CrossMark

ORIGINAL RESEARCH

Recommendation system techniques and related issues: a survey

Pushpendra Kumar[1] · **Ramjeevan Singh Thakur**[1]

Abstract Nowadays, e-commerce websites are emerging as a new market and allow the millions of product to the user for sale. The selection of product from millions of product requires an additional tool called recommendation system. The recommendation system (RS) helps the user to find the items they are looking for. Collaborative filtering is one of the techniques used in the RSs that is widely studied and used to make recommendation. In this paper, a review of the various methods, algorithms used in the recommender system, the metrics used in RSs and the challenges of recommendation system such as Cold-start, Data sparsity, Scalability, Privacy etc. have been discussed.

Keywords Recommendation system (RS) · Collaborative filtering (CF) · Cold-start · Data sparsity

1 Introduction

With the rapid growth of information on the Internet, it has become a pervasive problem for finding relevant information from the Internet. In this era of Internet, e-commerce has been growing rapidly and allowing millions of product for sale. E-commerce users are suffering from problems of selection of a product from millions of product. Recommendation system helps the e-commerce user to select the items from millions of items [1]. A Recommender system (RS) collects information from a customer about the items

he/she is interested in and recommends that items or products [2]. Nowadays, RS is used on almost every E-commerce websites, assisting millions of users. E-Commerce sites such as Netflix and Movielens [3] for moves, Amazon [4] for books, CD's and many other products, Entree for restaurant and Jester [5] for jokes uses recommender system to assist his consumer. The result of the recommender system is used for both E-Commerce organization as well as users [6] i.e., RS not only assist the customer in getting the preferred item but also increase the revenue of the organization by selling more products. Recommendation System can be classified into three categories based on how the recommendation is performed: Content-based recommendation, Collaborative filtering (CF), and Hybrid approaches. Collaborative Filtering is one of the most widely used and successful technique to recommend an item. It recommends the item to the particular user based on the rating of the other user in the system. Content based approaches perform prediction based on characteristics of the item from their past history, for example recommending a movie that has been categorized as "Action Movies" to a user who likes action movies. Hybrid approach is a combination of content based and collaborative filtering technique in different ways. Figure 1 shows the framework of Memory based collaborative filtering recommendation system. The framework shows that millions of individual users are using E-Commerce sites and their reviews regarding products are store in corresponding server. The E-Commerce organizations collect the data from server and preprocessed into require format (User Rating Matrix). The rating matrix is used to compute the similarity between users by using the different similarity algorithm such as Pearson correlation coefficient, cosine similarity, adjusted cosine similarity and Jaccard's similarity etc. then rating for a particular item that active

✉ Pushpendra Kumar
   pushpendra7589@gmail.com

1 Department of Computer Applications, Maulana Azad National Institute of Technology, Bhopal, Madhya Pradesh, India

Springer

496

Int. j. inf. tecnol. (December 2018) 10(4):495–501

Fig. 1 Framework of memory based collaborative filtering recommendation system
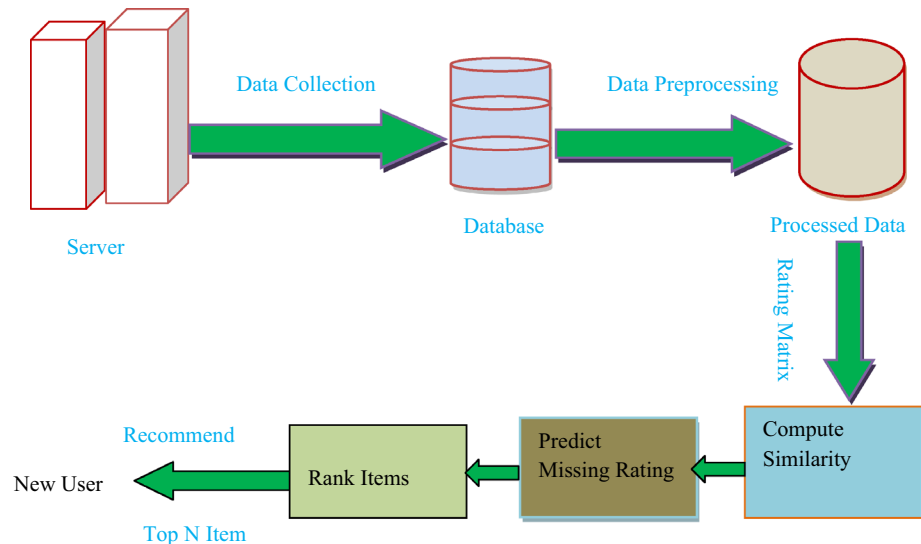
user has not yet been given is predicted and top N predicted items are recommended to the active user. The rest of paper is organized as follows. In Sect. 2 Related work is discussed. In Sect. 3 Recommendation system techniques are discussed. In Sect. 4 Issues related to Recommendation system are discussed. In Sect. 5 Evaluation metrics of Recommendation system are discussed. Finally, In Sect. 6 conclusion is given.

2 Related work

Rodrigues et al. [6] proposed a framework which combines the item-based collaborative filtering (CF) with user demographic information in cluster weighted mechanism to solve the cold start and data sparsity issues. This system provides the good recommendation to the new users which makes the user experience great and also increases the organization revenue. Better recommendations can be provided by making the cluster based on cross domain data. For example, if a user who likes romantic songs, the system can recommend him love story movies.

Ji et al. [7] introduced a scalable CF algorithm based on matrix factorization, performed prediction using two decision matrices: user-category and user-keyword instead of using single user-item rating matrix. The proposed algorithm is implemented on real data set and the result shows that model has good scalability for new items.

Gu et al. [8] a simple collaborative filtering suffers from data sparsity problem because of the explosive growth of users and items in e-commerce. This paper introduced a dynamic-weighted CF technique (DWCF) to resolve data sparsity and adaptive issues. In this approach similarity between user and items is found then a weight controlling method is proposed to find the impacts of user & item

similarity. So this method outperforms under various situations of data sparsity.

Koohi et al. [9] Collaborative filtering suffers from data sparsity and high dimensionality problem. In this paper, author solves these issues by finding the neighbor user using subspace clustering approach. The author constructs the different subset of a rated matrix as Interested (I), Neither Interested Nor Uninterested (NIU) and Uninterested (U). Based on these subsets three level of the tree is created for the neighbors of an active user. This method is efficient in dealing with sparse data.

Verma et al. [10] proposed a recommendation system using collaborative filtering (CF) technique and fuzzy c-means (FCM) clustering algorithm. FCM clustering is used for item clustering and CF is used for rating prediction. FCM performs better than K means clustering because K means has a restriction that one item belongs to single cluster where as one item may similar to more than one group of items.

Kumar et al. [11] proposed a hybrid collaborative filtering method to resolve the issues of sparsity and scalabilities provide more personalized recommendations. The proposed method works in two phases, in the first phase resolve the sparsity using Case based reasoning (CBR) followed average filling and in the second phase resolve the scalability using clustering into the group by Self-organizing map optimized with a Genetic algorithm.

Koohi et al. [12] Proposed a Collaborative Filtering recommendation system using fuzzy C-means clustering algorithm, performance against the K-means and SOM clustering approaches have been evaluated. The experimental result shows that fuzzy c-means clustering outperforms another clustering in terms of accuracy, precision and recall.

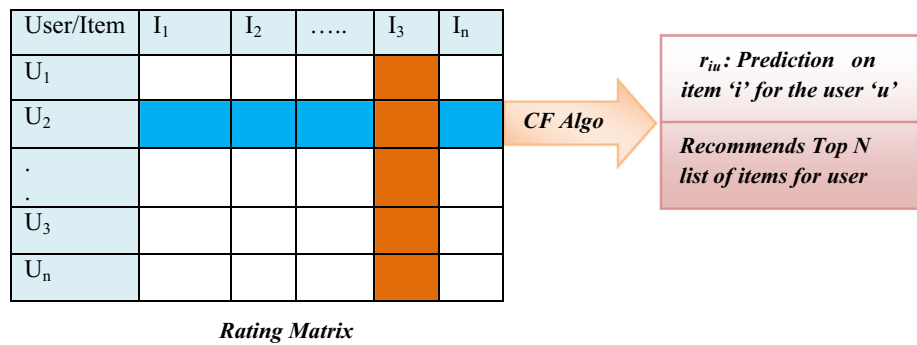**Table 1** Summarized information of literature review

References	Author	Issues name	Types of filtering	Technique used	Dataset used	Advantages
[6]	C. M. Rodrigues, et al. 2016	Cold start and real-time prediction problem	Hybrid collaborative filtering	Item-based CF algorithm combined with user demographic based CF algorithm in clusters weighted mechanism	Movie Lens	Save resources and time by predicting the relevant data only
[7]	Ke Ji, et al. 2104	Scalability problem	Collaborative filtering	Matrix factorization based approach	Real Dataset published by KDD CUP 2012	Good scalability for new items
[8]	Liang Gu, et al. 2014	Data sparsity and adaptivity issues	Collaborative filtering	Dynamic-weighted collaborative filtering approach (DWCF)	Movie Lens	Under various situations of data sparsity, DWCF performs good recommendations than traditional hybrid approach
[9]	Hamidreza Koohi, et al. 2017	Data sparsity and high dimensionality	Collaborative filtering	Finding neighbor user using subspace clustering method & new similarity measure is used to compute similarity value	Movie lens and Jester	Efficient in dealing with sparse data and performs better than the conventional method in terms of accuracy, precision and recall.
[10]	S. K. Verma, et al. 2013	Sparsity and cold start problem	Collaborative filtering	Fuzzy C means clustering	Movie Lens	Cold- start problem is resolve very efficiently
[11]	N. P. Kumar, et al. 2015	Data sparsity and scalability	Hybrid user-item based collaborative filtering	For data sparsity case based reasoning and for scalability self organizing map optimized with GA is used	Movie lens	Performs enhanced personalized recommendations with better rating quality
[12]	Hamidreza Koohi, et al. 2016	Recommendation accuracy	Collaborative filtering	Fuzzy C means clustering	Movie lens	Fuzzy C- means clustering algorithm produced better results compare to SOM and K-means clustering
[13]	O-Joun Lee, et al. 2015	Data sparsity and scalability	Collaborative filtering	Markov model and fuzzy clustering	Movie lens	Improves performance instability
[14]	Kyoung-jae Kim, et al. 2008	Data sparsity	Case based reasoning (CBR)	GA K- means clustering	Real-world data set from the internet shopping mall	By providing good optimal solutions in less time, efficiently segment the online shopping customer
[15]	Yilmaz Ar, et al. 2016	Prediction error	Collaborative filtering	Genetic approach	Movielens	Reducing the prediction error and helps in recommending better product

Lee et al. [13] introduce a Predictive Clustering-based Collaborative Filtering (PCCF) that combines the Markov model and fuzzy clustering with Clustering based CF (CBCF). This method solves the issue of reduced coverage and of unstable performance by tracking the changes in user preferences and bridging the gap between the static model and dynamic model.

Kim et al. [14] proposed a recommender system for online shopping market using GA K-means clustering. In this system, the author tries to segment the online shopping

498

Int. j. inf. tecnol. (December 2018) 10(4):495–501

Fig. 2 Collaborative filtering process

User/Item	I₁	I₂	…..	I₃	Iₙ
U₁					
U₂					
. .					
U₃					
Uₙ					

CF Algo

r_{iu}: Prediction on item 'i' for the user 'u'

Recommends Top N list of items for user

Rating Matrix

user according to their buying behavior. GA is used to resolve the local optima problem found in K-means clustering & provide a method of finding the relevant groups more efficiently.

Ar et al. [15] proposed an approach that reduces the prediction error that occurs in collaborative filtering RS. The conventional CF method uses similarity values directly for the rating prediction of an item whereas in proposed approach author uses a genetic algorithm before using the prediction process to get the better result. The statical analysis performed on various similarity matrices such as Vector Cosine Similarity, Pearson's Correlation and Extended Jaccard Coefficient and result shows that evolutionary approach has reduced the prediction error. Table 1 shows the summary of the works that have done by different authors in the field of the recommendation system.

3 Recommendation system techniques

3.1 Collaborative filtering

Collaborative filtering (CF) technique recommends an item to the particular user based on the rating/opinions of the other user [16, 18]. CF system performs recommendation by building a database of preferences for items by the user. The system then finds the user with similar interest and preferences by calculating similarities between the user profiles [17], build a group of similar user called neighborhood. A user gets the recommendation to those products that he has not rated/purchased but his neighbors are rated. Collaborative filtering performs predictions or recommendations, the prediction is a numerical value and recommendation is a list of top N items that the user will like the most [17] as shown in the Fig. 2. Collaborative filtering technique can be classified into two broad categories (a) Memory-based technique (b) Model-based technique [16–18]. Memory-based technique identify the similarity between an active user to all other user using similarity measures such as Pearson correlation, Cosine similarity, Jaccard coefficient etc. Then missing rating of an active user is predicted and the top k

rated item is recommended to the active user. In the model-based technique, previous ratings are used to develop a model using machine learning technique. Once the model is developed predictions can be made for an individual user.

3.2 Content-based filtering

Content-based (CB) approach performs recommendation of those items that are similar in characteristic to the item that the users have already used in their past. CB approach performs more analysis on the attribute of the item in order to produce recommendations. CB filtering (CBF) technique is most successful in webpages, publications and news recommendation.

CBF system automatically creates personalized profiles of the user based on his feedback and type of item likes. In order to generate meaningful recommendations, collected user information is compared against the characteristic of the item examine [19] as shown in the Fig. 3.

3.3 Hybrid filtering

Hybrid filtering system achieves by a combination of two or more recommendation system in order to get better performance over collaborative filtering and content-based filtering. It is possible to combine CF and CBF technique in a different way to obtain hybrid filtering system, which may produce several outputs. Hybridization process categorized into seven different types [17] such as (1) Weighted (2) switching (3) Mixed approach (4) Feature combination (5) feature augmentation (6) Cascade and (7) Meta-level.

4 Issue related to recommendation system

4.1 Limited content analysis

Content-based filtering (CBF) techniques are restricted by the characteristic that is explicitly concerned with the item that is recommended. So in order to obtain enough number of characteristic, the content must be in the form that can
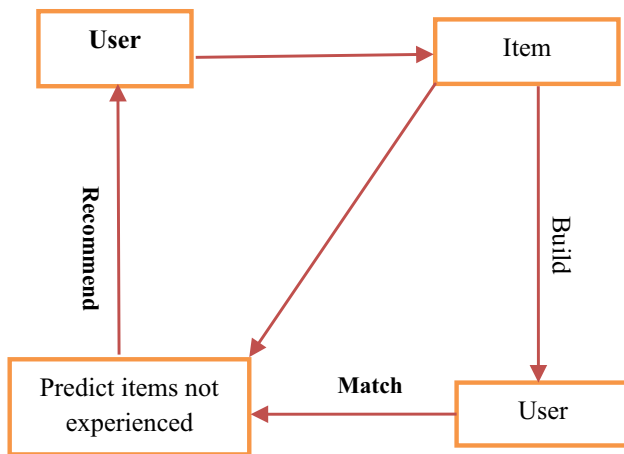
Fig. 3 Content-based filtering process

be parsed automatically or the characteristics should be assigned manually [24]. CBF also facing another problem that is when two different items having the same characteristics are not distinguishable to the system.

4.2 Cold-start problem

It refers to the situation where it is difficult to make recommendations for a new user and items. Because of lack of sufficient rating information, it is difficult to find similarity between users and items. So, neither the taste of the new users can be predicted nor the new items be rated or purchased by the users, this situation leads to less accurate recommendations. The cold start problem can be solved in many ways such as (a) Ask the new user at the beginning to rate some items. (b) Ask to state the taste of new users explicitly. (c). Recommends items to the new user based on the collected demographic information.

4.3 Data sparsity problem

This is the problem that occurs when a majority of the users do not rate most of the items and consequently, the user-item matrix becomes very sparse. So, the chance of getting a set of users with the similar rating decreases. Collaborative filtering uses the nearest neighbor approach to recommend items and less rating makes difficult to make accurate predictions about items.

4.4 Scalability problem

Recommender system is facing one of the vital and foremost issues with the large real-world dataset are called scalability. If the size of dataset grows with the number of user and items the computation also grows linearly. i.e., when the dataset is small algorithm works well but unable

to generate the satisfactory result for a large volume of the dataset. Thus, it is very difficult to apply recommendation technique with huge and dynamic data sets produced by item-users interaction. Scalability problem can be solved using Dimensionality reduction, Bayesian Network and Clustering etc.

4.5 Privacy issue

Recommendation algorithm requires input from the user population to produce quality personalized recommendations; this may lead to issues of data privacy and security. Thus a technique required to be designed that can reasonably and carefully use the user data by assuring that information about the user-item rating can't be freely available to the malicious users.

4.6 Synonymy

It is the situation which refers similar items having different names or entries. RS algorithms are unable to find the difference between closely related items such as "comedy movie" and "comedy film". The extreme usage of synonym words decreases the performance of collaborative filtering recommendation. Synonymy problem can be solved by using these methods (a) Construction of a thesaurus (b) Singular Value Decomposition (SVD) (c) Latent Semantic Indexing.

5 Evaluation metrics of RSs

The quality of a recommendation system algorithm can be assessed using the different method. The type of metrics used depends on the types of filtering technique. The assessment of prediction and recommendation has been considered essential so that the user can have the best experience with RSs. Evaluation metrics can be classified as follows:

5.1 Statistical accuracy metric

It evaluates the accuracy by comparing the predicted rating with the actual rating. The commonly used metrics are Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Correlation.

5.1.1 Mean absolute error

It is an average of the absolute deviation between predicted rating and actual rating. The lower MAE value shows the better prediction [20]. Let r_1, r_2, r3,...., r_n are the actual

500

Int. j. inf. tecnol. (December 2018) 10(4):495–501

ratings and the corresponding $p_1, p_2, p_3, ..., p_n$ are the predicted ratings.

It is defined as follows

$$MAE = \frac{\sum_{i=1}^{n} |r_i - p_i|}{n}. \tag{1}$$

5.1.2 Root mean square error

It is also used for the measure of model performance. RMSE is obtained by squaring the difference between predicted rating and actual rating, adding those together, dividing that by the no of test points and then taking the square root of the result [21]

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (r_i - p_i)^2}{n}}. \tag{2}$$

5.1.3 Correlation

Correlation analysis refers to the measure of the linear relationship between two variables. A higher correlation value shows more accurate rating prediction or recommendations

$$corr(p,r) = \frac{\sum_{i=1}^{n} (p_i - \bar{p})(r_i - \bar{r})}{\left[\sum_{i=1}^{n} (p_i - \bar{p})^2 \sum_{i=1}^{n} (r_i - \bar{r})^2 \right]^{1/2}}. \tag{3}$$

5.2 Classification accuracy metrics

The recommendation system makes decisions about whether an item is good or not by measuring the frequency [22]. Generally in a Recommendation system, binary rating is used that is the items are relevant to the user interest or not because rating dataset is extremely sparse compared to binary selection dataset. For this metrics, a rating dataset is transformed into the binary dataset. Three classification accuracy metrics Precision, Recall, and F-1 score are often used to assess the relevance between recommendations and user interest [23]

$$Precision = \frac{\text{Relevant Item Recommended}}{\text{Total Item Recommended}}, \tag{4}$$

$$Recall = \frac{\text{Relevant Item Recommended}}{\text{Total Relevant Items}}, \tag{5}$$

$$F1 = \frac{2 * Precision * Recall}{Pecision + Recall}. \tag{6}$$

6 Conclusion

Recommendation system has an ability to provide personalized information on the internet. In this era of internet, lots of RSs have been developed that are based on Content-based filtering, Collaborative filtering and Hybrid system and helps to reduce the problem of information overload. In this study authors found that CF recommendation system provide better recommendation but still facing problem of scalability and sparsity. So there is a possibility to improve the quality and performance of collaborative filtering based recommendation system by using the fuzzy clustering and the optimization technique.

References

1. Deshpande M, Karypis G (2004) Item-based top-N recommendation algorithms. ACM Trans Inf Syst 22(1):143–177
2. Sarwar B, Karypis G, Konstan J, Riedl J (2000) Analysis of recommendation algorithms for e-commerce. In: Proceedings of the 2nd ACM conference on Electronic and commerce, pp 158–167
3. Miller BN, Albert I, Lam SK, Konstan JA, Riedl J (2003) Movie Lens unplugged: experiences with an occasionally connected recommender system. In: Proceedings of the Int'l Conf. Intelligent user interfaces, Miami, Florida, USA, pp 223–266
4. Linden G, Smith B, York J (2003) Amazon.com recommendations: item-to-item collaborative filtering. IEEE Int Comput 7:76–80
5. Goldberg K, Roeder T, Gupta D, Perkins C (2001) Eigentaste: a constant time collaborative filtering algorithm. Inf Retr J 4:133–151
6. Rodrigues CM, Rathi S, Patil G (2016) An efficient system using item and user-based CF techniques to improve recommendation. In: Proc. of International Conf. on next generation computing technologies (NGCT), pp 569–574
7. Ji K, Shen H (2014) Using category and keyword for personalized recommendation: a scalable collaborative filtering algorithm. In: Proc. of sixth international symposium on parallel architectures, algorithms and programming, Beijing, pp 197–202
8. Gu L, Yang P, Dong Y (2014) An dynamic-weighted collaborative filtering approach to address sparsity and adaptivity issues. In: 2014 IEEE Congress on Evolutionary Computation (CEC), Beijing, pp 3044–3050
9. Koohi H, Kiani K (2017) A new method to find neighbor users that improves the performance of collaborative filtering. Expert Syst Appl 83:30–39
10. Verma SK, Mittal N, Agarwal B (2013) Hybrid recommender system based on fuzzy clustering and collaborative filtering. In: Proceedings of the 4th International Conference on Computer and Communication Technology (ICCCT), Allahabad, pp 116–120
11. Nitin PK, Fan Z (2015) Hybrid user-item based collaborative filtering. Procedia Comput Sci 60:1453–1461
12. Koohi H, Kiani K (2016) User based Collaborative Filtering using fuzzy C-means. Measurement 91:134–139
13. Lee OJ, Jung JJ, Eunsoon Y (2015) Predictive clustering for performance stability in collaborative filtering techniques. In: Proceedings of the IEEE 2nd International Conference on Cybernetics (CYBCONF), Gdynia, 2015, pp 48–55

14. Kim K-J, Ahn H (2008) A recommender system using GA - means clustering in an online shopping market. Expert Syst Appl 34(2):1200–1209

15. Ar Y, Bostanci E (2016) A genetic algorithm solution to the collaborative filtering problem. Expert Syst Appl 61:122–128

16. Kumar B, Sharma N (2016) Approaches, issues and challenges in recommender systems: a systematic review. Ind J Sci Technol 9(47). https://doi.org/10.17485/ijst/2016/v9i47/94892

17. Isinkaye FO, Folajimi YO, Ojokoh BA (2015) Recommendation systems: principles, methods and evaluation. Egypt Inform J 16:261–273

18. Khusro S, Ali Z, Ullah I (2016) Recommender systems: issues, challenges, and research opportunities. In: Kim K, Joukov N (eds) Information science and applications (ICISA) 2016. Lecture Notes in Electrical Engineering, vol 376. Springer, Singapore, pp 1179–1189

19. Sharma L, Gera A (2013) A survey of recommendation system: research challenges. Int J Eng Trends Technol 4(5):1989–1992

20. Gong S (2010) A collaborative filtering recommendation algorithm based on user clustering and item clustering. J Softw 5(7):745–752

21. Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature. Geosci Model Dev Discuss 7:1247–1250

22. Herlocker JL, Konstan JA, Terveen LG, Riedl JT (2004) Evaluating collaborative filtering recommender systems. ACM Trans Inf Syst (TOIS) 22(1):5–53

23. 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

24. Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans Knowl Data Eng 17(6):734–749