Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/64/20241444

Exploring the landscape of recommendation systems: A qualitative analysis of types, challenges, and potential solutions

Ruiyan Gao

The Hong Kong Polytechnic University

gry050201@outlook.com

Abstract. With the rapid development of computers and the Internet, recommendation systems have become an integral part of our daily lives, providing personalized suggestions for movies, products, etc. The aim of this paper is to summarize the types of recommendation systems, which are mainly classified into content-based recommendation systems and collaborative filtering-based recommendation systems. Content-based recommendation systems recommend products based on user interests and product descriptions, while collaborative filtering systems recommend products based on similarities between users or products. However, these systems face a number of challenges, including the cold-start problem, data sparsity, and scalability issues. The cold-start problem refers to the difficulty of making accurate recommendations for new users or products with limited information. To address this problem, researchers have proposed methods such as using user aspect models and integrating content information with collaborative filters. Data sparsity is another issue that affects the accuracy of the system when there is not enough data to infer user preferences. This can be mitigated by leveraging user behaviour from social networks and using a hybrid recommendation approach. Scalability is also an issue, as the system must handle exponential data growth while maintaining performance. Technologies such as user clustering and cloud computing are recommended for elastic data storage and scalable processing power. In conclusion, although recommendation systems have made significant progress, they still face challenges that need to be addressed. This paper qualitatively analyses these issues and compares different approaches proposed by researchers, providing a valuable resource for understanding the current state of recommendation systems and future directions.

Keywords: recommendation system, content-based recommendation, collaborative filtering based recommendation system, cold start problem.

1. Introduction

With the development of computers and computer networks, people's lives have undergone a series of changes. Most people will download many kinds of apps, such as shopping apps or video apps, and they will use the search engine in the apps to collect movies or their favourite products, etc., and they will also rate and comment on the apps after using them, thus generating a large amount of data. However, with the increase of data, there is an urgent need for a way to organize and categorize a vast amount of data, and recommendation systems were invented.

Gradually, people realize the importance of recommendation systems, which provide a kind of mass customization on the Internet [1]. In real life, for users, this kind of customization reduces the searching

DOI: 10.54254/2755-2721/64/20241444

workload, and for merchants, if they can accurately recommend the products that the users like, they can promote consumption and increase turnover. At the same time, the recommendation system can categorize different data, which indirectly reduces the amount of data processing and avoids the problem of data redundancy.

This paper will summarize the types of recommendation systems and the problems faced by recommendation systems at this stage. For the types of recommendation systems, this research will divide them into content-based recommendation systems and collaborative filtering-based recommendation systems. For the problems of recommendation systems, there are three main ones, cold start, data sparsity, and scalability.

Goyani and Chaurasiya authored a literature review titled ‘A Review of Movie recommendation Systems: Limitations, Surveys and Challenges’ [2]. This article explains a lot about the classification of recommendation systems and the challenges encountered, and also uses a lot of difficult mathematical formulas and data analysis to present the algorithms of recommendation systems, which are relatively difficult to understand. In contrast to that article, although the structure of the article is largely the same, this paper introduces only two widely used recommendation systems and the problems they may encounter, and provides qualitative analyses and comparisons of the theoretical results of different scientists without using large amounts of data for extrapolation, making them easy to understand. In addition to scientific analyses, this paper will compare recommendation systems with real-life examples, such as the solutions and pros and cons of existing apps for the cold-start problem mentioned later in the paper

2. Background

As Konstan and Riedl said, with the continuous development of computers, the recommendation systems came into being in the early nineties, which also provided personalized predictions for some large products [3]. Later, in 1997, Resnick and Varian gave a specific definition of the recommendation system [4]. People's recommendations are used as inputs in a standard recommendation system, which subsequently combines and sends the information to the appropriate receivers. In addition, as recommendation systems have fully entered people's lives, people are trying to understand their principles and broader applications. It is highly likely that recommendation systems are considered by most people as predicting and summarizing data. In fact, more methods have been applied to recommendation systems. According to Zhang et al., Deep Learning (DL) has performed well in the fields of computer vision and natural language processing in recent years, and is therefore widely regarded as having a promising application in the field of recommendation systems [5]. Meanwhile, Afsar et al. study also pointed out the sequential nature of the recommendation problem and suggested that Markov Decision Process (MDP) and Reinforcement Learning (RL) methods can be used to solve the problem [6]. In addition, they also proposed that the latest research combines RL with DL to form DRL, thus making it possible to apply reinforcement learning to recommendation problems with large state and action spaces.

3. Classification of recommendation systems

The growing interest in recommendation systems has led to numerous categorizations proposed by various researchers in the field. One prominent classification by Adomavicius and Tuzhilin identifies three primary types of recommendation systems based on their methodology: content-based recommendations, collaborative recommendations, and hybrid recommendations [7]. Concurrently, Bobadilla et al. outline four main filtering algorithms: content-based filtering, collaborative filtering, hybrid filtering, and demographic filtering [8]. Upon synthesizing these categorizations, it becomes evident that the two most representative types of recommendation systems are content-based recommendations and collaborative filtering-based recommendations.

Content-based recommendation systems rely on the characteristics and attributes of items to provide personalized suggestions [7]. By analyzing users' preferences and interests, these systems recommend items that exhibit similar features to those previously favored by the user. On the other hand,

DOI: 10.54254/2755-2721/64/20241444

collaborative filtering-based recommendations adopt a more social approach, basing suggestions on the collective preferences of users with similar tastes or behaviors [8]. By leveraging the wisdom of the crowd, these systems aim to predict users' preferences and provide tailored recommendations that align with their unique inclinations.

3.1. Content-based recommendation

According to Pazzani and Billsus, the content-based recommendation is to recommend corresponding products to users based on the description of the item and the user's interest [9]. The description of the item can be understood as quantitative, because an item will not change a significant number in a short time, so this research mainly study the user's behaviour. A user's behaviour is essentially a profile of a user in a database. Pazzani and Billsus believe that this study should mainly focus on two models [9]. One is the model of user preference. According to Jawaheer et al., people can think of it as explicit user feedback [10]. The model focuses on a function that predicts the n products that the user is most interested in, based on the user's preference for certain products. The other is the history of the user's interaction with the recommendation system, also known as implicit user feedback, which is mainly analyzed based on the user's time spent browsing the page, search history, and so on. The research on explicit and implicit feedback goes beyond this. Jawaheer et al. also propose a classification framework based on a set of different attributes including cognitive effort, user model, measurement scale and domain relevance for both forms of feedback [10].

3.2. Collaborative filtering based recommendation system

The idea of collaborative filtering is essentially based on the similarity between users or the similarity between subjects. Specifically, according to Yang et al., the algorithm of a collaborative filtering-based recommendation system mainly includes three steps [11].

3.2.1. Data collection/alteration

In this step, the collected data mainly includes four main categories:demographic data, production data, user behaviour and user rating [11]. Data collection is mainly reflected in the personal information entered by the user when using a new APP, which will be directly entered into the data back end for processing and categorization. At the same time, most APPs provide some fields for new users to choose from in the initial interface, and according to the fields chosen by the new users, they will be recommended in accordance with the degree of heat and the favourite degree of other users, which can increase the browsing volume of the user to get more However, this initial collection method also has great limitations, for example, with the passage of time, users may change their preferences, so most APPs have an editing interface, which is convenient for users to change their favourite content at any time, and carry out repetitive collection and change of data.

3.2.2. Data pre-processing

According to García et al., data pre-processing is mainly divided into transformation, integration, cleaning and normalization [12]. Some users may not provide part of the information, such as age, for convenience or privacy protection, so the APP has to delete these NA values when analyzing the data, and then carry out analyses such as outlier processing and chi-square test in order to get accurate results.

3.2.3. Collaborative Filtering

In this stage, the existing information is utilized to forecast the absent data; however, it is crucial to differentiate between the missing information in this context and the NA values encountered during the pre-processing phase. Here, the term "missing information" pertains to the lack of usage records for new users, rather than the null values addressed earlier in the process. The predictions derived from this filtering process are subsequently relayed back to the user interface, ultimately enhancing the accuracy of the results and promoting higher levels of user satisfaction.

DOI: 10.54254/2755-2721/64/20241444

Although both content-based recommendation systems and collaborative filtering-based recommendation systems essentially use known information to predict unknown information, Hameed et al. believe that collaborative filtering-based recommendation systems are better, and they believe that content-based recommendation systems suffer from two main problems: the first one is that some items do not have intrinsic content and cannot be accurately categorized in this way, and the second is that they can only return user comments related to the content, which has a smaller scope [13]. However, collaborative filtering based recommendation systems do not need to analyse content or store large amounts of data for each user, it only focuses on clustering algorithms. As a result, the analysis is more efficient and broader in scope. In fact, the content-based filtering approach is a good solution to the cold-start problem that collaborative filtering-based recommendation systems may encounter, as it can make recommendations based on the labeling and categorization of items without the need for access history.

4. Problems of recommendation system

Despite the fact that recommendation systems have been evolving for decades and have become more and more accurate as algorithms have been changed, there are still many problems with recommendation systems according to Goyani and Chaurasiya, such as the cold-start problem mentioned above, as well as problems with data sparsity and scalability [2].

4.1. Cold start problem

Cold start problem is one of the biggest problems encountered by the current recommendation system, It refers to the initial stage, the user and the product directly without direct interaction, so the system can not filter and classify different information, which may lead to the system recommended is not the user's favourite, and may even cause the user's disgust, which leads to the loss of some users. Cold start problems are mainly divided into three categories.

4.1.1. user cold start

The main reason for this problem is that newly registered users have not used the software, which can lead to blank user information, so it is impossible to filter recommendations.

4.1.2. Commodity cold start

This is due to the fact that the newly launched products do not have enough users to use and evaluate, so it is impossible to judge the product can be recommended to what group of people.

4.1.3. System Cold Start

This is the most serious cold start problem, there are not enough users nor enough data on the interaction between the product and the users.

4.1.4. Solution of cold start problem

There are already a great number of methods to solve the cold start problem, for example, Lam et al., proposed that the user's aspect model can be used to predict the rating of new users, and the method also modifies the Hoffman model [14]. Not only that, Leung et al. uses a hybrid recommendation approach that integrates content information from domain entries into collaborative filters using Cross-Level Association RulEs (CLARE) [15]. Therefore, for software, not only can it adopt the previously mentioned approach of letting users set their own favourite domains, but it can also use the user's already existing social network account to log in and analyse the user's behaviour on social networking sites to make recommendations.

4.2. Data sparsity

The data sparsity problem refers to the fact that there is not enough transaction and feedback data to infer the similarity of a particular user, thus affecting the accuracy and performance of the recommendation system [16]. For example, some merchants have many items, but it is impossible for

DOI: 10.54254/2755-2721/64/20241444

users to buy or browse all of them, then there are some items that cannot participate in the filtering of the recommendation system, resulting in the problem of data sparsity, which can be seen as a special kind of data sparsity problem, and the new items can also be regarded as not participating in the interactions with the users. Therefore, the solution to the cold start problem can also be the solution to the data sparsity problem. Not only that, integrated learning is also one of the solutions to the data sparsity problem, which can fuse the results of multiple models as a way to reduce the analysis error of a single model.

4.3. Scalability

The scalability of data inherently means that as the amount of information grows exponentially, processing may become inaccurate while trying to ensure data integrity and performance. However, current recommendation systems are already supported by powerful algorithms, so solving this problem is relatively simple. Georgiou and Tsapatsoulis propose the use of similar user clustering, which can effectively narrow down the scope of the data [17]. Not only that, but also the use of cloud services and resource management can be utilized to achieve elasticity of data storage and scalability of processing power by using cloud computing to allow dynamic allocation and release of resources. Furthermore, the simplest way is to delete part of the useless data. In real life, some merchants will observe the monthly or yearly profit of the product to decide whether to buy the product or not, and the product with a lower profit will be removed from the merchant, which not only saves the cost, but also reduces the amount of data and facilitates the data processing.

In addition to the above methods that can solve the problems that recommendation systems may encounter, there are some methods that have not been proven. For example, according to da Silva et al., the impact of matrix analysis and deep learning autoencoders on data sparsity cannot be determined yet [18]. Therefore experiments and analyses on these doubtful issues are still needed in future scientific research.

5. Conclusion

In summary, this paper outlines the types of recommendation systems, mainly content-based systems and collaborative filtering-based systems. The paper also discusses the major issues facing recommendation systems, including the cold-start problem, data sparsity, and scalability. For these three problems, the cold-start problem is mainly introduced because it is more difficult to deal with the cold-start problem as compared to the data scalability problem, while for the data sparsity problem, the cold-start problem is its special case, and the two are more similar.

Despite these challenges, recommendation systems have become an integral part of our daily lives, providing personalized advice and improving the user experience. This paper highlights the importance of solving the problem of recommendation systems and qualitatively analyses the theoretical results of different scientists, which is a valuable source for understanding the current state of recommendation systems and future directions.

References

[1] Ansari, A., Essegaier, S., & Kohli, R. (2000). Internet recommendation systems.
[2] Goyani, M., & Chaurasiya, N. (2020). A review of movie recommendation system: Limitations, Survey and Challenges. ELCVIA: electronic letters on computer vision and image analysis, 19(3), 0018-37.
[3] Konstan, J. A., & Riedl, J. (2012). Recommender systems: from algorithms to user experience. User modeling and user-adapted interaction, 22, 101-123.
[4] Resnick, P., & Varian, H. R. (1997). Recommender systems. Communications of the ACM, 40(3), 56-58.
[5] Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. ACM computing surveys (CSUR), 52(1), 1-38.

DOI: 10.54254/2755-2721/64/20241444

[6] Afsar, M. M., Crump, T., & Far, B. (2022). Reinforcement learning based recommender systems: A survey. ACM Computing Surveys, 55(7), 1-38.

[7] Adomavicius, G., & Tuzhilin, A. (2010). Context-aware recommender systems. In Recommender systems handbook (pp. 217-253). Boston, MA: Springer US.

[8] Bobadilla, J., Ortega, F., Hernando, A., & Bernal, J. (2012). A collaborative filtering approach to mitigate the new user cold start problem. Knowledge-based systems, 26, 225-238.

[9] Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In The adaptive web: methods and strategies of web personalization (pp. 325-341). Berlin, Heidelberg: Springer Berlin Heidelberg.

[10] Jawaheer, G., Weller, P., & Kostkova, P. (2014). Modeling user preferences in recommender systems: A classification framework for explicit and implicit user feedback. ACM Transactions on Interactive Intelligent Systems (TiiS), 4(2), 1-26.

[11] Yang, Z., Wu, B., Zheng, K., Wang, X., & Lei, L. (2016). A survey of collaborative filtering-based recommender systems for mobile internet applications. IEEE Access, 4, 3273-3287.

[12] García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., & Herrera, F. (2016). Big data preprocessing: methods and prospects. Big data analytics, 1, 1-22.

[13] Hameed, M. A., Al Jadaan, O., & Ramachandram, S. (2012). Collaborative filtering based recommendation system: A survey. International Journal on Computer Science and Engineering, 4(5), 859.

[14] Lam, X. N., Vu, T., Le, T. D., & Duong, A. D. (2008, January). Addressing cold-start problem in recommendation systems. In Proceedings of the 2nd international conference on Ubiquitous information management and communication (pp. 208-211).

[15] Leung, C. W. K., Chan, S. C. F., & Chung, F. L. (2008). An empirical study of a cross-level association rule mining approach to cold-start recommendations. Knowledge-Based Systems, 21(7), 515-529.

[16] Idrissi, N., & Zellou, A. (2020). A systematic literature review of sparsity issues in recommender systems. Social Network Analysis and Mining, 10(1), 15.

[17] Georgiou, O., & Tsapatsoulis, N. (2010, September). Improving the scalability of recommender systems by clustering using genetic algorithms. In International conference on artificial neural networks (pp. 442-449). Berlin, Heidelberg: Springer Berlin Heidelberg.

[18] da Silva, J. F. G., de Moura Junior, N. N., & Caloba, L. P. (2018, July). Effects of data sparsity on recommender systems based on collaborative filtering. In 2018 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.