Abstract: The current generation is finding it difficult to find the right information from the enormous amount of data they are presented with in the online platforms. It is hard to spend time online searching for information in such a scenario and it craves for the need of an information filtering system that could help them discover the information they seek. A research field that does this has emerged in the last few years called as recommender systems. A lot of extensive research is happening in the field which is trying to incorporate more attributes to give more precise and relevant personalised recommendations to a user. This paper is focused on reviewing some significant works in the three basic recommender system types including collaborative filtering, content based filtering and hybrid filtering. The paper also have identified and listed the major challenges faced by recommender systems. The main contribution of the paper is in proposing a novel hybrid recommender system which addresses the sparsity and serendipity drawback of recommender systems. The proposed method is expected to deliver more accurate, relevant and novel predictions.

Keywords: Recommender Systems; collaborative filtering; content based filtering; hybrid filtering; K-nearest neighbours

I. INTRODUCTION

The explosive growth of internet has resulted in a phenomenon known as information abundance. In a way we are drowning in information but starving for knowledge, and it is mainly due to influx of data into the internet caused by people on one side and the scarcity of techniques to process the data to knowledge on the other side. So the current scenario demands new techniques that can assist us to discover resources of interest among the enormous options we are presented with. All of this paved way for the introduction of recommender systems which attempt to recommend items of interest to particular users by predicting a user’s interest in an item based on related information about the items, the users and the interactions between items and users[1].

The first research paper in recommender systems came out in the mid 1990s and since then research in this area got diversified and various approaches [2] were introduced to present better recommendations. Recommender System algorithms basically performs information filtering and can be classified into three types, namely collaborative filtering, content based filtering and hybrid filtering [3][4]. With time newer strategies evolved from the basic categories with improved recommendations by including the social information, information from internet of things, location information, and genetic algorithm based methods etc. A lot of work has happened in this area over the last decade on both industry and academia. Recommender systems still remains an area of high interest as it constitutes a problem-rich area and the possibilities it offer for practical applications. A wide range of applications including recommendations in web search, books, movies, music, restaurants, food, apparels, vehicles, targeted advertisements, medicines, news, potential customers for companies and many more. Recommender systems are widely used by e-commerce sites to improve the user experience and there by benefitting the stores. The system is able to convert browsers to buyers and cross-sell more items by means of suggestions while shopping. It increases the user loyalty by enabling them to purchase items in fewer clicks and also providing the frequent customers with good deals and offers. In short a recommender system is able to attract the interest of the customers by providing them fast and accurate recommendations.

A. Recommender System Fundamentals

Recommender Systems are able to neutralize the effect of information overload to a great extend by filtering vital information fragment from the large amount of dynamically generated information. A recommender system is intelligent enough to predict for a user his preference of one item over another [5]. It is this property of Recommender systems that enables it to give personalized recommendations to users. A recommender system takes into account a combination of multiple factors to provide good recommendations. They include the type of data available for the system, the algorithm used for filtering, model used, the technique used including Bayesian networks, genetic algorithms, probabilistic approaches, nearest neighbor strategy etc. The results of the recommender system is also influenced by the system performance, sparsity of the database, objective of the system and finally the quality of results the system has targeted [6].

B. Types of Recommender Systems

A broad classification of recommender system classifies it into three categories, namely, Collaborative filtering, Content-based filtering and Hybrid filtering. Figure 1 gives an outline of the classification.
which gives the details of the items with which the user has interacted in the past along with its attributes. The more commonly occurring attributes among multiple objects for the user is weighted high over the others. These attribute weighting along with the history of the user is used to make a user preference model. This model is compared with all objects in the database and scores are assigned based on its similarity to the user profile. Recommendations are made based on this scoring [12] [13].

c) **Hybrid Filtering**

This method of filtering combines the advantages of both collaborative and content based filtering and can avoid their individual limitation [7]. There are different possible ways of combining collaborative and content based filtering methods into a hybrid system. The classifications are as follows:

1. collaborative and content based filtering implemented separately and later combining their results.
2. Use of collaborative filtering properties in content based method and vice versa.
3. A combination model which developed which combines both collaborative filtering and content based filtering properties.

This paper attempts to review the various algorithms that were proposed for providing recommendations and classifying them based on the methodology used and the application area. Many papers in the area of recommender systems that were published in the past few years were collected and studied. The review also proposes a novel recommender technique which is expected to offer better results over the existing popular methods.

The various sections in the paper are organized as section II discusses the various existing methods followed by the proposed work presented in section III. In the next section various challenges addressed by recommender systems are listed. And finally in section V the conclusion and future scope is presented.

II. **BACKGROUND STUDY**

The initial works on recommenders were using collaborative filtering that recommended news articles to users [14] and music album and artist recommendations from social information [15]. It was followed by a lot of works in the field of recommender systems which helped users to find products, services or content such as books, movies, music, television shows, electronic or digital products etc by applying various algorithms which reviews the different users and items to give proper suggestions [16] [17] [18]. The research works in the field of recommender systems will be organized in this paper based on its type.

A. **Collaborative filtering**

Majority of the works in recommender systems is concentrated on collaborative filtering based techniques. A work proposed by Sarwar et al [19] makes use of the entire user database and also applies statistical methods on the database to find out similar users who share similar interest. G Zhuo et al [20] proposed a framework which combines both collaborative filtering and case based reasoning to improve the recommendations of the system. They have made use of two different algorithms MIFA and RAA to ensure the improved performance and validated the same. In [21] the method was able to predict the votes of the active user based on partial information about the user and the weights calculated from user.

![Classification of Recommender System approaches.](image-url)
A personalized collaborative filtering was proposed in [23] that apply to web services implemented by computing the similarity. A hybrid collaborative technique was developed by them which combines both user and item based concept. Qian Wang et al [24] developed a user model which uses a combination of demographic information and item combination features. The model searches for a set of neighboring users who share similar interest. The accuracy is improved by using genetic algorithm to compute the weights resembling similarity among users. An Association Cluster Filtering (ACF) was proposed in [25] which uses ratings matrix to establish cluster models and assumes users in the same cluster share similar interests and different users in different clusters have less interests in common. Unknown rating prediction is possible if an item in a cluster has more ratings to its credit. It will also enable to deduce conclusions about the item. This works well on a sparse dataset. In [26] a cascading hybrid approach was proposed which combines the features, demographic information and ratings about an item and claimed to have addressed the shortcomings of both collaborative and content based filtering. A method that was proposed in [27] adds the concept of time context to its collaborative filtering algorithm. This enhancement has improved the performance and accuracy of the recommendations. Another method [28] effective on sparse data was proposed by Ibrahim et. al. used a combination of global data and item based values to provide better results. This score was used in objection to the explicit ratings which were normally used. The results showed significant improvement over the Netflix’s system for movie recommendations. Netflix also conducted a very popular competition [32] aimed at improving its existing algorithm.

B. Content-based Filtering

The very first works on content based filtering was expected to be the contributions of [29] [30] [31] which were information retrieval and filtering techniques and later on it was extended by other researchers to introduce more innovations. Normally content based techniques are used on text based data for their recommendations and the content being mainly contributed by the keywords. A Fab system proposed by M. Balabanovic et al. [33] recommends web pages and it achieves its recommendation by representing web page content with the 100 most important words. Another work [34] again that recommends documents uses the most informative 128 words to represent a document. The importance of a keyword is calculated by using a weighting scheme which can be implemented in many ways, but the popular one being term frequency/inverse document frequency (TF-IDF) measure [35]. Content based recommender system derives its recommendations mainly based on the previous ratings of the user and hence it maintains a content based profile for every user. In order to build the content based profile many techniques are available one of which Rocchio algorithm [36] and it is using an averaging approach that calculates the average vector from individual content vectors. Another work [34] estimates the probability of an item to be liked by the user using Bayesian classifier. If an item has listed many features a work by N. Littlestone et al. [37] has exhibited good results. Other machine learning techniques like clustering and neural networks can also be used [34]. Other works in the field of text retrieval has also contributed to content based filtering research, one such being adaptive filtering [38][39] which identifies relevant documents by scanning the documents one by one from a stream of documents. Another work by S. Robertson and S. Walker [40] uses a threshold to determine the relevance of a document to the user. The query has to satisfy a certain degree of match with the documents to become relevant.

C. Hybrid Recommender Systems

Hybrid recommender system is gaining popularity recently as it is not confined to one method alone and uses a combination of methods to offer better results and accuracy. It is able to counter the disadvantage caused by using a single method. Robin Burke [49] observed that any hybrid technique that is used will fall under one of the seven categories namely weighted, switching, mixed, feature combination, cascade, feature augmentation and meta-level.

A work in [41] used a combination of single Valued Decomposition technique and demographic information to improve the collaborative filtering technique. A.B. Barragáns-Martínez et. al. [42] proposed a method which combines the properties of both collaborative and content based filtering. Genetic algorithms [43] have also inspired works on hybrid filtering. Another technique proposed by Al-Shamri et. al. in [44] is a hybrid system. It made use of a fuzzy based genetic approach. A method demonstrated by M. Lee and Y. Woo [45] used a collaboration of neural networks and collaborative filtering concepts. An effective use of Bayesian networks was used in [46] for implementing a hybrid approach with offered better results over the existing collaborative or content based schemes on individual implementation. A clustering algorithm based on centering-bunching was used in [47] to implement a hybrid personalized recommender system. M. Saranya and T. Atsuhiro [48] came up with their version of hybrid system by using latent features which was highly appreciated.

III. PROPOSED WORK

A close review of different works in recommender systems has revealed many shortcomings and in this paper, we propose a hybrid recommender approach which we believe will be able to address some of the major shortcomings of recommender systems. Figure 2 illustrates the proposed system.

The proposed system uses the user item rating matrix to start with and then generates fuzzy similarity matrix for both the user and item. The fuzzy similarity matrix will hold more information about the user as well as item capable of giving better predictions. Also, the fuzzy matrix tends to be less sparse when compared to the normal user rating matrix. Then we apply dimensionality reduction on the matrix to reduce the sparsity further. The resultant data is used to perform fuzzy C means clustering which will help us identify and group similar users and items. So the neighbor identification will reveal more related items and users. In the next step we generate individual predictions for both the users as well as items. Their results are combined in the aggregation phase where we rank the results and then retrieve the top N recommendations so as to provide them as recommendations to the user. In this proposed method close attention has been provided in various intermediate stages to address and counter the challenges that affect the performance of a good recommender system. We expect that the counter measures will eventually lead to better predictions which are novel and more accurate.
IV. CHALLENGES IN RECOMMENDER SYSTEMS

Recommender System recommendations are not perfect and it faces many shortcomings out of which few are listed below.

A. Cold Start Problem

It refers to a situation where the recommender system is not able to make relevant predictions or recommendations due to the lack of initial ratings about a user or an item [49]. It can commonly occur in two situations; a new item or a new user gets added into the system. The new item problem occurs because a new item added in the system is not having any ratings initially [50][51]. The probability of recommending an unrated item is very low and hence they might go unnoticed. One of the possible ways of tackling this situation is by having a set of motivated users who will be responsible for rating every new item.

The reason for new user problem is the lack of ratings for a user new into the environment. In that scenario, it is not possible to recommend anything to such a user [52] [53]. Also when the user enters their first ratings into the system they expect to start getting recommendations which does not happen. It is because the number of ratings given to the system by the user is not sufficient enough to make good recommendations. So the probability of a new user leaving the system is high.

B. Sparsity

Sparsity is yet another problem encountered by recommender systems and this occurs mainly due to the fact that the no of items available to be rated is very high when compared to the number of items already rated by the user. So when a user item matrix is populated only a very few entries will be marked which causes the matrix to be sparse leading to poor recommendations [54]. One of the possible solutions to this problem is by giving recommendations to a user by referring to the similarity in user profiles which assumes that if two users share similar interests, it is not really necessary to deduct conclusions solely based on the similarity of items they rated. This type of filtering is known as demographic filtering [55]. Another method of addressing the sparsity problem was proposed in [56], which used Singular Value Decomposition (SVD) to reduce the dimensionality of sparse rating matrix.

C. Scalability

Scalability issue arises as the number of users, items and ratings information grows day by day. Even with the growing amount of information recommender systems are expected to respond quickly with recommendations for the online customers and it demands a higher scalability. The implementation of such system becomes complex and costly. The key challenge is in designing an efficient learning algorithm which is capable of handling such large datasets which keeps on growing. One of the solutions proposed is to use an online learning algorithm [57] which processes the updates related to each user immediately and sequentially. Another method [58] proposed to address the scalability issue uses a distributed algorithm where the computations are done in parallel in multiple machines.

D. Overspecialization

This is one of those challenges of recommender system which causes a user to lose interest of the system. Here the items similar to those rated high by the user is given as recommendations which also means that the user might have already bought or experienced the item. Hence the recommendations will not shed much interest on the user and there is a very high probability that user might leave using the system because it is not able to be of much use to the user. One way of handling the issue was proposed in [59] which uses a neighborhood based collaborative filtering. Other solutions include introducing some randomness using various randomness measures, using genetic algorithms or by eliminating similar items.

E. Serendipity

Serendipity is a very crucial objective that every recommender system strives to achieve. It is all about gaining the user trust and loyalty. The user will be provided with novel and relevant recommendations which are significantly different from the items that the user has already rated. It is difficult to conceive the idea of serendipity completely as the concept itself is very subjective and such encounters are very rare in real world scenarios. It is worth noting that there is no consensus on which serendipity definition and evaluation metric to be used. Various solutions have been proposed which attempts to introduce serendipity in the recommendations; they include re-ranking the results of any accuracy oriented algorithms [59] to produce relevant scores, Full Auralist [60] algorithm which
generates rank and integrates them into the ranked list of items using their linear combination etc.

V. CONCLUSION AND FUTURE WORK

The research in recommender system is directed on the right path of improving the relevance and accuracy of personalized recommendations. Many promising works are also proposed and implemented in the last few years. But the challenges faced by recommender systems were not addressed completely and there is a lot of room for improvement. This paper has attempted to list some of the significant works in the field and propose a novel hybrid approach that can confront some of the drawbacks of recommender systems. This paper has made use of fuzziness, dimensionality reduction and clustering approach to improve the recommendation quality. Also rather than focusing just on the user or item at a time this method has used information from both and aggregated them to deliver better results. At the last phase top n recommendations approach was also used which is a well proven recommendation method. Future works can explore more possibilities of using other heuristic techniques and genetic algorithms in improving the accuracy of recommendations. Also data mining techniques can also be experimented with to improve the initial database filtering process for a better input. Always a good input leads to better results.

VI. REFERENCES

[1] J. Bobadilla, F. Ortega, A. Hernandez, A. Gutierrez, Recommender systems survey, Knowledge-Based Systems, 46 (2013) 109-132.

[2] Deshpande, M., and Karypis, G. Item-based top-n recommendation algorithms. ACM Trans. Inf. Syst. 22, 1 (2004), 143-177.

[3] M. Pazzani, A framework for collaborative, content-based, and demographic filtering, Artificial Intelligence Review-Special Issue on Data Mining on the Internet 13 (5-6) (1999) 393–408.

[4] R. Burke, Hybrid recommender systems: survey and experiments, User Modeling and User-Adapted Interaction 12 (4) (2002) 331–370.

[5] Pan C, Li W. Research paper recommendation with topic analysis. In Computer Design and Applications IEEE 2010;4, pp. V4-264.

[6] J. Bobadilla, F. Ortega, A. Hernandez, A. Gutierrez, Recommender systems survey, Knowledge based systems, Vol. 43, 2013, pp 109-132.

[7] M. Balabanovic and Y. Shoham, “Fab: Content-Based, Collaborative Recommendation,” Comm. ACM, vol. 40, no. 3, pp. 66-72, 1997.

[8] Burke R. Hybrid recommender systems: survey and experiments. User Model User-adapted Interact 2002;12(4):331–70.

[9] Breese J, Heckerma D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the 14th conference on uncertainty in artificial intelligence (UAI-98); 1998, p. 43–52.

[10] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. 2001. Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international Conference on World Wide Web (Hong Kong, Hong Kong, May 01 - 05, 2001). ACM, New York, NY, 285-295. DOI= http://doi.acm.org/10.1145/371920.372071

[11] Xiao yuan Su, Taghi M. Khoshgoftaar, A survey of collaborative filtering techniques, Advances in Artificial Intelligence archive, 2009.

[12] Aggarwal, Charu C. (2016). Recommender Systems: The Textbook. Springer. ISBN 9783319296579.

[13] Peter Brasilovsky (2007). The Adaptive Web. p. 325. ISBN 978-3-540-72078-2.

[14] Resnick, P., Iakouvo, N., Sushak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of netnews. Computer Supported Cooperative Work Conf.

[15] Shardanand, U., & Maes, P. (1995). Social information filtering: Algorithms for automating 'Word of Mouth'. Human Factors in Computing Systems Conf.

[16] Frias-Martinez, E., Chen, S. Y., & Liu, X. (2009). Evaluation of a personalized digital library based on cognitive styles: Adaptivity vs. adaptability. International Journal of Information Management, 29, 48–56.

[17] Frias-Martinez, E., Magoulas, G., Chen, S. Y., & Macredie, R. (2006). Automated user modeling for personalized digital libraries. International Journal of Information Management, 26, 234–248.

[18] Kim, J. K., Kim, H. K., Oh, H. Y., & Ryu, Y. U. (2010). A group recommendation system for online communities. International Journal of Information Management, 30, 212– 219.

[19] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, “Item based collaborative filtering recommendation algorithms,” in Proceedings of the 10th International Conference on World Wide Web (WWW '01), pp. 285–295, May 2001.

[20] Guangping Zhuo, Jingyu Sun and Xue Li Yu “A Framework for Multi-Type Recommendations”, Eighth International Conference on Fuzzy Systems and Knowledge Discovery, 2007.

[21] John S., Breese, David Heckerman and Carl Kadie. 1998. (July), Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence.

[22] Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., and Riedl, J. 1997. GroupLens: applying collaborative filtering to Usenet news. Comm. ACM 40, 3 (Mar. 1997), 77-87.

[23] Yechun Jiang, Jianxun Liu, Mingdong Tang and Xiaoqing (Frank) Liu “An Effective Web Service Recommendation Method based on Personalized Collaborative Filtering”, 2011 IEEE International Conference on Web Services.

[24] Qian Wang, Xianhu Yuan, Min Sun “Collaborative Filtering Recommendation Algorithm based on Hybrid User Model”, FSKD, 2010.

[25] Chuangguang Huang and Jian Yin “Effective Association Clusters Filtering to Cold-Start Recommendations”, 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery.

[26] Mustansar Ali Ghazanfar and Adam Prugel-Bennett, ” A Scalable, Accurate Hybrid Recommender System”, 2010 Third International Conference on Knowledge Discovery and Data Mining.

[27] Liang He and Faqing Wu, “A Time-context-based Collaborative Filtering Algorithm”, IEEE International Conference on Granular Computing, 2009

[28] Ibrahim A. Almosallam and Yi Shang “A New Adaptive Framework for Collaborative Filtering Prediction”, 2008 IEEE Congress on Evolutionary Computation (CEC 2008).

[29] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. Addison-Wesley, 1999.

[30] G. Salton, Automatic Text Processing. Addison-Wesley, 1989.

[31] N. Belkin and B. Croft, “Information Filtering and Information Retrieval,” Comm. ACM, vol. 35, no. 12, pp. 29-37, 1992.

[32] James Bennett; Stan Lanning (August 12, 2007). “The Netflix Prize”. Proceedings of KDD Cup and Workshop 2007.
...