекстно-зависимый анализ данных.

Abstract: Personalized Recommendation Technology is a tool to help users quickly learn information they are likely to be interested in. Because of network news's inherent characteristics, such as data massif, heterogeneity, rapid updates and change, timeliness and strong geographic awareness, and so on, the progress of personalized news recommendations technology lags behind other recommendation systems (music, movies, etc.). Therefore, the main objective of today's news recommendation system is to integrate existing personalized recommendation technologies into the area of news recommendations, explore ways to handle massive, diverse news data, build the optimal user preference model, and improve overall performance, personalized news recommendations. This article presents the latest personalized news recommendation technologies of recent years, as well as the advantages and disadvantages of mainstream technology.

Keywords: news recommendation, data collection method, classification, personalized news recommendation, context-dependent data analysis.

Введение. В настоящее время большой объем информации (включая данные, изображения, видео и другой контент) является общедоступным в социальных сетях. Такой значительный обмен информацией вводит проблему информационной перегрузки для пользователей. Таким образом, некоторые веб-сайты используют системы рекомендаций, чтобы преодолеть эту проблему и предложить полезную информацию для целевых пользователей. Рекомендательные системы были широко изучены в середине 1990-х годов. Они используются для рекомендации книг, фильмов, музыки, продуктов и телевизионных программ. Системы рекомендаций могут сократить время и

затраты пользователей и, одновременно с этим, улучшить процесс, качество и стратегию принятия решений для поставщиков контента.

В отличие от двух других инструментов, используемых для решения проблем информационной перегрузки, таких как таксономия и поисковые системы, технология персонализированных рекомендаций не требует от пользователей предоставления явных и точных требований [1]. Вместо этого он строит модель предпочтения интересов пользователей, анализируя их исторические записи поведения и другую соответствующую информацию, а затем активно рекомендует информацию, которая наилучшим образом соответствует их интересам и потребностям пользователя на основе этой модели.

В отличие от рекомендательных систем в области электронной коммерции, кино, музыки и т. д., структура персонифицированных новых рекомендательных технологий является более сложной и трудной из-за особенностей самих новостей (сильная контекстуальная корреляция, быстрые изменения популярности, сильная своевременность исполнения, социальные факторы воздействия и т. д.) и релевантности между новостями (новости не являются независимыми). Кроме того, из-за огромного количества новостной информации в Интернете, большого числа читателей и их быстрого роста, это приносит высокий спрос на массовые возможности обработки данных персонализированной технологии рекомендаций новостей. Хорошая персонализированная система рекомендаций новостей также должна обладать сильной адаптивностью и масштабируемостью, чтобы обеспечить пользователям наименьшую задержку времени отклика в службах рекомендаций новостей [2]. Но сейчас такие новостные рекомендательные системы с высоким рекомендательным эффектом и хорошей производительностью все еще трудно полностью реализовать, используя существующие методы. Таким образом, есть еще аспекты, которые необходимо изучить и улучшить в персонализированной рекомендации новостей.

Далее описаны современные персонализированные новостные рекомендательные технологии последних лет.

Методы персонализированной рекомендации. В настоящее время существует много исследований по персонализированной рекомендации, и появилось много ультимативных методов рекомендации. Наиболее известными методами являются следующие: рекомендация на основе ассоциативных правил (*association rules*), коллаборативная фильтрация (*collaborative filtering*), фильтрация на основе контента (*content-based filtering*), социальная фильтрация (*social filtering*) и гибридная рекомендация (*hybrid recommendation*).

1) Рекомендация на основе ассоциативных правил.

Интеллектуальный анализ ассоциативных правил относится к поиску корреляции между элементами в большом объеме данных. Основная идея состоит в том, чтобы обнаружить правила и паттерны, а также отношения между этими паттернами, содержащимися в данных, которые отвечают определенным уровням поддержки и доверия, путем интеллектуального анализа уже известных данных. Рекомендация, основанная на ассоциативных правилах, заключается в том, чтобы установить соответствующие отношения между элементами и моделями интересов пользователей с помощью различных ассоциативных правил, а затем предсказать интересы пользователей на основе этих отношений, после чего дать рекомендации [3].

Этот метод прямолинеен и прост в реализации. Он также обладает сильной универсальностью в различных областях и лучшей производительностью в режиме рекомендаций реального времени. Однако существует серьезная проблема холодного запуска. Как правило, новый добавленный элемент трудно найти в системе, и он не может быть рекомендован из-за отсутствия данных о взаимодействии с ним. С увеличением количества элементов и пользователей в системе ассоциативные правила быстро увеличиваются в количестве. Таким образом, затраты на обслуживание правил также увеличиваются, что снижает эффективность работы системы.

2) Коллаборативная фильтрация.

Рекомендательная система, использующая коллаборативную фильтрацию основана на предпочтениях группы единомышленников. На основе полученной информации система рекомендует те элементы, которыми определенная группа со схожими интересами уже интересовалась, а конкретный человек — еще нет. Данный метод находит релевантность между элементами или пользователями, анализируя поведение групп пользователей, а затем рекомендует конкретным пользователям похожие элементы [4]. Это самый ранний предложенный и наиболее широко используемый метод в рекомендательных системах. Его основная идея — групповой интеллект. Коллаборативную фильтрацию обычно можно разделить на два метода: коллаборативная фильтрация на основе пользователя и коллаборативная фильтрация на основе элемента.

2.1) Коллаборативная фильтрация на основе пользователя.

Этот метод находит сходства между пользователями с помощью анализа предпочтений пользователей по элементам, а затем дает рекомендации на основе похожих пользователей. В частности, он сначала находит набор пользователей, схожих по интересу с целевым пользователем, в соответствии с данными о конкретных предпочтениях и интересах, а затем отфильтровывает те элементы, которые предпочтительны для аналогичных пользователей, но не были просмотрены целевым пользователем, и, наконец, рекомендует эти элементы целевому пользователю. Точность рекомендаций с использованием коллаборативной фильтрации на основе пользователя высока в случае полного и богатого набора данных. Кроме того, коллаборативная фильтрация на основе пользователя может неявно и прозрачно выявлять релевантность между рекомендуемыми объектами и пользовательскими предпочтениями. Однако с увеличением числа пользователей в системе этот метод требует поддержания огромной матрицы подобию пользователей, что увеличивает время расчета рекомендаций и снижает эффективность системы. Он также не может решить проблему холодного запуска. Новые элементы не смогут быть рекомендованы целевым пользователям из-за отсутствия необходимых данных о поведении пользователей.

2.2) Коллаборативная фильтрация на основе элемента.

Этот метод обнаруживает сходство в элементах, а затем рекомендует похожие элементы на основе существующего выбора пользователя. В частности, коллаборативная фильтрация на основе элемента сначала вычисляет сходство между элементами в соответствии с данными о поведении пользователя, а затем генерирует список рекомендаций для целевого пользователя на основе сходства элементов и поведения пользователя [5]. Данный метод рекомендаций не требует от новых пользователей, данных об их интересах. Как только новые пользователи выбирают тот или иной элемент в системе, им можно рекомендовать те элементы, которые похожи на выбранный элемент. В то же время метод коллаборативной фильтрации на основе элемента способен быстро реагировать на поведение пользователей. Иными словами, новое поведение пользователя (выбор нового элемента, не похожего на остальные) приведет к изменению результата рекомендации в реальном времени. Однако именно потому, что этот метод использует поведение пользователей для изучения сходства между элементами без учета различий интересов разных пользователей, а также не учитывает релевантность контента элементов, точность метода коллаборативной фильтрации на основе элемента становится ниже, чем у коллаборативной фильтрации на основе пользователя [6]. Одновременно с этим он не имеет возможности рекомендовать пользователям новые элементы без обновления таблицы сходства элементов в автономном режиме. Именно поэтому он неэффективен в новостной системе рекомендаций, т. к. новости обновляются очень быстро, намного превышая скорость увеличения числа новых пользователей [7]. Этот метод требует поддержания и обновления огромной таблицы сходства новостей, что делает его очень дорогостоящим и неэффективным.

3) Фильтрация на основе контента.

Фильтрация на основе контента является расширением и развитием коллаборативной фильтрации. Этот метод извлекает и анализирует содержимое объектов рекомендаций, получает интересы пользователей на основе их

исторического поведения и рекомендует пользователям элементы, которые наилучшим образом соответствуют их моделям интересов. Суть этого метода заключается в интеллектуальном анализе содержательных особенностей объектов рекомендаций и построении моделей интересов на основе содержательных особенностей и поведения пользователей.

Этот метод включает в себя три этапа. Во-первых, извлекаются определенные функции для каждого элемента на основе содержимого для представления каждого элемента, называемого текстовым представлением объектов рекомендации. Затем интересы пользователей изучаются с помощью функций наборов элементов, которые нравились или не нравились пользователям в прошлом, данный этап носит название построение модели предпочтений пользователей. Наконец, набор наиболее релевантных элементов выбирается для рекомендации каждому пользователю на основе интересов конкретных пользователей и особенностей элементов-кандидатов, полученных на предыдущих двух шагах, называемых генерацией результатов рекомендаций.

4) Социальная фильтрация.

Если конкретный элемент нравится другим пользователям, имеющим пересечения с целевым пользователем в социальной сети, он может быть рекомендован целевому пользователю, это называется рекомендацией социальной фильтрации. Преимущество этого метода заключается в том, что он не требует взаимодействия между пользователями и элементами, но все же может давать рекомендации для новых пользователей или новых элементов [8]. То есть он способен решить проблему холодного запуска. Это связано с тем, что этот метод использует информацию о социальных сетях пользователей для анализа их предпочтений по интересам, а затем выборочно рекомендует им любимые элементы своих друзей, что несколько похоже на метод коллаборативной фильтрация на основе пользователя, за исключением того, что источники данных различны. Недостатком этого метода является то, что

разреженность данных в социальной сети все еще существует, и он не учитывает контекстуальную информацию.

5) Гибридная рекомендация.

По мере развития различных методов рекомендаций появляется ряд исследований о смешивании этих многочисленных методов. Гибридная рекомендация в основном использует такие стратегии смешивания: взвешенное слияние, комбинация признаков, каскад, увеличение признаков и т. д. для интеграции различные методы рекомендаций. Цель гибридной рекомендации состоит в том, чтобы компенсировать недостатки отдельных методов, максимизировать их преимущества сами по себе [9]. На практике обычно выбираются различные методы рекомендаций и используются соответствующие стратегии для их смешивания в соответствии с конкретными сценариями применения.

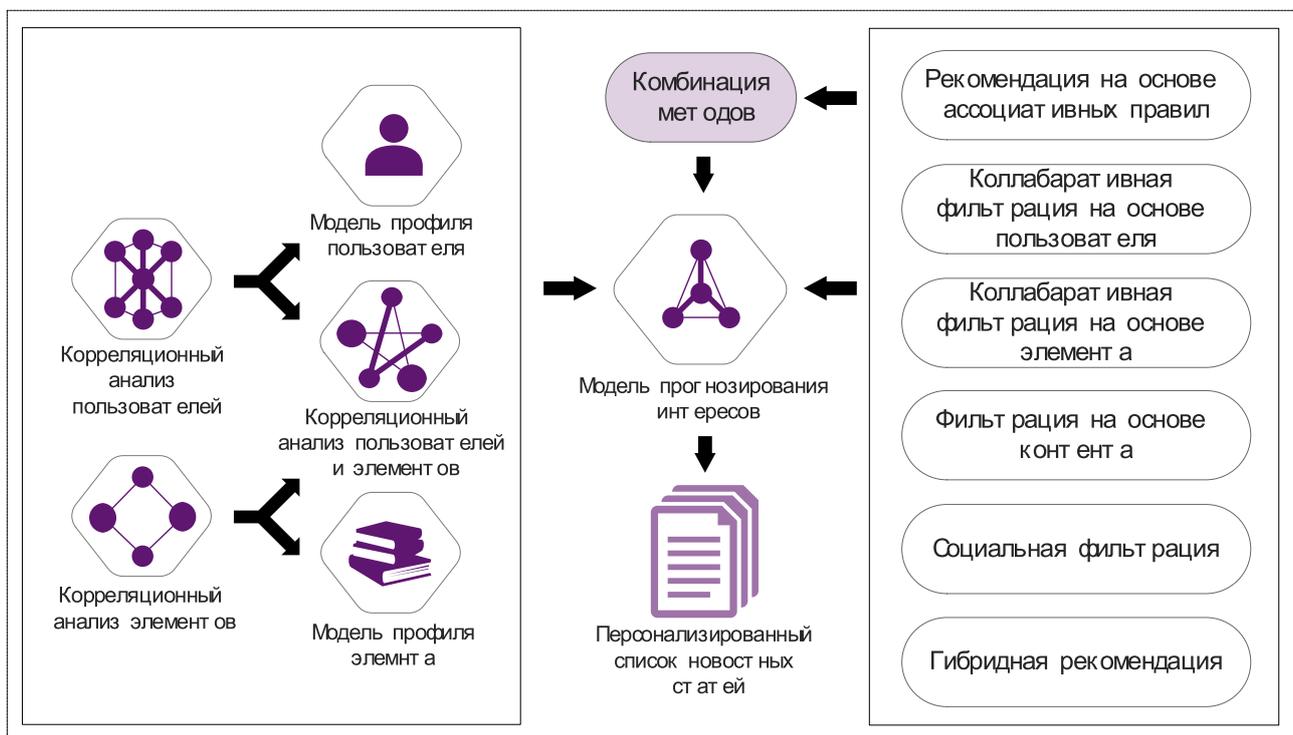


Рис. 1. Общая структура системы рекомендаций новостей, на основе общепринятых методов рекомендаций

С точки зрения текущих технологий рекомендаций, используемых в персонализированных новостных рекомендациях, можно построить блок-

схему, показанную на Рис. 1. Анализируя информацию о пользователе [10] и информацию об элементе (новостной статье), можно получить модель профиля пользователя, модель профиля элемента и модель корреляции пользователь-элемент. Затем модель прогнозирования интересов пользователей может быть извлечена путем обучения. После чего списки рекомендаций новостей могут быть рекомендованы пользователям, которые больше всего заинтересованы в этих новостях, основанных на модели прогнозирования. Более того, рекомендательная система будет показывать разные списки для разных пользователей или групп пользователей, чтобы достичь окончательного эффекта желательной рекомендации. Модель прогнозирования интереса содержит или только один или несколько алгоритмов рекомендаций, перечисленных в правом блоке на Рис. 1. Контекстная информация и информация о социальных сетях часто добавляются для повышения точности рекомендаций и уменьшения негативного влияния проблемы разреженности данных.

Выводы: Роль персонифицированных рекомендаций в процессе цифровизации всех областей интернета также становится все более важной, особенно в области новостей. Одной из важнейших задач сегодня является интеграции, существующих персонализированных рекомендательных технологий в интернет-новости; другая не менее важная задача — построить оптимальную модель предпочтений пользователей, всесторонне учитывающую множество факторов. Эта статья анализирует персонализированные методы рекомендации новостей, надеюсь, что данная информация может оказать некоторую полезную помощь исследователям и инженерно-техническому персоналу предприятий в смежных областях рекомендации новостей.

Библиографический список:

1. Zhang Y. GroRec: A group-centric intelligent recommender system integrating social, mobile and Big Data technologies // IEEE Trans. Services Comput., 2016, vol. 9, no. 5, pp. 786–795, Sep./Oct.

2. Okura S., Tagami Y., Ono S., Tajima A. Embedding-based News Recommendation for Millions of Users. // Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, 1933–1942.
3. Zuo Y., Zeng J., Gong M., Jiao L. Tag-aware recommender systems based on deep neural networks // Neurocomputing, 2016, vol. 204, pp. 51–60.
4. Portugal I., Alencar P., Cowan D. The use of machine learning algorithms in recommender systems: A systematic review // Expert Systems with Applications, 2018, vol. 97, pp. 205-227.
5. Wang Q., Zhang T., Rong Z. Collaborative Filtering Similarity Algorithm Using Common Items// WHICEB, 2017, vol. 16.
6. Клюквин Р.В., Белов Ю.С. Использование косинуса схожести в рекомендательных системах на основе коллаборативной фильтрации // Электронный журнал: наука, техника и образование. 2017. № 2 (12). С. 131-136.
7. Shams B., Haratizadeh S. Graph-based collaborative ranking // Expert Syst. Appl., 2017, vol. 67, no. C, pp. 59-70.
8. Celdran A. H., Perez M. G., Clemente F. J. G., Perez G. M. PRECISE: Privacy-aware recommender based on context information for cloud service environments // IEEE Commun. Mag., 2014, vol. 52, no. 8, pp. 90–96.
9. Wu Y., Qi M., Yang R. A news recommendation system based on an improved collaborative filtering algorithm // Comput. Eng. Sci., 2017, vol. 39, no. 6, pp. 1179–1185.
10. Крюкова Я.Э., Белов Ю.С. Прогнозирование поведения пользователей, основанное на машинном обучении // Актуальные вопросы науки. 2019. № 50. С. 165-167.