

Application of recall methods in recommendation systems

Yumeng Wang

Zhejiang University, Hangzhou, China

3190103183@zju.edu.cn

Abstract. In order to have a more comprehensive introduction and understanding of the research progress of recall strategies in recommender systems, this paper reviews the application of diverse recall methods in various recommender systems by different researchers. By searching and reading literature in major databases like Google Scholar, it is found that the recall method suitable for news recommendation system is also generally applicable in other recommendation systems. Therefore, this paper takes news recommendation system as an example to introduce traditional content-based recall and collaborative filtering-based methods. Hot-based recall and Embedding-based recall also developed in recent years. Furthermore, recall strategies (emotion-based recall and UIBB) that are specifically applicable to music and e-commerce recommendation systems are introduced. This paper briefly introduces these recall styles and collects researchers' evaluations and attitudes towards these recall styles, aiming to provide help for recommender system designers in optimizing recall methods.

Keywords: Recommendation System, Content-based Recall, Collaborative Filtering, Hot-based Recall, News Recommendation.

1. Introduction

In the era of big data, predicting user behavior and recommending content that users may be interested in is an essential function of software. Recall is a very important part of the recommendation system and there are different recall methods in different recommendation systems. In this regard, various researchers worked on this topic, such as Paul Covington et al. [1], M. Deshpande et al. [2], Peter Brusilovsky et al. [3], Adrián Valera et al. [4], Yousefian Jazi et al. [5], Farah Tawfiq Abdul Hussien et al. [6], Singh Mahesh Kumar et al. [7] and Zeqi Ruan et al. [8].

Recall is very important in recommendation systems. By filtering and sorting the data in the database, the materials that users are most likely to click, browse or use are obtained. In the era of big data with information explosion, the recommendation system is an indispensable tool for software to improve the duration and frequency of users. Obviously, there must be diverse recommendation systems for software with different functions and different object orientations, and the corresponding recall methods are also of diverse methods.

2. Methodology

This paper discusses the commonly used recommendation systems in various recommendation systems and applicable recall methods are classified and explained. The author searched for terms such as "recall", "recommendation system", "news recommendation", "collaborative filtering" and "multi-channel recall" in major thesis databases (Google Scholar, CNKI, Emerald Publishing, etc.) and sorted out a large number of early literature and recent research.

The purpose of this paper is to present a review on the use of the recall metric for evaluating information retrieval systems, especially news recommend system. The value of this paper is to collect, identify and analyse literature that is used in recall. In addition, different views of researchers about recall are identified.

As can be seen in Fig. 1, the relevant documents were grouped according to their years of publication. In alphabetical order, the author's name was included as a citation and the number of documents per year was indicated.

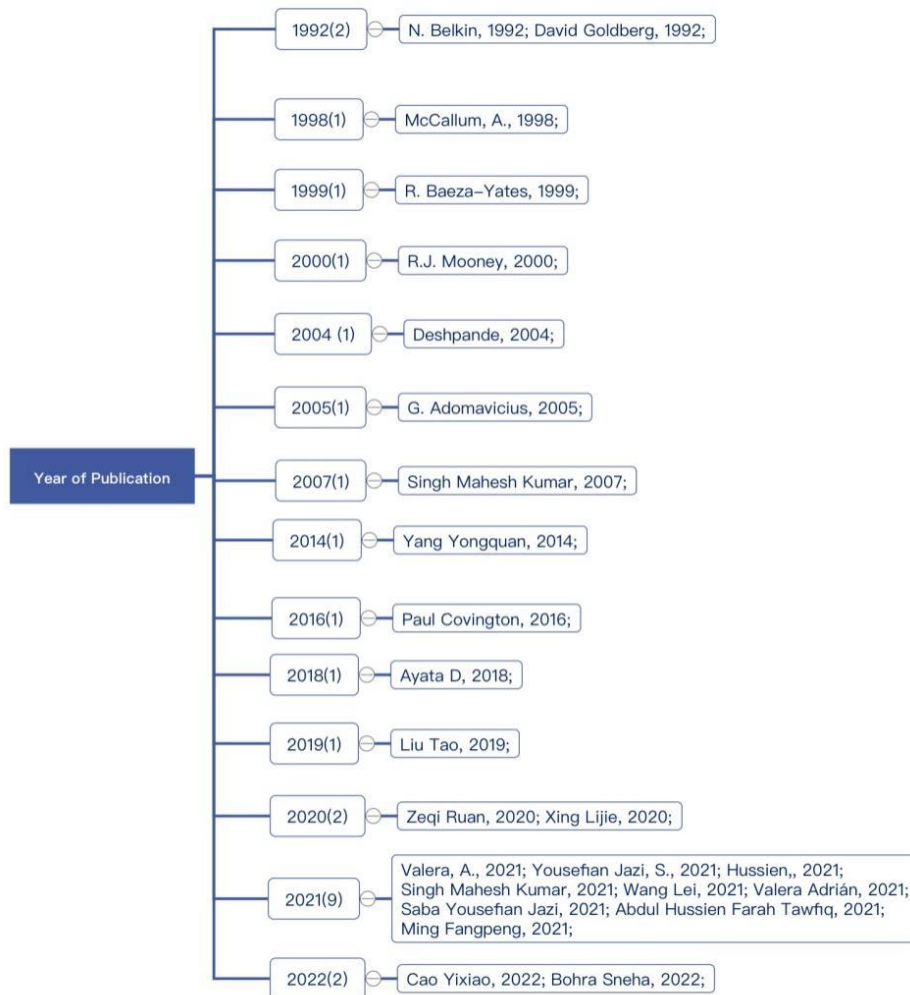


Figure 1. The identified relevant documents and the number of citation per year.

This paper mainly takes the news recommendation system as an example. Some unique recall methods for music and e-commerce recommendation systems are also mentioned. Each recall method is explained from four aspects: definition, formula, relevant elements and applicable scenarios.

3. Findings

This section mainly demonstrates about the four aspects of each recall methods, especially in news recommendation system. The recommended systems included are: news recommendation system, music recommendation system and e-commerce recommendation system.

3.1. Recalls in news recommendation system

Among the recommendation systems, the news recommendation system involves fewer elements, and the requirements are the most common basis. Therefore, there are a lot of research and documents related to recall algorithms. This section will mainly describe several recall methods commonly used in news recommendation systems. It is worth noting that the news recommendation system mentioned here includes recommendations of various news reading software (such as Google News, Toutiao), social software (such as Twitter, YouTube, Weibo) and a variety of real-time article browsing websites system.

After comprehensively sorting out the relevant literature, the recall methods used in the news recommendation system mainly include the following five: Content-based recall, Collaborative filtering, Hot-based recall, Embedding-based recall and Muti-recall (hybrid recommendation algorithm).

3.2. Content-based Recall

This recall method is inspired by the fact that people have subjective evaluations of some items in the past, and similar evaluations will be made to other similar items in the future. Therefore, the main idea of the content-based recommendation (CR) [9, 10] is to use the user's historical behavior data to analyze the user's behavioral habits, and then recommend articles to them according to the history data. The recommended articles are generally based on its title, tags, user-related comments, and additional information annotated manually to refine user behaviors. In software interaction, user behaviors are generally analyzed according to operations such as review, like, watch, browse, and click. It is necessary to extract and model the features of the above information, convert the information into measurable concrete indicators, and then use some methods to calculate the similarity for recall after reasonable quantification.

Content-based recall can be traced back to research in information retrieval [11] and information filtering [12]. R.J. Mooney and L. Roy [10] gave the specific definition of content-based recall and detailed the recall implementation process, which can be roughly summarized as three steps: extracting information and building a database, learning a profile, achieving recall.

Information data such as article titles are extracted from the user's historical data as slots, which can be used for article specific locating. Then the author extracted some non-text wrappers as additional description information, such as authors, synopses, published reviews, customer comments, related authors, related titles, and subject terms. These wrappers can be used as slots for locating articles in subsequent recalls, and together they form the database. Subsequently, during the learning profile process, the user's evaluation of historically browsed articles (review, like, watch, browse, and click, etc.) is weighted in the recall model. R.J. Mooney and L. Roy [9] used Bayes rules to calculate the posterior probability:

$$P(c_j|D) = \frac{P(c_j)}{P(D)} \prod_{i=1}^{|D|} P(a_i|c_j) \quad (1)$$

In this formula a_i represents the i th word in the article, and D means the length of the article in words. Articles will be ranked by their odds ratio:

$$\frac{P(c_1|D)}{P(c_0|D)} \quad (2)$$

When the odds are greater than 1, the example will be classified as c_1 , which represents positive class, and c_0 , negative otherwise. With this model, train on the articles in the training set.

Finally, according to the recall model, the probability of the articles in the material library being clicked by users is calculated, and the top k articles will be recommended. At this point, a round of recall is completed, and the recommendation system will continuously improve and correct the weights of the parameters in step 2 (learning profiles) based on the user's feedback on the recommended articles. Then recall and recommend again and again.

However, although the content-based recall method can meet the needs of recommendation systems to a certain extent, there are still many drawbacks. M.J. Pazzani and D. Billsus [10] argued that the CR system is limited by the information contained in the articles themselves in the material library. When the vocabulary and other information in the article cannot fully reflect the internal information conveyed, the recommendation will be affected. For example, CR can distinguish between digital articles and life articles by some keywords, but it would be difficult to distinguish between funny life articles and other life articles. In addition, G. Adomavicius and A. Tuzhilin [13] argued that under the CR system, users will only be recommended articles in related fields that they have read before. For example, for users who like the Picasso school, even if the articles of the Van Gogh school are excellently written, they will never be recommended. In this way, the system recommendation will be relatively single and lacks diversity.

Nevertheless, it is worth noting that the most important application of content-based recall lies in various collaborative filtering and other systems.

3.3. Collaborative Filtering Recall

The collaborative filtering algorithm was first proposed by Goldberg and other scholars in 1992, and was initially used for filtering in the mail system, named Tapestry [14]; the collaborative filtering algorithm supported by this system has been continuously applied in various recommendation systems. The recall strategy of collaborative filtering mainly considers the similarity between users in the system to generate recommendations.

This type of recall can also be subdivided into two types [15], one is proximity-based collaborative filtering, and the other is model-based collaborative filtering. The former can also be divided into user-based and item-based models, and the common models of the latter include graph models, clustering models, Bayesian models, latent factor models and so on.

The collaborative filtering recall process can be divided into three parts [16]: collecting user preferences, finding similar users or items, and generating recommendation results.

Taking a news recommendation system as an example, first collecting user behavior data such as ratings, comments, likes, and number of views and preprocess them. Through appropriate behavior analysis methods, a user preference matrix (that is, the matrix of user list * article list, corresponding value is the user's comprehensive evaluation of the article) is generated. When searching for similar users, the user similarity $sim(i,j)$ needs to be calculated, and the following three algorithms are generally used. Similarity based on correlation coefficient, using the Pearson correlation coefficient to measure the similarity between user i and user j:

$$sim(i, j) = \frac{\sum_{c \in I_{ij}} (R_{i,c} - \bar{R}_i)(R_{j,c} - \bar{R}_j)}{\sqrt{\sum_{c \in I_{ij}} (R_{i,c} - \bar{R}_i)^2} \sqrt{\sum_{c \in I_{ij}} (R_{j,c} - \bar{R}_j)^2}} \quad (3)$$

Similarity based on cosine:

$$sim(i, j) = \cos(I, J) = \frac{\sum_{c=1}^n R_{i,c} R_{j,c}}{\sqrt{\sum_{c=1}^n R_{i,c}^2} \sqrt{\sum_{c=1}^n R_{j,c}^2}} \quad (4)$$

Similarity calculation based on adjusted cosine:

$$sim(i, j) = \frac{\sum_{c \in I_{ij}} (R_{i,c} - \bar{R}_i)(R_{j,c} - \bar{R}_j)}{\sqrt{\sum_{c \in I_i} (R_{i,c} - \bar{R}_i)^2} \sqrt{\sum_{c \in I_j} (R_{j,c} - \bar{R}_j)^2}} \quad (5)$$

In the above formula, I_{ij} represents the set of articles that user i and user j have jointly rated. $R_{i,c}$ represents the rating of article c by user i and $R_{j,c}$ represents the rating of article c by user j . \bar{R}_i represents the average rating of articles by user i and \bar{R}_j represents the average rating of articles by user j . Finally, the recommendation result is generated. Based on the ratings of similar users, it predicts the ratings of the target users on articles that do not produce operational behaviors, and recommends N books with the highest ratings to the target users. The value of N is a constant and dynamically adjusted according to the rating value.

According to the papers of various researchers, the recall method based on collaborative filtering has the following advantages [15, 16]: better filtering information such as artwork and music that are difficult to carry out automatic content analysis by machine; sharing the experience of others; using the feedback information of similar users to dynamically improve the recommendation function of the recommendation system; providing users with new Recommended; being less intrusive to users when visiting pages, and easy to implement technology.

3.4. Hot-based Recall

In some news recommendation systems that pay attention to timeliness and popularity (such as popular social software like Facebook, Twitter, insgram, Weibo, etc.), recall based on article popularity is also very common. In these social software, there are often hot search rankings or trends, which provide users with relevant popular search recommendations, and hot-based recall comes in handy.

Popularity-based recall is more common to use the number of clicks on articles and the timeliness of clicks as parameters for calculating article popularity. The formula is presented as follows [17]:

$$h(a) = e^{-(t-t_a)} \frac{c_a - c_{min}}{c_{max} - c_{min}} \quad (6)$$

where c_a is the number of clicks on the article, c_{max} and c_{min} are the maximum and minimum clicks in all articles, t_a is the time when the article was clicked, and t is the current time. By calculating the popularity, the articles (or entries, items, etc.) that get TOPK in order will be recommended. It can be seen from the formula that the closer to the current time and the higher the number of clicks, the higher the popularity.

However, it is obvious that the recall method based on popularity is popular and diversified, and is not suitable for personalized recommendation.

3.5. Embedding-based Recall

The above three recall methods are all based on a data log basis, but when the data of users and projects are relatively blank, the cold start problem will be encountered. At this time, the recall based on the article Embedding can better solve this problem. A cold start problem is a situation where there is not enough underlying data for relevant calculations. It can be divided into three types: article cold start, user cold start and system cold start, which refer to three situations: the generation of new articles, the registration of new users and the establishment of new systems. The second one, the cold start problem of new users can be preliminarily solved by recommending popular items to users through hot-based recall. However, for the first situation, the author can only use the content information of the article to analyze the characteristics or attributes of the article content to obtain the vector representation of the article, so as to calculate the similarity of the article for recommendation. Such a recall method is called embedding-based recall.

The recall method based on article Embedding is divided into two parts: Embedding vector construction and similar article recall. First, it is necessary to perform natural language processing (NLP)

on the article, and use a suitable neural network model for Embedding vector calculation. After completing the vector construction, the article similarity calculation is similar to the similarity calculation in the previous collaborative filtering recall.

3.6. *Muti-recall*

In fact, each recall strategy has its own advantages and disadvantages. Multi-way recall is a strategy that runs multiple recall methods in parallel, and comprehensively considers the recall results. It combines the advantages of each recall to a large extent and makes better recommendations. There are also many studies on multi-way recall [18, 19].

3.7. *Recalls in other recommendation system*

In addition to some universal recall methods, diverse recommendation systems will also have specific recall methods fitting for application scenarios. For example, in music recommendation systems, in addition to the traditional content-based and collaborative filtering recall [20, 21] strategies, there are also emotion-based recall [22]. What is more, there is recommendation system based on dynamic analysis of customer behavior in e-commerce platform [23], in addition to the above strategies [7, 24].

3.8. *Music recommendation system*

Those traditional recall methods are also fitting music recommendation system, for which this section will mainly discuss special recommendation strategies for music. It can be known that people's current mood will affect the choice of music [25]. The system recognizes the user as different emotional states (anger, surprise, sadness, boredom and happiness) according to the user's sitting posture, voice's signal, mouse clicks pattern and other signals while listening to music.

Emotional states are represented by two-dimensional coordinates, and emotional states such as anger, surprise, sadness, boredom, and happiness are all points in the two-dimensional coordinates [20]. In the same way, the emotions conveyed by the song can also be represented by points in two-dimensional coordinates. Then a hypothetical line is drawn between the user's current emotion and the target emotion (preferable emotion), and the user's emotional state is transformed by recommending songs on the theoretical line to the user.

According to research by Saba Yousefian Jazi et al., this recall method is more accurate in dealing with happiness emotions, and needs to be improved in dealing with other emotions

3.9. *E-commerce recommendation system*

In [23], Abdul Hussien Farah Tawfiq and other scholars proposed UIBB recall methods that are superior to CFUB (Traditional CF based on users), CFIB (CF based on item) and FPP (CF-based and SPA-based). UIBB is a proposed recall method, represents user item behavior-based recommendation.

The system constructs a global matrix for each user, including all products and the calculation of the characteristic parameters of each product. If there are any changes to these characteristic parameters, the matrix will be updated. Each item of system parameters is a list, including user operation parameters (such as browsing, purchase, collection, price, etc.) used for calculation, and then use these parameters to build a user preference matrix, which is the dynamic update global matrix mentioned above. After that, a recommendation list is generated for each user based on their preference behavior.

According to experiments [23], this improved recall method has better results from the perspectives of precision, recall and F1 function. In addition, user behavior other than purchase information is fully considered. Finally, UIBB shows a good recommendation effect.

4. **Conclusion**

Taking news recommendation system as an example, this paper briefly outlines five recall methods commonly used in recommendation systems: content-based, collaborative filtering, hot-based, embedding-based and multi-recall. In addition, two special recall methods are introduced, which are emo-

tion-based recall in music recommendation system and user item behavior-based recall in e-commerce recommendation system.

In addition to the recall methods mentioned in the article, there are many excellent algorithms suitable for different recommendation scenarios. Scholars in the field of machine learning are also constantly working on this. This paper briefly introduces these recall styles and collects researchers' evaluations and attitudes towards these recall styles, aiming to provide help for recommender system designers in optimizing recall methods and create a more convenient era of big data.

References

- [1] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In Proceedings of RecSys . ACM, 191--198.
- [2] Deshpande, Mukund and George Karypis. Item-based top-N recommendation algorithms. ACM Trans. Inf. Syst. 22 (2004): 143-177.
- [3] Pazzani M J and Billsus D 2007 Content-based recommendation systems (The adaptive web. Springer Berlin Heidelberg) pp. 325-341.
- [4] Valera, A., Lozano Murciego, A., Moreno-García, M.N. Context-Aware Music Recommendation Systems for Groups: A Comparative Study. Information 2021, 12, 506.
- [5] Yousefian Jazi, S., Kaedi, M. & Fatemi, A. An emotion-aware music recommendation system: bridging the user's interaction and music recommendation. Multimed Tools Appl 80, 13559–13574 (2021).
- [6] Hussien, Farah Tawfiq Abdul, Rahma, Abdul Monem S. and Abdulwahab, Hala B., (2021), An E-Commerce Recommendation System Based on Dynamic Analysis of Customer Behavior, Sustainability, 13, issue 19, pp. 1-21.
- [7] Singh Mahesh Kumar, Rishi Om Prakash, Singh Akhilesh Kumar, Singh Pushpendra, Choudhary Pushpa. Implementation of Knowledge based Collaborative Filtering and Machine Learning for E-Commerce Recommendation System[J]. Journal of Physics: Conference Series,2021,2007(1).
- [8] Zeqi Ruan, Chen Yibo, Shen Zhangguo. Video Recommendation System in Internet Era[J]. IOP Conference Series: Earth and Environmental Science,2020,598(1).
- [9] R.J. Mooney, L. Roy, "Content-Based Book Recommending Using Learning for Text Categorization," Proc. Fifth ACM Conf. Digital Libraries (DL '00), pp. 195-204, 2000.
- [10] M.J. Pazzani, D. Billsus, "Content-Based Recommendation Systems," The Adaptive Web: Methods and Strategies of Web Personalization, pp. 325-341, Springer-Verlag, 2007.
- [11] R. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval. Addison-Wesley, 1999.
- [12] N. Belkin, B. Croft, "Information Filtering and Information Retrieval," Comm. ACM, vol. 35, no. 12, pp. 29-37, 1992.
- [13] G. Adomavicius, A. Tuzhilin, "Toward the Next Generation of Recommendation Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749, June 2005.
- [14] David Goldberg, David Nichols, Brian M. Oki, Douglas Terry. Using collaborative filtering to weave an information tapestry[J]. Communications of the ACM,1992,35(12).
- [15] Liu Tao, College Book Recommendation System Based on Collaborative Filtering [J]. Modern Computer (Professional Edition), 2019(02):87-90.
- [16] Yang Yongquan. Research on personalized book recommendation system based on collaborative filtering technology [J]. Henan Library Journal, 2014,34(06):119-122.
- [17] Wang Lei. Research on Recommendation Algorithms and Systems Based on Deep Learning [D]. Beijing University of Posts and Telecommunications, 2021. DOI: 10.26969/d.cnki.gbydu.2021.000946: 26-27.
- [18] Xing Lijie, Lijie Xing, Xiwei Feng, Haiming Chen, Ying Wang, Yue Zhang. Research on fused sorting based on logical regression in news recommendation system[J]. IOP Conference Series: Earth and Environmental Science,2020,510(6).

- [19] Ming Fangpeng, Tan Liang, Cheng Xiaofan. Hybrid Recommendation Scheme Based on Deep Learning[J]. *Mathematical Problems in Engineering*, 2021, 2021.
- [20] Saba Yousefian Jazi, Marjan Kaedi, Afsaneh Fatemi. An emotion-aware music recommendation system: bridging the user's interaction and music recommendation[J]. *Multimedia Tools and Applications*, 2021 (prepublish).
- [21] Cao Yixiao, Liu Peng. Personalized Music Hybrid Recommendation Algorithms Fusing Gene Features[J]. *Mathematical Problems in Engineering*, 2022, 2022.
- [22] Ayata D, Yaslan Y, Kamasak ME (2018) Emotion based music recommendation system using wearable physiological sensors. *IEEE Trans Consumer Electron* 64(2):196–203.
- [23] Abdul Hussien Farah Tawfiq, Rahma Abdul Monem S., Abdulwahab Hala B.. An E-Commerce Recommendation System Based on Dynamic Analysis of Customer Behavior[J]. *Sustainability*, 2021, 13(19).
- [24] Bohra Sneha, Bartere Mahip. Implementing a Hybrid Recommendation System to Personalize Customer Experience in E-Commerce Domain[J]. *Electrochemical Society Transactions*, 2022, 107(1).
- [25] Valera Adrián, Lozano Murciego Álvaro, MorenoGarcía María N.. Context-Aware Music Recommendation Systems for Groups: A Comparative Study[J]. *Information*, 2021, 12(12).