http://doi.org/10.35784/iapgos.6203

received: 07.05.2024 | revised: 09.09.2024 | accepted: 16.09.2024 | available online: 30.09.2024

ANALYSIS OF CONTENT RECOMMENDATION METHODS IN INFORMATION SERVICES

Oleksandr Necheporuk[1], Svitlana Vashchenko[1], Nataliia Fedotova[1], Iryna Baranova[1], Yaroslava Dehtiarenko[2]

[1]Sumy State University, Faculty of Electronics and Information Technologies, Department of Information Technology, Sumy, Ukraine,
[2]Lublin University of Technology, Faculty of Electrical Engineering and Computer Science, Lublin, Poland

Abstract. *The object of the research is the process of selecting a content recommendation method in information services. The study's relevance stems from the rapid development of informational and entertainment resources and the increasing volume of data they operate on, thus prompting the utilisation of recommendation systems to maintain user engagement. Considering the different types of content, it is necessary to address the problem of data filtration based on their characteristics and user preferences. To solve this task, we analysed content-based and collaborative filtering methods using various techniques (model-based, memory-based, and hybrid collaborative filtering techniques), knowledge-based filtering, and hybrid filtering methods. Considering each method's advantages and disadvantages, we chose a hybrid method using model-based collaborative filtering and content-based filtering for the future development of our universal recommendation system.*

Keywords: content-based recommender system, collaborative recommender system, hybrid recommender system

ANALIZA METOD REKOMENDACJI TREŚCI W SERWISACH INFORMACYJNYCH

Streszczenie. *Przedmiotem badań jest proces wyboru metody rekomendacji treści w serwisach informacyjnych. Trafność badania wynika z szybkiego rozwoju zasobów informacyjnych i rozrywkowych oraz wzrostu ilości danych, na których działają, dlatego w celu utrzymania uwagi użytkownika wykorzystywane są systemy rekomendacyjne. Biorąc pod uwagę różne rodzaje treści, konieczne jest rozwiązanie problemu filtrowania danych na podstawie ich charakterystyki i preferencji użytkownika. Aby rozwiązać problem, przeanalizowano metody filtrowania treści, filtrowania kooperacyjnego z wykorzystaniem różnych technik (technika oparta na modelu, technika oparta na pamięci i hybrydowa technika filtrowania kolaboracyjnego), filtrowanie oparte na wiedzy oraz metody filtrowania hybrydowego. Biorąc pod uwagę zalety i wady każdej metody, wybrano metodę hybrydową wykorzystującą filtrowanie kolaboracyjne oparte na modelu i filtrowanie oparte na treści do przyszłego rozwoju proponowanego uniwersalnego systemu rekomendacji.*

Słowa kluczowe: system rekomendacji oparty na treści, system rekomendacji oparty na współpracy, hybrydowy system rekomendacji

Introduction

Today, there is a multitude of informational and entertainment services, such as social networks, messengers, streaming platforms, and music services. Considering their commercial interest and the expenses involved in maintaining hardware and software, most of them have advertising and/or a paid subscription system to monetise their businesses. Thus, there is a need for a mechanism that encourages users to utilise the resource for as long as possible. For this purpose, various content recommendation models exist. However, taking into account their diversity and the limited applicability of each of them, there is a need to analyse the methods used in existing models for the future development of a universal model capable of meeting the needs of any information services, regardless of the content suggested to the end user.

1. Review of recent research and publications

Modern information services and entertainment platforms with different types of content were analyzed during the study of content recommendation methods. In the work [1], the recommendation system of the Netflix movie and series streaming service (NRS) was investigated using reverse engineering, and the main principles of its operation were detained. The study's authors [2] compared various collaborative filtering algorithms and proposed using the singular value decomposition approach to improve the accuracy of NRS prediction. The article [3] describes the basic principles of operation and methods of recommendation systems, particularly YouTube. At the same time, the research [4] provides a detailed description of the recommendation diversification model based on Determinantal Point Processes (DPP). The Meta corporation has its own open-source deep learning recommendation model (DLRM) [5] integrated into Facebook. The authors [6] developed a ranking system for Facebook news feed items using the Group-wise Follow-the-Regularized-Leader (G-FTRL) learning algorithm. The research [7] proposed a hybrid recommendation model for Instagram, considering images' visual similarity detected by a neural network. An article [8] provides a detailed description of collaborative filtering principles using the example of the Spotify music service. While Reddit, an entertainment-informational service, does not have a recommendation system, research [9] describes methods for predicting the popularity of individual posts on the platform using machine learning techniques. Many studies are dedicated to comparing content personalisation methods [10–14], as well as their improvement [15, 17, 18].

Authors [19] investigated the creation of a universal content recommendation model, proposing the use of deep GRU networks. At the same time, [12] utilised a classical item-based collaborative filtering algorithm with its inherent "cold start" problem.

The study aims to analyse content recommendation methods in contemporary information and entertainment services, taking into account the diversity of data types of recommended items.

2. Analyze of Content-based filtering methods

2.1. Classification of filtration methods

Modern digital distribution services, online stores, social networks, music platforms, news websites, and other information and entertainment resources share a common goal - to attract more users and increase interest in their product or service. On the one hand, attracting potential users is a relatively easy task, having interesting and, most importantly, necessary content, as targeted advertising and marketing technologies are generally well-developed and popular. However, on the other hand, it is not enough to simply attract people to a specific website; it is also necessary to satisfy their needs so that, based on a positive experience, users return to the service in the future. Users have different needs depending on the service type, such as purchasing a specific product, watching a movie, or finding relevant news. There are recommendation systems to enhance user experience with such resources. They help find the necessary information while considering the content consumer's needs and preferences. Examples of such systems include recommendations of similar products, automatic music playlists, unified social media feeds or news, and predictions of films or series that you may like. This functionality is utilized in the most popular modern web services.

Thanks to recommendation systems, the convenience of using a service is enhanced, thereby improving the resource's competitiveness. Given this, a lot of recommender systems have been created for specific tasks. However, they can be classified according to basic filtering methods (Fig. 1) [10].

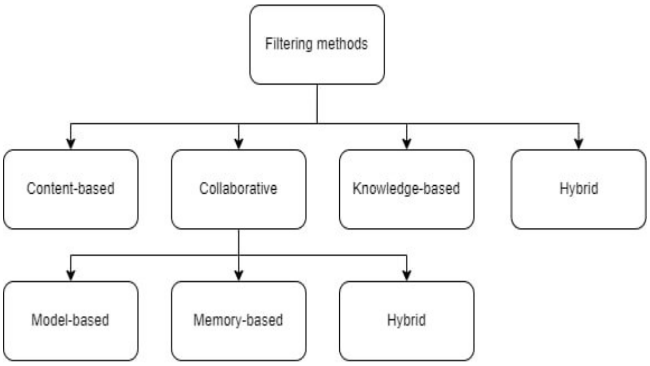

Fig. 1. Classification of content filtering methods

2.2. Method of content filtering

Content-based filtering is based on the user's interaction with the elements of the service, utilizing the objects' properties to improve future recommendations. Based on ratings or implicit indicators (such as viewing time, comments left, or "share" button usage), the system collects and accumulates data to create a list of preferred content features. Considering the gathered data, a dataset is formed with properties likable to the user. The method is schematically depicted in Fig. 2.

Content-based recommender systems use various models to find similarities between elements, such as vector-based models (for example, Term Frequency Inverse Document Frequency-TFIDF; latent semantic analysis [21]) or probabilistic models (decision trees; neural networks or Naive Bayes Classifier) [11]. Properties used for similarity analysis may include object category, genre (of a movie), author, duration (of video or audio), textual description, and other metadata. In addition to this data, user data may also be used, such as region, gender, age, and others, if provided by the service.

Considering the characteristics of the content-based filteringmethod, there is a dependency between the type of content and the associated metadata inherent to it. Therefore, there is a scalability issue in recommendation systems. As a result, its usage is often limited to specific domains, making the method narrow-focused [22].

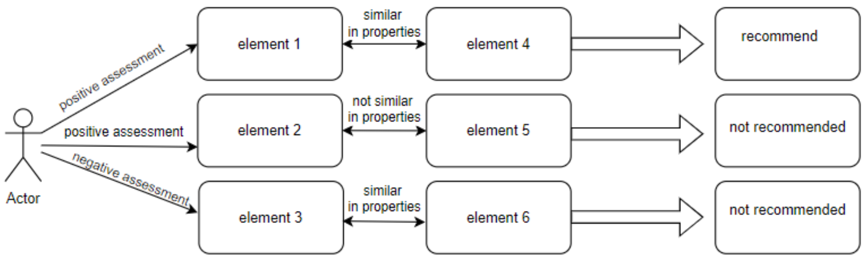

Fig. 2. Scheme of recommendation using a content filtering method

2.3. Collaborative Filtering Method

Unlike content-based filtering, the collaborative filtering method accumulates and uses data on user interaction with system elements to predict recommendations. This method is popular among online stores because rating and reviews systems are prevalent in e-commerce services, making the method more efficacious [23].

According to the algorithms used, collaborative filtering methods are classified as Memory-Based, Model-Based, and hybrid methods.

Memory-based techniques (also known as neighbourhood-based approaches) rely on finding similar elements or users based on their usage experience. The system stores user-item relationships in an interaction matrix, from which K nearest neighbours (K-Nearest Neighbours) are selected, and N objects are generated based on their weighted average ratings [14]. Notably, explicit user ratings are used within this technique, represented by a rating scale most often.

Model-based collaborative filtering methods use a specific model of user interaction with content. These models are based on statistical methods or machine learning techniques, such as clustering, Bayesian, dimensionality reduction, probabilistic matrix factorization, and latent variable models [14]. The technique involves constructing a model based on user rating data or a system dataset for further recommendations without referring to the actual content data during service usage.

The hybrid technique of collaborative filtering combines approaches from memory-based methods with those based on models to address the drawbacks of both approaches and enhance the accuracy of recommendations. However, it is more complex to implement. The main drawback of collaborative filtering methods is the "cold start" problem [12], which arises due to the lack of information about preferences for new users or the scarcity of features for new items in the system. Additionally, there are problems with data sparsity, as many users do not provide explicit ratings, resulting in sparse rating matrices that complicate finding the nearest neighbours and forming recommendations. Model-based systems help alleviate the negative impact of data sparsity by incorporating implicit item evaluations. Furthermore, the problem of "grey sheep" is prevalent, wherein a user's preferences do not align with those of their neighbours.

One of the methods to solve the "cold start" and "grey sheep" problems is knowledge-based filtering. This method allows to recommend new items in the system, or those that do not have enough interactions or user rating, by using deep knowledge of the subject area and user requirements. Knowledge-based recommendations are generated based on two approaches: case-based and constraint-based. In the first case, objects similar to those that meet the user requirements are recommended, formed from experience and represented by "cases", which, in turn, consist of a "problem-solution" pair. Regarding the second approach, recommendations are formed using rules that define the relationship between user requirements and object characteristics. Such recommendation approaches are suitable for systems where the user is willing to spend more time defining rules to obtain more accurate and personalized recommendations [24].

Thus, a knowledge-based method addresses the problem of diversity, enabling the broadening of recommendations. However, it is narrow-focused and requires the creation of rules or case formation dependent on the service domain and demands deep knowledge about the recommended objects.

2.4. Hybrid Filtering Method

Hybrid methods involve using two or more methods within a single recommendation system to reduce the disadvantages of individual methods while simultaneously leveraging their strengths. For example, this could entail combining content-based and collaborative filtering methods or integrating collaborative filtering with a knowledge-based approach, among other combinations.

Work [14] mentions three aspects of the operation of the hybrid model: content and collaborative filtering work independently of each other, and their evaluations are combined; some characteristics of collaborative filtering are used in the content model, and the content model calculates the final rating; or the use of content-based filtering characteristics in the collaborative model to determine recommendations. However, in work [16], seven aspects of using the hybrid model are mentioned, such as combining ratings of each method using weights for each of the results; switching between models depending on the situation (which in turn adds flexibility to services with different types of represented content); simultaneous display of elements using different methods within the same service; combining functions of different approaches into a single recommendation; cascading refinement of the results of one method by another; expanding the array of input data of one approach by the output results of another; and the meta-level, where the entire model created using one method is used by another.

It is evident the use of multiple filtering methods within a single recommendation system allows for the usage of all available information regarding both content units and the entire service system. Hence, there are discrepancies in the classification of hybrid recommendation system usage aspects.

Along with the wide variability in the use of hybrid methods comes complexity in implementation, namely developing a mathematical model, considering all necessary parameters and factors during implementation, and scalability of the developed system. Considering the wide possibilities of using different approaches and system data, it is possible to develop a custom recommendation system that will be free from the main known drawbacks of individual content filtering methods while simultaneously leveraging their advantages. At the same time, using data regarding the type and properties of content units allows for flexible use of the developed model without being limited solely to knowledge about content or interaction between users and system elements.

Therefore, for the future developed recommendation system, which will be designed as an API for web service developers, it is advisable to use a hybrid method, combining collaborative filtering based on the model of user interaction with content, combined with content-based filtering considering the content properties including searching for similar items based on both explicit and implicit indicators.

3. Conclusions

The result of the research is a description and analysis of methods used for content filtering and recommendation formation in recommendation systems. The usage of recommendation systems in modern information services and entertainment platforms with different types of content was analyzed. Classification of content filtering methods based on basic filtering techniques was presented, detailing the operation principle of each (content-based filtering, collaborative filtering, knowledge-based filtering, and hybrid filtering methods) with their advantages and disadvantages. To meet the need for a filtering method for further development of a universal recommendation system that can be used by any service regardless of the content type to recommend, a hybrid content recommendation model combining collaborative filtering based on the model and content-based filtering was chosen.

References

[1] Akbar A., Agarwal P., Obaid A. J.: Recommendation engines-neural embedding to graph-based: Techniques and evaluations. International Journal of Nonlinear Analysis and Applications 13(1), 2022, 2411–2423 [https://doi.org/10.22075/IJNAA.2022.5941].

[2] Aldossary A.: Recommender Systems Principles and methods in Web 2 Applications An analytical view of the filters used in YouTube. Multi-Knowledge Electronic Comprehensive Journal For Education And Science Publication (MECSJ), 37, 2020.

[3] Ameen A.: Knowledge based Recommendation System in Semantic Web-A Survey. International Journal of Computer Applications 182(43), 2019, 20–25.

[4] Bertini M. et al.: Keeping up with the Influencers: Improving User Recommendation in Instagram using Visual Content. UMAP 2020 Adjunct – Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization, 2020, 29–34 [https://doi.org/10.1145/3386392.3397594].

[5] Chu W. T., Tsai Y. L.: A hybrid recommendation system considering visual information for predicting favorite restaurants. World Wide Web 20(6), 2017, 1313–1331 [https://doi.org/10.1007/S11280-017-0437-1/METRICS].

[6] Gohari F. S., Tarokh M. J.: Classification and Comparison of the Hybrid Collaborative Filtering Systems. International Journal of Research in Industrial Engineering 6(2), 2017, 129–148 [https://doi.org/10.22105/RIEJ.2017.49158].

[7] Korotaev A., Lyadova L.: Method for the development of recommendation systems, customizable to domains, with deep GRU network. Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management – IC3K 2018, 2018, 231–236 [https://doi.org/10.5220/0006933302310236].

[8] Kwan C., Koh's M. Q., Jasser M. B.: A Comparison Study Between Content-Based and Popularity-Based Filtering via Implementing a Book Recommendation System. International Journal Of Advanced Research In Engineering & Technology 11(12), 2020, 1121–1135 [https://www.researchgate.net/publication/348190964_A_Comparison_Study_Between_Content-Based_and_Popularity-Based_Filtering_via_Implementing_a_Book_Recommendation_System] (available: 10.05.2016)

[9] Madathil M.: Music Recommendation System Spotify-Collaborative Filtering Mithun Madathil, 2017.

[10] Mohammed Al Mani I. A.: Collaborative filtering recommendation system: Comparison study (M.Sc. Thesis, Altinbaş University). 2018. [http://openaccess.altinbas.edu.tr/xmlui/handle/20.500.12939/1724] (available: 10.05.2016).

[11] Mustafa N. et al.: Collaborative filtering: Techniques and applications. Proceedings of International Conference on Communication, Control, Computing and Electronics Engineering – ICCCCEE 2017 [https://doi.org/10.1109/ICCCCEE.2017.7867668].

[12] Naumov M. et al.: Deep Learning Recommendation Model for Personalization and Recommendation Systems. 2019.

[13] Ni X. et al.: Feature selection for Facebook feed ranking system via a group-sparsity-regularized training algorithm. Proceedings of International Conference on Information and Knowledge Management, 2019, 2085–2088 [https://doi.org/10.1145/3357384.3358114].

[14] Nilashi M., Ibrahim O., Bagherifard K.: A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques. Expert Systems with Applications 92, 2018, 507–520 [https://doi.org/10.1016/J.ESWA.2017.09.058].

[15] Pajkovic N.: Algorithms and taste-making: Exposing the Netflix Recommender System's operational logics. Convergence 28(1), 2022, 214–235 [https://doi.org/10.1177/13548565211014464].

[16] Parfenenko Yu., Kovtun A., Verbytska A.: Recommended information system for video search. Scientific journal "Transactions of Kremenchuk Mykhailo Ostrohradskyi National University" 5(118), 2019, 97–102 [https://doi.org/10.30929/1995-0519.2019.5.97-102].

[17] Patil M., Brid S., Dhebar S.: Comparison Of Different Music Recommendation System Algorithms. International Journal of Engineering Applied Sciences and Technology 5 (6), 2020 [https://doi.org/10.33564/IJEAST.2020.v05i06.036].

[18] Raghuwanshi S. K., Pateriya R. K.: Recommendation systems: Techniques, challenges, application, and evaluation. Advances in Intelligent Systems and Computing 817, 2019, 151–164 [https://doi.org/10.1007/978-981-13-1595-4_12/COVER].

[19] Ranjan A. A. et al.: An Approach for Netflix Recommendation System using Singular Value Decomposition. An International Research Journal 10(4), 2019, 774–779 [https://doi.org/10.29055/jcms/1063].

[20] Shahbazi Z., Byun Y.-C.: Improving the Product Recommendation System based-on Customer Interest for Online Shopping Using Deep Reinforcement Learning. Soft Computing and Machine Intelligence Journal 1(1), 2021 [https://doi.org/10.22995/scmi.2021.1.1.02].

[21] Shuaibi A.: Predicting the Popularity of Reddit Posts. 2019.

[22] Wilhelm M. et al.: Practical Diversified Recommendations on YouTube with Determinantal Point Processes. ACM Reference Format: Gillenwater, 2018. [https://doi.org/10.1145/3269206.3272018].

[23] Zarzour H. et al.: A new collaborative filtering recommendation algorithm based on dimensionality reduction and clustering techniques. 9th International Conference on Information and Communication Systems – ICICS 2018, January 2018, 102–106 [https://doi.org/10.1109/IACS.2018.8355449].

[24] Mashkovskyi S.: Latent-semantic analysis, social networks and non-structured data: interaction method. Visnyk Universytetu "Ukraina" 2(23), 2019 [https://doi.org/10.36994/2707-4110-2019-2-23-29].

M.Sc. Oleksandr Necheporuk
e-mail: oleksandr.necheporuk@student.sumdu.edu.ua

Ph.D. student of Department of Information Technology, Sumy State University, Ukraine.
Research interests: software development, advanced system modeling, and the integration of artificial intelligence to craft intelligent and efficient application software.

https://orcid.org/0000-0002-9905-031X

Ph.D. Svitlana Vashchenko
e-mail: s.vashchenko@cs.sumdu.edu.ua

Associate professor in Department Information Technology, Sumy State University, Ukraine. Author and co-author of more than 50 scientific papers. Research interests: software engineering, algorithmization, use of modern information technologies (in particular, system modeling, structural-functional analysis) for the development of application software in various fields of activity.

https://orcid.org/0000-0002-7021-2629

Ph. D. Nataliia Fedotova
e-mail: n.fedotova@cs.sumdu.edu.ua

Associate Professor of the Department of Information Technology, Sumy State University, Ukraine. Author and co-author of more than 30 scientific papers. The author`s research area focuses on e-learning systems, information technologies of design and management in complex systems, 3D modeling, visualization and animation, real-time computer graphics and simulations.

https://orcid.org/0000-0001-9304-1693

Ph.D. Iryna Baranova
e-mail: i.baranova@cs.sumdu.edu.ua

Associate professor of the Department of Information Technology, Sumy State University, Ukraine. Author and co-author of more than 30 scientific papers. Research interests: 3D modeling, visualization, animation, software application for the development of systems of automated design, e-learning systems.

https://orcid.org/0000-0002-3767-8099

Yaroslava Dehtiarenko
e-mail: s99277@pollub.edu.pl

Student of the Faculty of Electrical Engineering and Computer Science, Lublin University of Technology, Poland. The author`s research area focuses on optimization methods, human-computer interaction, real-time systems, computer graphics and simulations.

https://orcid.org/0009-0004-9455-7092