International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Recommendation System Using Community Detection for Social Media E-Commerce

Sundar Rajan S[1], Dr. K.Thirunadana Sikamani[2]

[1] Professor, School of Information Technology, SRM University Sikkim, Gangtok, East Sikkim.
[2] Professor, Department of Computer Science & Engineering, St. Peter's Institute of Higher Education and Research, Chennai.

Abstract
The Recommendation System or the personalization system in social networks like Facebook plays vital role in product Marketing. People are mostly addicted nowadays over social networks such as Facebook, Twitter, Instagram, and so on. Thus, the data gathered from social networks like Facebook can be leveraged for recommendation systems or any other systems that requires knowledge about users. In social networks like Facebook, users reveal considerable information about their preferences, feelings, activities, etc. This information can be very valuable in determining the actual needs and preferences of users. The main objective of this research is to design and develop a recommendation system for social networks through community detection especially in Facebook. Communities are mined by the influential algorithm, that can be one of the powerful algorithm to mine out the perfect community to forecast the products. In this paper, we study and analyze various clustering techniques used for product recommendation using social information, which are used to identify the concept to address the data sparsity problem, cold start issues, and in turn to improve the prediction accuracy for product recommendation.

Keywords: Collaborative Filtering, Content-based Filtering, Recommender Systems, Social Networks.

1. Introduction

Recommendation systems are one of the successful and widespread application of machine learning technologies in business. These algorithms are designed to anticipate the rating or preference a user might assign to a particular item. Their primary objective is to deliver the most pertinent information by identifying meaningful patterns within a dataset, ensuring a personalized and efficient user experience. Recommendation systems are mostly concentrating in the fewer areas such as products, movies, books, research articles, and social tags in general. There are also recommendation systems for social networking users, financial services and life insurance. In this view, Recommendation systems are typically classified into two categories such as content based filtering and collaborative filtering methods. Although some recommendation system combines both approaches, they found some statistical modification. Instead of providing a static experience in which users search for and potentially buy products, past purchases and searches, and on other users' behavior, we shall recommend product based on users purchase with their before & after consumption ratings and review, other user's top rating, viewing products.

In social networking, especially Facebook, targeting the group of people based on community can be used to recommend the product to gain huge response towards the product success. Selecting the community based on the product can be categorized into two most popular approaches among Recommendation System [26], they are collaborative and content-based recommendation algorithms. Here, the recommendations are performed based on the products that people with similar preferences and interests, involvement and their activities preferred previously. While in content-based filtering method, recommended items are with content exactly matched to previously preferred items by a target user [26]. The influential algorithm in this recommendation system are used to filter or mine the exact community, by analyzing the group of user's conversation, preference, likes, dislikes, and their activities. The proposed algorithm is used to mine frequent item set and newer version of association rule learning over social networking databases. The process begins by detecting frequent interactions among individuals within a group regarding specific items in the database. These interactions are then expanded into larger item sets that occur with significant regularity, ensuring valuable insights are derived from recurring patterns.

Content-based Recommender System, on the other hand, analyze item descriptions to find items that are of particular interest to the user [26]. They are comfortable with a well arranged framework for comparing user interests or preference with the items' specifications to suggest the most suitable item to a target user. Although this kind of recommendation methods facing the new items' issues, they quiet suffer from the cold-start problem in situations when new users are involved.

Most of the systems have more number of items rather than users that typically rates very small part of the items, which results in sparsity problem. Recommender systems that uses collaborative filtering experience this sparsity problem along with cold-start problem, which is another commonly faced one by the recommender system.

The cold-start problem is categorized as user based and item based. Cold-start challenge is caused by new items that are supposed to be recommended to users while there are not enough previously submitted ratings about them. Cold-start issue arises only when a new user who has already joined an E-shopping environment has presented just few opinions about the product. In this kind of situations, there is no interaction or communication happens between the new user and the other ones [26], and so it's not possible to find out the difference or similarity between them. As a result, the recommendation systems are unable to make reliable product recommendations. Both collaborative and content-based recommendation systems have a shortcoming related to cold-start problem. For reliable and effective recommendations, they require strict records of previous item ratings, need to know whether they are trusted user, the user how much related to product or well-known user. Because, collaborative methods cannot function properly for the new items that exist with no records on previous ratings. Although this cold-start problem happens, they need to recommend the new items too. This issue has been mitigated to some extent by content-based recommendation systems, which can predict item relevance even in the absence of prior ratings. Nevertheless, even content-based recommendation system suffers from cold-start issue, as they were unable to recommend products to new comers, if there were no records of interactions with the system previously.

One way to solve this data sparsity and cold-start problem is that a system can gather related information from other possible sources based on specific community to ensure accurate recommendations by replacing missing rating data. Likewise, it can use information about the preferences of the friends of new users.

2. Related Work

In this section, we review relevant research literature pertaining to collaborative filtering, recommender systems, data mining, and personalization. Tapestry is one of the oldest implementation of collaborative filtering based recommender systems. This system relies on the explicit opinions of people from a close-knit community, such as an office workgroup [18]. However, recommender system for large communities cannot depend on each person knowing the others. Later, several ratings based automated recommender systems were developed. The GroupLens research system [12,16] provides a pseudonymous collaborative filtering solution for Usenet news and movies. Ringo [16] and Video Recommender [7] are email and web based systems that generate recommendations on music and movies respectively. A special issue of communications of the ACM presents a various number of recommender systems where other technologies have also been applied to recommender systems, including Bayesian Networks, Clustering, and Horting. Bayesian Networks create a model based on a training set with a decision tree at each node and edges representing user information [27]. The model can be built off-line over a matter of hours. The resulting model is very small, very fast, and essentially as accurate as nearest neighbor methods [3]. Bayesian Networks may prove practically for environments in which knowledge of user preference changes slowly with respect to the time needed to build the model [1, 28]. But they are not suitable for environments where user preference model has to be updated frequently [19]. Clustering techniques work by identifying group of users who appear to have similar preferences. Once the clusters are formed, prediction for an individual can be made by aggregating the opinions of the other users in that cluster. Some clustering techniques represent each user with partial participation in several clusters [27]. The prediction is then an average across the clusters, weighted by degree of participation [28]. Clustering techniques usually produce less personal recommendations than other methods, and in some cases, the clusters have worse accuracy than nearest neighbor algorithms [3]. Since the size of the group that must be analyzed is much smaller, the cluster performance will be very good. Clustering techniques can also be applied as a "first step" for shrinking the candidate set in a nearest neighbor algorithm [31] or for distributing nearest neighbor computation across several recommender engines. While dividing the population into clusters may affect the accuracy or recommendations to users near the fringes of their assigned cluster [29]. Pre-clustering may be a worthwhile trade-off between accuracy and throughput [28]. Horting is a graph based technique in which nodes are users, and edges between nodes indicate degree of similarity [1, 27] between two users. Predictions are produced by walking through the graph to nearby nodes and combining the opinions of the nearby users. Horting differs from nearest neighbor as the graph may be walked through other users [28] who have not rated the item in question, thus exploring transitive relationships that nearest neighbor algorithms do not consider [19]. Using synthetic data, Horting produces better predictions than a nearest neighbor algorithm. Schafer et al., present a detailed taxonomy and examples of recommender systems used in E-commerce and how they can provide one-to-one personalization and at the same time can capture customer loyalty [18]. Although these systems have been successful in the past, their widespread use has exposed some of their limitations such as the problems of sparsity in the data set, problems associated with high dimensionality and so on. Sparsity problem in recommender system has been addressed in. The problems associated with high dimensionality in recommender systems have been discussed in [18], and application of dimensionality reduction techniques to address these issues has been dealt with. However, with the discovery of data mining tools and knowledge [4] which inherently is associated with databases more advanced and efficient approaches can be utilized to analyze large datasets and generate recommendations more swiftly and accurately. Our work explores the extent to which item-based recommenders, a new class of recommender algorithms, are able to solve these problems [19].

Ma et al. (Ma et al. 2011) proposes a method that forms a social composition, according to the common interests among the friends of a particular user. Following the same idea, a recommendation is made for a particular user based on the social influence of both close and distant friends (He & Jianming 2010). Oliveira et al. (Oliveira et al. 2012) built a Recommender System based on trust among friends, called trustaware recommender systems. In the same fashion, but considering the cold-start problem, Caron, and Bhagat (Caron & Bhagat 2013) use the information of a new user's friends in e-commerce and propose a learning model of user preference in the social network. Lalwani et al. (Lalwani et al. 2015) assume users that are part of a community have similar preferences, or are influenced by it. They utilize social interactions (friend connections) to detect Facebook communities and to simulate cold-start scenarios, recommending items already ranked by users in the community to the user without an assessment history.

Maniktala et al. (Maniktala et al. 2015) also assume users prefer items already acquired by friends, but the friends' influence depends on how strong the friendship relation is. They propose techniques to classify social relationships as strong or weak and recommend items acquired by users with strong friend connection to a new user. However, they consider cold-start for a user that has consumed up to 5 items, that can be considered to be a moderate cold-start once there is some information about the user. Felicio et al. (Felicio et al. 2016) use models that has been already built in systems for users to suggest items to a new user. They built a personalized model for a cold-start user by selecting prediction models from a set of strongly linked users. They handle several social network connection weight metrics to classify links among users.

The data mining field has considerably advanced in recent years due to technological advances providing the processing and storage of a large volume and variety of data. In particular, social networks are considered essential to this change, driving the creation of such data, which is generated by different users. Thus, individual efforts regarding techniques for extracting information on specific data types, such as text mining, become relevant for achieving results (Aggarwal & Zhai 2012) (Hu & Liu 2012) (Xia et al. 2016).

3. Methodology

3.1.Recommendation System

In social networks especially in Facebook, the Recommendation systems (RSs) are vastly used to provide products with high preference and highly rated quality products to the user from a large number of choices. Fig. 3.1 shows the overview of recommendation system on social network. Collaborative filtering (CF) is a commonly used technique to provide recommendations. Two major issues are arising ever in recommending products to the user. Those are, Data sparsity arises from the phenomenon that users in general rate only a limited number of items, Cold start refers to the difficulty in bootstrapping the RSs for new users or new items.

Figure 3.1: Overview of Recommender System in Social Networks

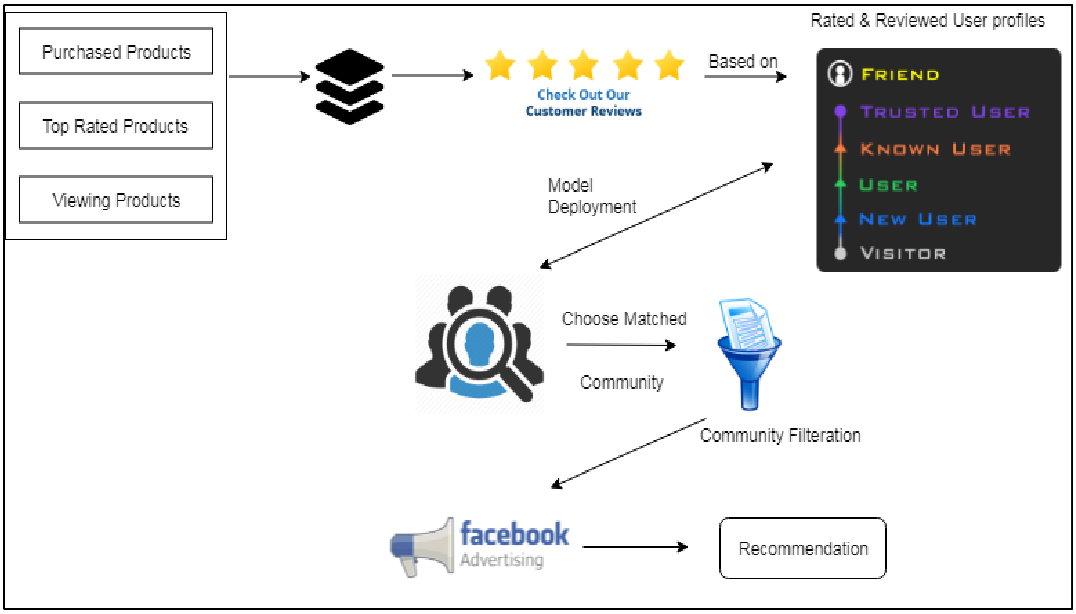


Table 3.1: Review Status and Rating of Users

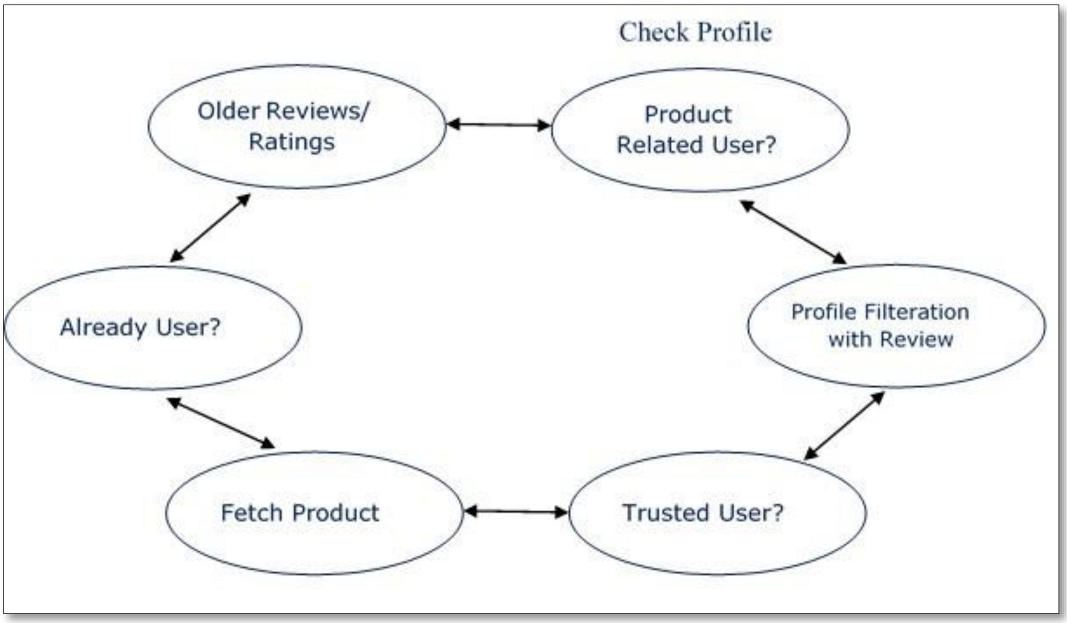
Users	Place	Status	Rating	Review State	Trust
Deva	Chennai	Buyer	5	Positive	Certified Buyer
Shabna	Jaipur	Buyer	2	Negative	Certified Buyer
Sonal	Mumbai	Viewer	4	Positive	Un Certified Buyer
Sankar	Chennai	Buyer	4	Positive	Certified Buyer
Akash	Rajasthan	Viewer	4	Positive	Un Certified Buyer
Jeethu	Delhi	Buyer	5	Positive	Un Certified Buyer

The Table 3.1 shows the review status and its users rating. Based on the user rating, the review state will change it may be negative or positive. For each user, the trust can be made by the review report.

3.2.Product Mining

The following categories of products are to be mined to obtaining brand loyalty, good quantity and quality of the product, customer satisfaction providing positive customer experience, increasing the number of product purchases and service subscriptions are the goals of any recommendation. This product mining should be based on reviews and ratings of the product. Before fetching the product, they need to check whether the reviews and ratings comes from the trusted well-known user by checking user profiles. These strong data from the user are used to overcome the cold-start issue and sparsity problem. The Fig. 3.2 shows the verification of reviews and ratings of the product buying the customer. The verification can be made by checking his old review and ratings. The profile filtration is based on the reviews only.

Figure 3.2: Verification of Reviews and Ratings



3.3.Face Book Community Fetching & Advertising

Facebook is to keep community Pages in the hands of their respective brand owners or by any other common people. Community Pages are used to provide opportunities to express their interest, involvement and creativity, while allowing for official Pages to continue representing official entities such as businesses, bands or celebrities. Community Pages are a type of Facebook Page dedicated to their preferred conversation or experience that is owned collectively by the community connected to it. Community Pages are make you connect with others who share similar interests and experiences. For making Beneficial of face book user, need to gather the user preference, feelings and activities. Initially, check the user is joining in which preferred community and monitor their activities continuously whether he/she would like to share their opinion positively or negatively in this community, whereas, how long he/she gives their interest or involvement on the same thing. Based on these criteria community are to be mined. Here, belongs the user interest the product advertisement is to be recommended. Table 3.2 shows the community mining in Facebook.

Table 3.2: Community Mining in Facebook

Community	**No. of Members**	**Involvement Towards**	**Spending Time/min**	**Major Sharing Opinion**
Sports	20	IPL	10	Lack of players
School	25	Books	15	Exams
Entertainment	42	Scary Movies	5	Movies
Fashion	18	Earrings	8	Varieties

4. Algorithm

Facebook data on the recommended domain k-NN

Similarity between users is computed using related data to the domain of Recommendation of product from collected Facebook preferences. Thus the similarity between two users, m and n, is defined as the mutual preferences normalized by the number of their aggregated preferences. This method attempts to use only Facebook data and does not involve any numeric rating provided by the user.

$$W_{m,n} = |I_m \cap I_n| / |I_m \cup I_n| \qquad (1)$$

Where, I_m represents the subset of items that have been found in the Facebook profile of user m. The system returns all data that are published by the users' friends with similarity above 0.5 (set empirically).

Singular value decomposition (SVD)

SVD CF methods converts both the users and items to a joint latent factor space. It becomes the main method of choice when prediction of users ratings is highly preferred. This singular value decomposition uses stochastic gradient descent procedure to solve least square error problem [29]. The predicted rating is calculated using the formula,

$$P_{a,i} = r'' + d_i + d_a + q_i^T p_a \qquad (2)$$

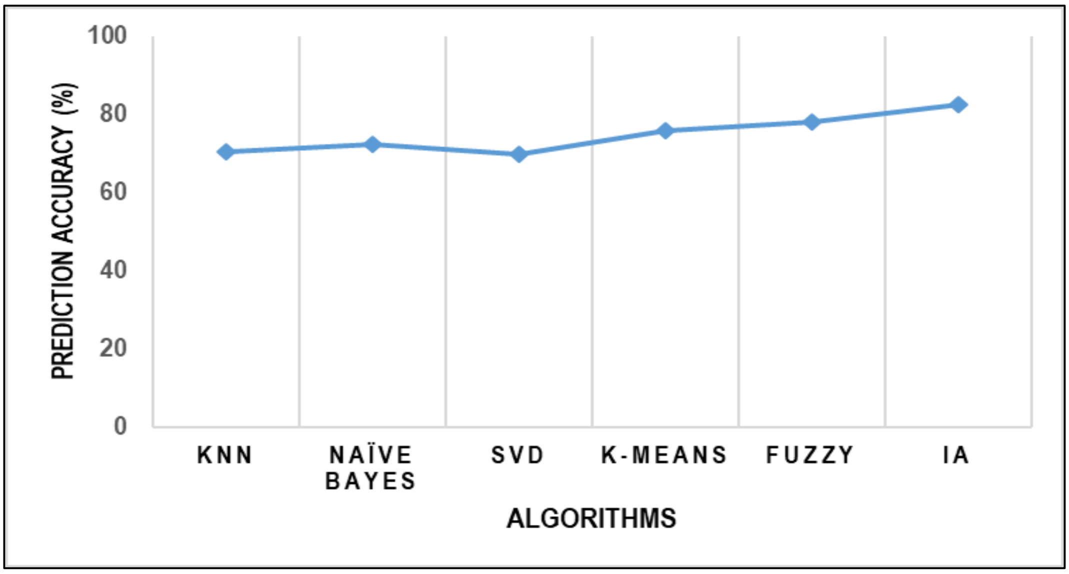
Where d_a and d_i represents the average deviations of user a and item i. Likewise r" implies the overall average rating. The vectors q_i, p_a evaluates how the user or item relate to every one of the latent factors. The experiments reveals that it provides best results when the number latent factors set to 5. It works well over a very large data set.

Our proposed algorithm works on two major steps, Joining and Pruning. In Influential Algorithm, the set of items which has minimum support are considered to be frequent itemset. And those are join with the property of any subset of frequent itemset must be frequent. To find a set of candidate item sets, it is generated by joining with itself. In proposed algorithm the set of priorities are to be taken for fetching the results.

5. Results and Discussion

Generation of Facebook data by the users have no more control over the quality of the data or the format that it uses for the content. To extract proper and accurate items from Facebook requires a more refined application. Those application will apply sentiment analysis based on communities to identify positive notifications or comments. To excerpt relevant domain based user preference information, we have viewed for certain information from the Facebook users account like profile data and links published by them which contains relevant keywords. The result of this process was, for each and every user an aggregated database was created which contains a set of preferred items based on domain interest. Finally, the result was tested by partitioning the data set randomly to different training and test sets.

Figure 5.1: Comparison of Prediction Accuracy of Product Recommendation over Various Algorithms



For the Recommendation system, we have tested the data sets with various algorithms like KNN, SVD, Naïve Bayes, k-means and proposed influential algorithm. The influencial algorithm gives more prediction accuracy than all those algorithm. The above figure represents the results of comparison of prediction accuracy of product recommendation produced by different algorithms.

Figure 5.2: Comparison Chart Depicting Time Taken for Mining Data by Different Algorithms Over Different Communities

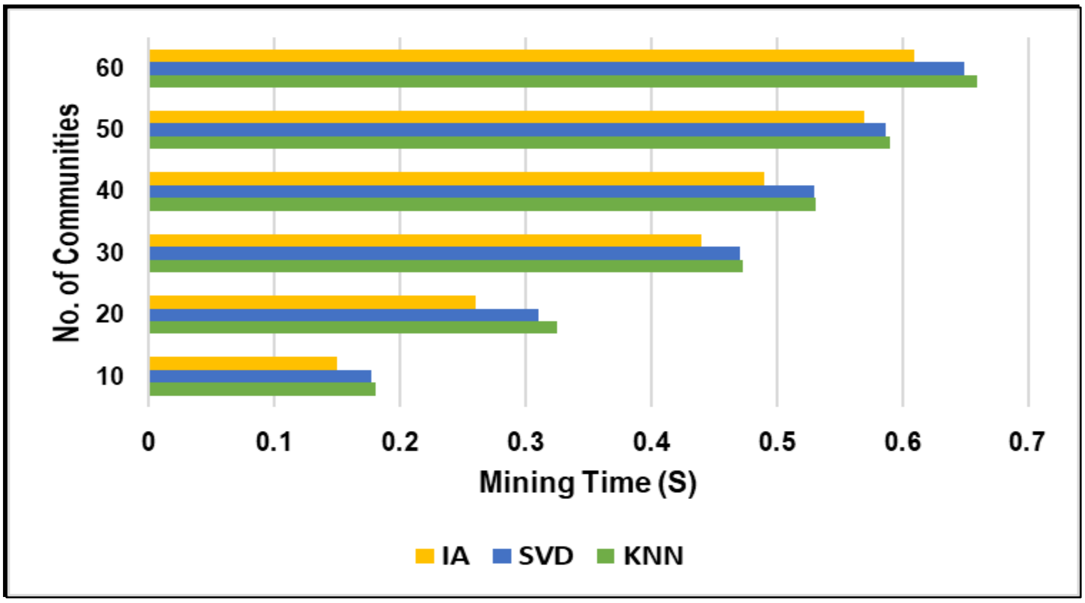


Figure 5.2 illustrares the time taken to mine matched data from Facebook communities. Compared to other algorithms, time taken by the influential algorithm is lesser than KNN and SVD algorithms. When comparing the time taken by KNN and SVD, there is no much difference. From the above results we can conclude that community based recommendation can provide better accuracy.

6. Conclusion

Recommendation or personalization systems are a powerful technology for extracting additional value from its user database. These systems help users find items they want to buy for their personal or professional use. This personalization system is rapidly becoming a crucial tool in E-Shopping on the Web. And also it is being stressed by the huge amount of user data in existing Facebook database. Hence, we chose the community page to recommend or advertise the products to make benefit of user, which is based on user's interaction, involvement and feelings, activities in those pages. Experimental results on this data set proves that our proposed system provides better prediction accuracy of recommender systems over the other algorithms. And also provides remedy to the data sparsity and cold-start issues inherent in collaborative filtering by gathering related information from other possible sources based on specific community to ensure accurate recommendations by replacing missing rating data and by using information about preferences of the friends of new users.

References

1. Aggarwal, C. C., Wolf, J. L., Wu K., and Yu, P. S. (1999). Horting Hatches an Egg: A New Graph-theoretic Approach to Collaborative Filtering. In Proceedings of the ACM KDD'99 Conference. San Diego, CA. pp. 201-212.
2. Aggarwal, C.C., and Zhai, C. (2012). An Introduction to Text Mining. In Mining Text Data. Boston, MA: Springer US, pp. 1–10.
3. Breese, J. S., Heckerman, D., and Kadie, C. (1998). Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, pp. 43-52.
4. Demiriz, A., (2004). Enhancing Product Recommender Systems on Sparse Binary Data. Data Mining and Knowledge Discovery, 9(2), 147–170. ISSN 1384-5810.
5. Deshpande, M., Karypis, G. (2004). Item-based Top-N recommendation algorithms. ACM Transactions on Information Systems, 22(1), 143–177. ISSN 1046-8188.
6. Goldberg, D., Nichols, D., Oki, B. M., and Terry, D. (1992). Using Collaborative Filtering to Weave an Information Tapestry. Communications of the ACM. December.
7. Hill, W., Stead, L., Rosenstein, M., and Furnas, G. (1995). Recommending and Evaluating Choices in a Virtual Community of Use. In Proceedings of CHI '95.
8. Hu, Y., Koren, Y., and Volinsky, C. (2008), Collaborative filtering for implicit feedback datasets. In ICDM, pages 263-272.
9. Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., and Riedl, J. (1997). GroupLens: Applying Collaborative Filtering to Usenet News. Communications of the ACM, 40(3), pp. 77-87.
10. Ma, H. (2011). Recommender systems with social regularization. In Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11. New York, New York, USA: ACM Press, p. 287.
11. Maniktala, M. (2015). Finding the most informational friends in a Social Network based Recommender System. In Annual IEEE India Conference (INDICON). New Delhi, pp. 1–6.
12. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In Proceedings of CSCW '94, Chapel Hill, NC.
13. Schafer, J. B., Konstan, J., and Riedl, J. (1999). Recommender Systems in E-Commerce. In Proceedings of ACM E-Commerce 1999 conference.

14. Schein, A.I., Popescul, A., Ungar, L.H., and Pennock, D.M. (2002). Methods and metrics for cold-start recommendations. In Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, pages 253–260.
15. Sedhain, S., Sanner, S., Braziunas, D., Xie, L., and Christensen. J. (2014). Social collaborative filtering for cold-start recommendations. In Proceedings of the 8th ACM Conference on Recommender systems, pages 345–348.
16. Shardanand, U., and Maes, P. (1995). Social Information Filtering: Algorithms for Automating ’Word of Mouth’. In Proceedings of CHI ’95. Denver, CO.
17. Su, X., and Khoshgoftaar, T.M (2009). A survey of collaborative filtering techniques. In Adv. In Artificial Intelligence.
18. Yu Li, Liu Lu, Li Xuefeng (2005). A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce. Expert Systems with Applications, 28(1), 67-77.
19. Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl (2001). Item-based Collaborative Filtering Recommendation Algorithms. Appears in WWW10, May 1-5, 2001, Hong Kong.
20. J. Ben Schafer, Joseph Konstan, John Riedl (1999). Recommender Systems in E-Commerce. In the Proceedings of the 1st ACM conference on Electronic commerce, Pages 158 – 166, https://doi.org/10.1145/336992.337035.
21. Nicole B. Ellison, Charles Steinfield, Cliff Lampe (2007). The Benefits of Facebook “Friends:” Social Capital and College Students’ Use of Online Social Network Sites. Journal of Computer-Mediated Communication 12(4):1143-1168.
22. Allan Douglas Rodrigues Oliveira, Leonardo Bessa, Thiago Rodrigues Andrade, Lucia Filgueiras (2012). Trust-Based Recommendation for the Social Web. IEEE Latin America Transactions 10(2): 1661-1666.
23. Caron, S. & Bhagat, S., (2013). Mixing bandits. In Proceedings of the 7th Workshop on Social Network Mining and Analysis - SNAKDD ’13. New York, New York, USA: ACM Press, pp. 1–9.
24. Deepika Lalwani, Dvln Somayajulu, Pisipati Radha Krishna (2015). A Community Driven Social Recommendation System. IEEE International Conference on Big Data. DOI: 10.1109/BigData.2015.7363828.
25. Felicio, C.Z. et al., (2016). Preference-Like Score to Cope with Cold-Start User in Recommender Systems. In 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE, pp. 62–69.
26. Vala Ali Rohani et al., (2014). An Effective Recommender Algorithm for Cold-Start Problem in Academic Social Networks. Mathematical Problems in Engineering.
27. Oloyede I., Hassan I.A. (2019). Evaluation of Accuracy Between Item-Based and Matrix Factorization Recommender System. Annals. Computer Science Series. 17th Tome 1st Fasc.
28. Charalampos Vassiliou, Dimitris Stamoulis, Drakoulis Martakos (2006). A hybrid content-based clustering architecture: minimising uncertainty in personalised multimedia content. International Journal of Intelligent Systems Technologies and Applications, Vol. 1, No. 3/4.
29. Chhavi Rana, Sanjay Kumar Jain (2014). An Extended Evolutionary Clustering Algorithm for an Adaptive Recommender System. Social Network analysis and Mining, 4(164).
30. Zan Huang et al., (2003). A Graph Model for E-Commerce Recommender Systems. Journal of the American Society for Information Science and Technology, 55(3): 259-274.
31. Ben Schafer J., Joseph A. Konstan, John Riedl (2001). E-Commerce Recommendation Applications. Data mining and Knowledge Discovery, Volume 5, pages 115–153.