

# Exploration of Collaboration Filtering Techniques for Product Recommendation

Ajith Kumar, P. Madhavan

*Abstract :Today, recommendation system has been globally adopted as the most effective and reliable search engine for knowledge extraction in the field of education, economics and scientific research. Collaborative filtering is a proven techniques used in recommender system to make predictions or recommendations of the unknown preferences for users based on the known user preferences. In this paper, collaborative filtering task and their challenges are explored, study the different recommendation techniques and evaluate their performance using different metrics.*

**Keywords:** Collaborative Filtering, Knowledge Extraction, Recommender System

## I. INTRODUCTION

Recommendation has become the most important tool used to change the method of communication between web users and web sites. Recommender system has been widely used in the field of economics, education and scientific research [1].

Today, the volume of data has tremendously increased that makes it extremely difficult for a user to find information they needed. With the increase in the volume of data, problem of data deficiency have been eliminated but leading to the issue of overload [2]. The excessive data leading to overload in the area of research makes it very difficult for researchers to properly judge the relevance of the items retrieved as the search result for making the right decision [3]. To alleviate the issue of overload an effective filtering mechanism is required to determine useful information from a large volume of data. Collaborative filtering as a classic filtering algorithm for the development of recommender system is divided into item based collaborative filtering and user based collaborative filtering. The user based collaborative filtering makes recommendations by searching similar users to the target user and uses the favorite items of the similar users to make recommendations to the target users whereas the item based collaborative filtering tries to locate items similar to the items purchased by the target user based on the user purchase or browse history [4].

**Revised Manuscript Received on February 05, 2020.**

**Ajith Kumar**, Master of Technology, Department of Computer Science and Engineering, SRM Institute of Science and Technology, KTR Campus, Chennai, Tamilnadu, India.

**Dr. P. Madhavan**, Associate Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, KTR Campus, Chennai, Tamilnadu, India.

## A. Characteristics and Challenges of Collaborative Filtering

The ability of the recommender system to provide accurate recommendations to customers is a great challenge that relies on how well the challenges are addressed. The speed and accuracy of the recommender system in E-commerce attracts the attention of the customers and also boost the benefit of the company. The major characteristic of collaborative filtering task includes:

- **Data Sparsity:** in e-commerce recommender systems evaluates a large number of product/item set that are distributed across different locations. The performance of predictions or recommendations made by the collaborative filtering method is a great challenge due to the sparsity of the user item matrix used by the CF method. A cold start problem also occurs when a new user or item is added to the system. In this situation it is difficult to find a similar user as there is no enough information about the user. Also, the newly added item cannot be recommended to a user as it has no got any rating from the previous users or purchase history. A reduced coverage also occurs due to a small number of ratings when compared to the large number of items/products in the system. In this case the recommender system may not be able to generate a prediction for the items. Thirdly, neighbor transitivity may reduce the effectiveness of the recommendation if the databases are sparse. In this case users with similar taste may not be identified [2].
- **Scalability:** Count of users and items of commercial system is significantly increasing. When the number of users and items increases, the traditional collaborative filtering algorithms undergoes a serious scalability problem as the computational resources exceeds practical or acceptable level [4].
- **Synonym:** a number of similar items may have different names or entries within the system. In this case the recommender system may consider these items with similar name or entries as differently without discovering the latent association between the products. The relevance of synonyms reduces the performance of the predictions made by the recommender system.
- **Grey Sheep:** the opinion of the user may not consistently agree or disagree with the opinion of any user group. In this case the user may not benefit from the recommender system.

In view of the challenges faced by the recommender system, a number of techniques were proposed by researchers and more of these researches are in place to address the issues.

### II. RELATED WORKS

In [5], authors proposed a hybrid recommendation system to surpass the difficulties in providing the right product based on customer preferences. The method combines context-based, data mining techniques and collaborative filtering. The algorithm begins by obtaining the similar group of customers based on customer lifetime value.

In [6], authors developed a recommendation system using two recommender systems: collaborative and demographic analysis that provide an effective product recommendation.

In [7], authors provide the overview of different collaborative filtering techniques giving some example, illustrative charts, and new trends with respect to time. Authors diagrammatically represent the techniques and explore future research directions. In this paper authors conclude that extra effort needs be put in place to overcome the limitations of the current techniques used.

In [8], authors proposed a structural balance theory based recommendation on E-Commerce large rating data. In this method the enemy (dissimilar user) of the target user is first obtained and then the possible friends of the E-Commerce target user based on the found enemy's enemy (Friends).

In [9], authors use the common singularity ratings provided by users to determine the similarity between the users. Author's uses lower probability to choose a neighbor that best fit the preferences of a group.

In [3], authors classify the most commonly used method for scientific paper recommendation, deeply analyze the evaluation metrics used in paper recommendation system and provide the summary of the challenges and problems faced in paper recommender system.

In [10], authors proposed an aspect sentiment collaborative filtering technique that obtained the users attributes towards aspect of the item through fined grained sentiment analysis on user's transaction history. The degree of the user desire and importance on the item is analyzed for each feature using fuzzy kano model.

In [13], authors proposed a recommendation algorithm that merges trusted relevance and matrix factorization. The proposed method establishes an effective trust metric model that integrates the user information into the proposed model. The direct or indirect trust relationships are considered using a concentric hierarchical model and integrate more information into the matrix factorization recommendation algorithm. Authors consider trust factors and interest similar facts to design the trust relevance.

In [11], authors proposed a high recommendation algorithm that exploit the user's attributes and cluster the attributes into a number of clusters. In each of the generated cluster a virtual opinion leader is conceived that is used to represents the entire cluster which significantly reduced the dimension

of the initial user item matrix. A method called One-VU is designed and applied to virtual opinion leader item matrix to obtain the results of the recommendation.

### III. COLLABORATIVE FILTERING METHODS

In this section, we discuss in detail the collaborative filtering methods used for generating recommendations for items to users of the e commerce.

#### Memory-Based Collaborative Filtering Technique

The memory-based collaborative filtering technique used the available data stored in the memory to predict the new rating using similarity between users or items. The set of similar users forms neighbor that is used to extract objects or products of similar ratings. Assumption here is that two users with similar rating history for some similar items will likely have the same preferences for other items. This algorithm is divided into user based and item-based algorithms based on the using the target user or target item/product. The user based and item-based methods were proposed to address the issue of scalability in the system [12].

#### Model Based Collaborative Filtering Technique

The model based collaborative filtering technique adopts various techniques on the training sets, in order to obtain pattern on a data and learn a model to predict new rating for user items. This model uses techniques like Fuzzy systems, slope, Bayesian Classifier, various probabilistic relational models, linear regression, Markov chain model, Latent Dirichet Allocation, Singular Value Decomposition, Clustering Models, Neural Networks, maximum entropy model and so on [12].

#### Hybrid Recommendation Technique

Various combinations of content based and collaborative filtering techniques exist that can be used to exploit user r item information, ratings and similarities of various users and items. The four ways of combining content based and collaborative filtering techniques includes: separate implementation of content based and collaborative filtering and them merging the result, Boosting the collaborative filtering algorithm using some features of the content based method, boosting the content based method using some characteristics of the collaborative filtering method and unify the collaborative filtering and the content based method into a single recommender system [12].

**Table 1: Overview of CF Methods**

CF Method	Techniques Used	Advantages	Drawback
Memory Based CF	Neighbor Based CF	<ul style="list-style-type: none"> <li>- It's very easy to implement</li> <li>- New data can be added incrementally</li> </ul>	<ul style="list-style-type: none"> <li>- Its performance decreases on sparse databases</li> <li>- it fails to recommend for new user or item</li> <li>- it depends on human ratings</li> </ul>
Model Based Cf	Fuzzy systems, slope, Bayesian Classifier, various probabilistic relational models, linear regression, Markov chain model, Latent Dirichet Allocation, Singular Value Decomposition, Clustering Models, Neural Networks, maximum entropy model	<ul style="list-style-type: none"> <li>- Improve the performance of predictions</li> <li>- It better address the problem of scalability, sparsity and other related problems</li> </ul>	<ul style="list-style-type: none"> <li>- Useful information are lost in dimensionality reduction</li> <li>- Prediction performance and scalability tradeoff</li> </ul>
Hybrid Based	Content Boosted CF, Merged Content Based and CF Algorithm,	<ul style="list-style-type: none"> <li>- Improved the performance of prediction</li> <li>- Overcomes the problems of other techniques like sparsity, gray sheep and so on</li> </ul>	<ul style="list-style-type: none"> <li>- Increased in complexity and cost of implementation</li> <li>- require external information at are usually not available</li> </ul>

#### IV. EVALUATION METRICS

We categorize evaluation metrics under two major categories memory-based approach and model-based approach. Under memory-based approach we have user-based, Item-based, RA, UOS, MLCF, ULPE, IGPE. Under model-based approach we have slope 1, w-slope 1, RSVD, NMF, PMF, BPMF, NLPMF. Accuracy based metrics (NMAE) Normalized Mean Absolute Error, (NRMSE) Normalized Root Mean Square Error, (P@N) Precision, (R@N) Recall, ( $\tau$ ) Kendall tau correlation, (F1) score has inverse correlation to precision and recall. We denote testing and training time in seconds. (RA) Resource Allocation in CF is used for link prediction. (UOS) User Opinion Spreading works in spreading user's preferences for more accurate ratings. (MLCF) Multilevel Collaborative Filtering is used to enhance similarity calculation of unknown items. (ULPE) User Local Profile Expansion is method of user's ratings based on neighbors' ratings on items by active users. (IGPE) Item Global Profile Expansion is recommendation method based on similar item's ratings already rated by user. (SLOPE 1) is a method of computing average difference between two items rated by same user. (Weighted Slope One) is an enhanced Slope one which also uses number of observed ratings. (RSVD) Randomized Singular Value Decomposition is used to find low rank value with

fast probabilistic method of calculation. (NMF) Non-Negative Matrix Factorization used for creation of user and item-based matrix with non-negative values to calculate ratings. (PMF) Probabilistic Matrix Factorization is proven method for large dataset works based on linear probabilistic model. (BPMF) Bayesian Probabilistic Matrix Factorization is updated PMF with automated control over parameters of all model. (NLPMF) Non-Linear Probabilistic Matrix Factorization uses Gaussian process variable models for recommendation.

V. EXPERIMENT AND RESULT

In this experiment, a machine with Intel(R) Core i5, CPU of 2.0 GHz, RAM of 4GB running JAVA

language on windows 10 operating system was used to compare the performance of some RS in terms of various evaluation metrics.

Table 2: Comparison of Various CF Algorithms

Metrics	NMAE	NRMSE	$\tau$	P@10	R@10	F1	Training	Testing
User-Based	0.183	0.246	0.413	0.013	0.009	0.010	8.056	169.37
Item-Based	0.196	0.285	0.392	0.087	0.107	0.096	18.503	4012.3
RA	0.187	0.248	0.125	0.026	0.015	0.019	9.245	274.76
UOS	0.238	0.324	0.254	0.149	0.172	0.160	9.923	129.43
MLCF	0.160	0.212	0.457	0.011	0.007	0.008	8.752	157.04
ULPE	0.154	0.202	0.732	0.005	0.002	0.003	13.071	184.11
IGPE	0.169	0.246	0.851	0.012	0.081	0.021	139.21	217.11
SLOPE 1	0.225	0.318	0.312	0.047	0.054	0.050	24.201	138.22
W-SLPOE 1	0.213	0.297	0.306	0.073	0.071	0.072	24.705	136.57
RSVD	0.143	0.195	0.285	0.105	0.101	0.103	71.103	109.32
NMF	0.201	0.252	0.349	0.142	0.087	0.108	19.140	106.21
PMF	0.213	0.274	0.316	0.113	0.093	0.102	893.07	123.14
BPMF	0.172	0.246	0.274	0.097	0.009	0.016	21.514	111.52
NLPMF	0.235	0.310	0.339	0.108	0.103	0.105	181.12	397.61

VI. CONCLUSION

In this paper, the well-known memory-based and content based collaborative filtering techniques were implemented and evaluated using different metrics, discuss the advantages and drawbacks of the well-known methods. Experiment shows that the memory-based CF have a faster query time compared to other techniques and the model based have more scalability and perform better on sparse databases. However, the later model is expensive to build and result to increase in time for the training. Therefore, a tradeoff between performance and scalability needs to be addressed.

REFERENCES

1. J. Zhang, D. Chen, and M. Lu, "Combining Sentiment Analysis With a Fuzzy Kano Model for Product Aspect Preference Recommendation," IEEE Access, vol. 6, pp. 59163–59172, 2018.
2. X. Su and T. M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques," Advances in Artificial Intelligence, vol.
3. "Scientific Paper Recommendation: A Survey," IEEE ACCESS, pp. 9324–9339, Jan. 2019.
4. J. Han, M. Kamber, and J. Pei, Data mining: concepts and techniques. Amsterdam: Elsevier/Morgan Kaufmann, 2012.
5. F. Rodrigues and B. Ferreira, "Product Recommendation based on Shared Customers Behaviour," Procedia Computer Science, vol. 100, pp. 136–146, 2016.
6. "An Effective Product Recommendation System for E-Commerce Website Using Hybrid Recommendation Systems," IJCSSE, pp. 81–88.
7. S. S. Sohail, "Product Recommendation Techniques for Ecommerce," IJAR CET, vol. 1, no. 9, pp. 219–225.
8. L. Qi, X. Xu, X. Zhang, W. Dou, C. Hu, Y. Zhou, and J. Yu, "Structural Balance Theory-Based E-Commerce Recommendation over Big Rating Data," IEEE Transactions on Big Data, vol. 4, no. 3, pp. 301–312, Jan. 2018.
9. F. Ortega, R. Hurtado, J. Bobadilla, and R. Bojorque, "Recommendation to Groups of Users Using the Singularities Concept," IEEE Access, vol. 6, pp. 39745–39761, 2018.
10. J. Zhang, D. Chen, and M. Lu, "Combining Sentiment Analysis With a Fuzzy Kano Model for Product Aspect Preference Recommendation," IEEE Access, vol. 6, pp. 59163–59172, 2018.
11. J. Zhang, Y. Wang, Z. Yuan, and Q. Jin, "Personalized real-time movie recommendation system: Practical prototype and evaluation," Tsinghua Science and Technology, vol. 25, no. 2, pp. 180–191, 2020.
12. M. Jalili, S. Ahmadian, M. Izadi, P. Moradi, and M. Salehi, "Evaluating Collaborative Filtering Recommender Algorithms: A Survey," IEEE Access, vol. 6, pp. 74003–74024, 2018.

13. W. Li, X. Zhou, S. Shimizu, M. Xin, J. Jiang, H. Gao, and Q. Jin, "Personalization Recommendation Algorithm Based on Trust Correlation Degree and Matrix Factorization," IEEE Access, vol. 7, pp. 45451–45459, 2019.

AUTHORS PROFILE



Ajith Kumar received the BTech degree in computer science and engineering from SRM Institute of science and technology, Chennai, Tamilnadu, India in 2018. He is currently pursuing the MTech degree in computer science and engineering. His research interests include recommender systems and machine learning.



Dr.P.Madhavan obtained his Bachelor’s Degree-BE in Electronics and Communication Engineering and Master’s Degree-ME in Multimedia Technology. He has completed his Ph.D degree from Anna University, Chennai in the field of Wireless Adhoc Network. He started his carrier as Lecturer in the year 2002. At present, He is working as Associate Professor at SRM University, Chennai. He has more than 13 years of

teaching experience in Engineering College. He has published more than 20 papers in International and National Journals, 30 International and National Conferences. He is the Life member of Indian Society for Technical Education (ISTE), International Association of Engineers (IAENG). His areas of Interest include Wireless adhoc and sensor networks, Multimedia Communication, Fuzzy and Neural Computing.

