An Analysis of Data Sparsity Resolution Algorithms Used in Recommender Systems

Shivani Bhardwaj , **Kushal Kanwar** , **and Gaurav Gupta**

Abstract In previous years, a substantial rise has been observed in the use of YouTube, Netflix, Amazon, and other similar web services that have enabled the utilization of recommender systems. Ranging from e-commerce to online advertisements, recommender systems tend to be unavoidable in routine journeys. Fundamentally, such systems are emphasized by suggesting relevant items to end-users and are of great use in enhancing the engagement rate. However, there are several problems such as Cold Start and Data Sparsity that impact the efficacy of the recommender system. A wide range of factors is responsible for such issues. In a similar context, this paper focuses on conducting a systematic literature review and analysis of different algorithms that have been proposed by researchers and practitioners in the field of recommender systems. The preliminary purpose of this paper is associated with obtaining an in-depth and succinct understanding of a wide range of solutions that have emerged from different studies. Based on the algorithms mentioned in the literature, it has been observed that Singular Value Decomposition Plus Plus offers the best possible resolution to data sparsity issues.

Keywords Recommender system · Collaborate-based filtering · Sparsity · Accuracy · SVD

1 Introduction

In the last decade, recommendation systems have assumed an important role in online social media, e-commerce, and entertainment platform such as LinkedIn, YouTube, Research Gate, etc. [1]. Earlier, it was very difficult to find a suitable and efficient recommendation for the users. With the development of technology, recommender systems have grown exponentially in various fields of information and web applications. The number of datasets is available pertaining to the recommender systems which facilitates in generating and different preferences.

S. Bhardwaj (✉) · K. Kanwar · G. Gupta
Yogananda School of AI, Computer and Data Science, Shoolini University, Solan, H.P., India
e-mail: shivanibhardwaj2687@gmail.com

Recommender system is accumulation of effective tools that can be used for recommending future preferences of a set of products to consumers and appropriately predict the most probable items [2]. The recommender system consists of various approaches such as content-based filtering, collaborate-based filtering, and hybrid filtering. Collaborative Filtering (CF) plays a vital role in recommendation systems which can help to make recommendations based on users' interests and preferences by using the previous history. Most developers tend to utilize collaborative filtering as this technique provides the best preferences. To get efficient and appropriate preferences, different models have been encompassed in CF. Further, a large number of datasets is available that can potentially create data sparsity and scalability issues [3]. To analyze this problem and improve the quality of the data. Machine Learning techniques can be utilized. Further, recommender system and their techniques are discussed to make appropriate preferences as per users' needs. With the help of such techniques and models, better performance of the recommender system can be attained.

This paper has been fragmented into multiple segments, each of which potentially focuses on different aspects of Recommender System. The paper initiates with a fundamental and rudimentary description of recommender systems. Different types of systems have been thoroughly discussed in the primary section of the paper. Major classifications discussed in the paper include content-based filtering, collaborative-based filtering, and hybrid filtering. Further, the paper also emphasizes on the discussion of two major types of memory-based filtering which includes item-based filtering and user-based filtering. After a comprehensive description of the basic types of recommender systems, analysis and findings of data sparsity algorithms have been thoroughly discussed in the paper. Overall, the research paper provides a thorough and in-depth description about the concept of recommender system and the algorithms used for the purpose of resolution of data sparsity.

2 Related Work

In this section, we will discuss the recommender system and their types are discussed further in detail.

2.1 Recommender System: Classifications

Recommender system has evolved as a revolutionary concept that provides end users with the suggestions of information that would be highly useful to them [4]. The recommender system offers appropriate ways for providing personalized results. This system was predominantly employed in e-commerce and entertainment but nowadays it has grown in the field of research and academics. Fundamentally, recommender systems are those systems that predict the future of any items based upon
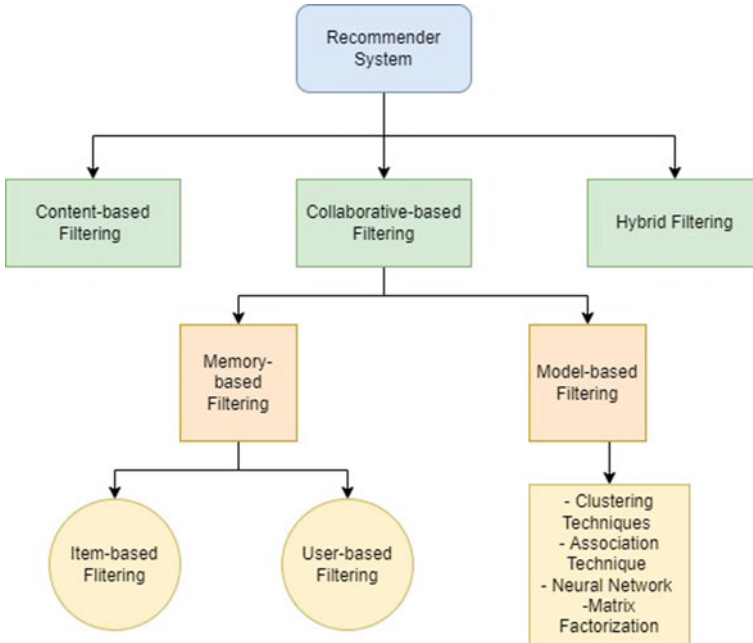
Fig. 1 Types and classifications of recommender system

the past behavior of end-user's [5]. In these days' machine learning introduced so many algorithms to predict recommendation as per previous preferences. All the classifications of recommender system are represented in Fig. 1 given and further explained in details.

2.2 Content-Based Filtering

The content-based recommender system is developed on single-user preferences. For example, on e-commerce websites, every individual tends to search according to their interest, and this user history is recorded as their past behavior. Further, the system examines the user's search history and then recommends similar choices to users [6].

2.3 Collaborating-Based Filtering

Collaborating is built upon users' historical behavior that including star ratings and reviews. This system used to build a frequent change in user's preferences. All your previous information is gathered over the internet then the system will make a

recommendation based on its analysis. This proves that collaborating filtering secures an important place in recommender systems [1].

Model-based collaborative filtering

This technique is beneficial for calculating the matrix factorization and this technique will be more efficient than memory-based collaborative filtering. Some Machine Learning approaches are included to make accurate predictions. Approaches include associate rule, decision tree, clustering, matrix techniques, etc. [7].

Memory-Based Collaborative Filtering

It entirely works with users' previous database to make a single prediction. For every single prediction, it consists of a preference database of user-item filtering and item-item filtering. A memory-based collaborating system is beneficial for making similarities between the two, due to the sparsity and scalability issue that comes under this scenario [7].

2.4 Hybrid Filtering

Different approaches are introduced in the recommender system and each of which has its functions and parameters. Still, there are some lags in the recommender system to improve for which hybrid filtering can be potentially used. Hybrid filtering is the merger of both content-based filtering and collaborating filtering and it works on system performance. The hybrid technique is used for resolving issues and enhancing the performance of the recommender system [8].

Althbiti et al. [7] the author has proposed item-based collaborative filtering and made use of the movies dataset to do a comparison between different clustering approaches. With the help of such approaches, the author wants to reduce the unpleasant data to remove the sparsity and scalability issues. A developer proposed a novel recommender system to improve the performance of reliability measures. This method is beneficial for collecting unreliable ratings and evaluating the results of reliability measures. In Anand and Bharadwaj [9], a simple approach is introduced to calculate the statistical classification of users' ratings and behavior. It uses both user-based and item-based which forms a hybrid approach and generates more accurate predictions that improve the performance time. In Guo et al. [10], a method is proposed to extract the unpleasant data to reduce the sparsity issues. As users tend to buy items based on personalized but improper and ineffective feedback could impact the process. To overcome this, the author introduced two methods: the first one is linear regression and the other is multidimensional similarities to tackle the sparsity data and make it reliable. An author proposed a technique particularly cross-domain and transfer learning to facilitate the similarity between distinct user profiles. Model-based and Memory-based collaborative filtering comes together to deal with the sparsity issue in the recommender system. To improve the performance of the recommender system by targeting certain user preferences to predict accurate results.

The author Zhang et al. [11], identifies that the collaborating filtering is suffering from sparsity issues as the number of products selling rapidly increases sparsity and rebuilding the bipartite graph to improve the accuracy and density of the network in the graph than the original one. Further, the author proposed clustering algorithms to handle the performance and accuracy. In Men et al. [12], approaches are introduced to check sparsity under different scenarios and compare those approaches at different levels to enhance input. The long short-term memory algorithm is introduced to analyze the performance time under the same circumstances. Those items which are preferable to sparse values are eliminated from the recommender system to get more accurate recommendations. In Sharifi et al. [13], the author proposed a new algorithm to handle the sparsity issue by using non-negative matrix factorization to predict better results as compared to real data. a method is proposed to modify the collaborative based recommender system with the help of matrix factorization. This method proposed the incorporation Based recommender system to improve the issue of sparsity.

As rating data is not appropriate, using such an algorithm improves the prediction and provides accurate values to each. This will improve the best accuracy and recommendations.

3 Data Sparsity Resolution Algorithm in Recommender Systems and Their Analysis and Findings

In this section, we will compare algorithms which are used in the resolution of data sparsity. Discuss the algorithms used by different paper and then compare their results based on Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). This will help to overcome the sparsity issues in future research work and Figure out among all of the algorithms used by different papers which gave results. All algorithms and their results are discussing Table 1.

A sample of research is completed on the sparsity issue, to reduce the issue researchers propose so many learning algorithms such as SVD, k-means, ANN. Still, our research is not completing each and every paper of sparsity in the recommender system. Above table demonstrate the analysis of the algorithms to overcome the sparsity issue. With the help of this techniques 80% of sparsity issues are removed but there is still 20% sparse data which can potentially affect the recommendation process. Based on the evaluation of different studies, appropriate comparisons have been made between various algorithms. Each algorithm has been explained with their results but it has been observed that singular value decomposition plus plus (SVD++) results provide best accuracy among all the techniques. This approach gives less error in terms of Root Mean Square Error (RMSE)—0.92 and Mean Absolute Error (MAE) is 0.72 [13].

Table 1 Comparison of algorithms

Author	Algorithms	Description	Results
Anwar and Uma [1]	Singular value decomposition plus plus (SVD++)	Comparing proposed algorithm with some machine learning approaches name as SVD and KNN	Provide less root mean square error (0.92) and mean absolute error (0.721) will improve the sparsity issue and cold-start issue
Gong et al. [14]	Memory-based collaborative filtering and model-based collaborative filtering	Comparing these two algorithm on basis of rating of items which include sparsity and scalability issues	The performance of the real time collaborative based recommender system is better than mean absolute error collaborative filtering
Koohi and Kiani [15]	Fuzzy c-means clustering algorithm is used	Comparing fuzzy c-mean clustering with K-mean and SOM clustering to get better accuracy and precision	By using fuzzy c-mean clustering improving the accuracy of the recommender system
Koohi and Kiani [16]	Proposed map reducing and clustering approach	Proposed approaches are implemented on different datasets	Such techniques are helpful to check the accuracy of recommendation based on different data set
Xie et al. [17]	N/A	The paper has conducted a systematic literature review to investigate, assess, evaluate, and discuss contributions pertaining to concepts and potential tools to eliminate data sparsity	The author emphasized on the importance of sparsity in recommender systems to propose solutions for enhancing overall efficacy and accuracy of the system
Kolahkaj et al. [18]	Dynamic contextual modeling approach	Use of density-based clustering algorithms to discover key area-of-interest to alleviate sparsity problem	Proposed mechanism outperforms previous works on metrics of precision, recall, and F-score
Althbiti et al. [7]	Proposed hybrid filtering model with ANN and Clustering algorithms	Use the k-fold and cross validation to improve the overfitting	Improve the precision, recall, ROC and F-score while adding the value in ANN

(continued)

Table 1 (continued)

Author	Algorithms	Description	Results
Zhang et al. [11]	Collaborative filtering with social behavior of users	Use of social data pertaining to different users to deal with data sparsity	The proposed method employs computational model to make predictions by calculation of sum of rating and assessment of rating patterns
Sharifi et al. [13]	Algorithm proposed non negative matrix and SVD	This algorithm is used to predict the sparse rating value and check their accuracy	Compare the non-negative matrix with SVD to get better results
Sahu and Dwivedi [19]	User profile as bridge in cross-domain recommender systems (UP-CDRSs)	UP-CDRSs transfers knowledge between various domains with the help of user profiles. In addition, probabilistic graphical model has been used to maximize posterior probability	The results tend to demonstrated that proposed method outperformed others with and without learning methods to a great extent in terms of accuracy

4   Conclusion

To summary, this paper has focused on conducting a thorough and rigorous analysis and evaluation of multivariate research papers. The primary objective of this paper has been to understand the potential trends pertaining to sparsity-related investigations and gain a strong understanding of key solutions that have been proposed by different authors. This work is an output of a broad search strategy which has been executed on a complete range of platforms to gather succinct information for performing the necessary analysis and evaluation. The study has heavily focused on understanding different algorithms proposed by a wide range of authors. In future research the algorithms can be deployed to make comparison between algorithm in case of data sparsity issue, So, that data sparsity issue get resolved. For further research, a more practical approach can be employed to gain a strong and in-depth understanding regarding the working of various algorithms to propose effective solutions.

References

1. Anwar T, Uma V (2021) Comparative study of recommender system approaches and movie recommendation using collaborative filtering. Int J Syst Assur Eng Manage 12(3):426–436. Available https://doi.org/10.1007/s13198-021-01087-x. Accessed 25 Nov 2021
2. Geetha M, Renuka D (2019) Research on recommendation systems using deep learning models.

Int J Recent Technol Eng (IJRTE) 8(4):10544–10551. Available https://doi.org/10.35940/ijrte.d4609.118419. Accessed 2 July 2022

3. Ahmadian S, Afsharchi M, Meghdadi M (2019) A novel approach based on multi-view reliability measures to alleviate data sparsity in recommender systems. Multimedia Tools Appl 78(13):17763–17798. Available https://doi.org/10.1007/s11042-018-7079-x. Accessed 25 Nov 2021

4. Dhanda M, Verma V (2016) Recommender system for academic literature with incremental dataset. Proc Comput Sci 89:483–491. Available https://doi.org/10.1016/j.procs.2016.06.109. Accessed 2 July 2022

5. Lu J, Wu D, Mao M, Wang W, Zhang G (2015) Recommender system application developments: a survey. Dec Support Syst 74:12–32. Available https://doi.org/10.1016/j.dss.2015.03.008. Accessed 25 Nov 2021

6. Lops P, Jannach D, Musto C, Bogers T, Koolen M (2019) Trends in content-based recommendation. User Model User-Adapted Interaction 29(2):239–249. Available https://doi.org/10.1007/s11257-019-09231-w. Accessed 25 Nov 2021

7. Althbiti A, Alshamrani R, Alghamdi T, Lee S, Ma X (2021) Addressing data sparsity in collaborative filtering based recommender systems using clustering and artificial neural network. In: 2021 IEEE 11th annual computing and communication workshop and conference (CCWC), 2021. Available https://doi.org/10.1109/ccwc51732.2021.9376008. Accessed 25 Nov 2021

8. Logesh R, Subramaniyaswamy V (2021) Exploring hybrid recommender systems for personalized travel applications. Cognitive Inform Soft Comput:535–544. Available https://doi.org/10.1007/978-981-13-0617-4_52. Accessed 25 Nov 2021

9. Anand D, Bharadwaj K (2021) Utilizing various sparsity measures for enhancing accuracy of collaborative recommender systems based on local and global similarities. Expert Syst Appl 38(5):5101–5109. Available https://doi.org/10.1016/j.eswa.2010.09.141. Accessed 25 Nov 2021

10. Guo G, Qiu H, Tan Z, Liu Y, Ma J, Wang X (2017) Resolving data sparsity by multi-type auxiliary implicit feedback for recommender systems. Knowl-Based Syst 138:202–207. Available https://doi.org/10.1016/j.knosys.2017.10.005. Accessed 25 Nov 2021

11. Zhang F, Qi S, Liu Q, Mao M, Zeng A (2020) Alleviating the data sparsity problem of recommender systems by clustering nodes in bipartite networks. Expert Syst Appl 149:113346. Available https://doi.org/10.1016/j.eswa.2020.113346. Accessed 25 Nov 2021

12. da Silva J, de Moura Junior N, Caloba L (2018) Effects of data sparsity on recommender systems based on collaborative filtering. In: 2018 international joint conference on neural networks (IJCNN). Available https://doi.org/10.1109/ijcnn.2018.8489095. Accessed 25 Nov 2021

13. Sharifi Z, Rezghi M, Nasiri M (2014) A new algorithm for solving data sparsity problem based on Non negative matrix factorization in recommender systems. In: 2014 4th international conference on computer and knowledge engineering (ICCKE). Available https://doi.org/10.1109/iccke.2014.6993356. Accessed 25 Nov 2021

14. Gong S, Ye H, Tan H (2009) Combining memory-based and model-based collaborative filtering in recommender system. In: 2009 Pacific-Asia conference on circuits, communications and systems, 2009. Available https://doi.org/10.1109/paccs.2009.66. Accessed 25 Nov 2021

15. Koohi H, Kiani K (2016) User based collaborative filtering using fuzzy C-means. Measurement 91:134–139. Available https://doi.org/10.1016/j.measurement.2016.05.058. Accessed 25 Nov 2021

16. Koohi H, Kiani K (2020) Two new collaborative filtering approaches to solve the sparsity problem. Cluster Comput 24(2):753–765. Available https://doi.org/10.1007/s10586-020-03155-6. Accessed 25 Nov 2021

17. Xie F, Chen Z, Shang J, Fox G (2014) Grey forecast model for accurate recommendation in presence of data sparsity and correlation. Knowl-Based Syst 69:179–190. Available https://doi.org/10.1016/j.knosys.2014.04.011. Accessed 25 Nov 2021

18. Kolahkaj M, Harounabadi A, Nikravanshalmani A, Chinipardaz R (2021) Incorporating multi-dimensional information into dynamic recommendation process to cope with cold start and

data sparsity problems. J Amb Intell Human Comput 12(10):9535–9554. Available https://doi.org/10.1007/s12652-020-02695-4. Accessed 25 Nov 2021

19. Sahu A, Dwivedi P (2019) User profile as a bridge in cross-domain recommender systems for sparsity reduction. Appl Intell 49(7):2461–2481. Available https://doi.org/10.1007/s10489-018-01402-3. Accessed 6 Dec 2021

20. Ribeiro J, Carmona J, Mısır M, Sebag M (2014) A recommender system for process discovery. Lecture notes in computer science, pp 67–83. Available https://doi.org/10.1007/978-3-319-10172-9_5. Accessed 2 July 2022