

A Hybrid Approach to Solve Cold Start Problem in Recommender Systems using Association Rules and Clustering Technique

Hridya Sobhanam
Dept. of Information Technology
Easwari Engineering College
Chennai, India

A.K.Mariappan
Dept. of Information Technology
Easwari Engineering College
Chennai, India

ABSTRACT

Number of people who use internet and websites for various purposes is increasing at an astonishing rate. More and more people rely on online sites for purchasing songs, apparels, books, rented movies etc. The competition between the online sites forced the web site owners to provide personalized services to their customers. So the recommender systems came into existence. Recommender systems are active information filtering systems that attempt to present to the user, information items in which the user is interested in. The websites implement recommender system feature using collaborative filtering, content based or hybrid approaches. The recommender systems also suffer from issues like cold start, sparsity and over specialization. Cold start problem is that the recommenders cannot draw inferences for users or items for which it does not have sufficient information. This paper attempts to propose a solution to the cold start problem by combining association rules and clustering technique. Comparison is done between the performance of the recommender system when association rule technique is used and the performance when association rule and clustering is combined. The experiments with the implemented system proved that accuracy can be improved when association rules and clustering is combined. An accuracy improvement of 36% was achieved by using the combination technique over the association rule technique.

General Terms

Data mining, recommender systems

Keywords

cold start, association rule, clustering, taxonomy, user profile.

1. INTRODUCTION

Recommender systems are now an integral part of online sites. They are very useful in recommending items or products to user according to their interests. The origin of recommenders can be traced back to methods like cognitive science, approximation theory, information retrieval and management science. The benefits of online sites having a recommender system in place are cross-selling, personalization, keeping the customers informed and customer retention. Some of the websites that use recommenders are Amazon, MovieLens, eBay, CDNow, MovieFinder. In collaborative filtering approach, the system recommends new items to the user by analyzing items purchased by similar users (Amazon.com). In content based

approach, they recommend items with similar contents to the items preferred by the target users (PandoraRadio). In hybrid approaches, both the content based and collaborative approaches are used to provide recommendations (Netflix). These approaches provide the customers with a number of recommendations.

Cold start problem (new user, new item) is one of the major issues which hinder the performance of recommenders. In the case of a new user, the number of ratings will be very less. This implies that the user profiles (consist of ratings given to the items) will be very short. The new user will be given non personalized recommendations till an adequate number of ratings are collected for the user. For a new item updated in the system, initially there will be no ratings. The possibility that this item will be recommended to the user is minimal. These problems should be addressed because the initial recommendations given to a new user plays an important role in deciding the user satisfaction and retention. Only if good quality recommendation is given, the users will come back to the site.

Various methods exist for addressing the cold start problem. Some of these solutions are based on association rules, clustering, classification etc. Many hybrid recommenders also exist for solving this issue. In this paper, association rules and clustering technique is used for solving cold start. Association rules are used to create and expand the user profile so that it will contain more number of ratings/domains of interest so as to solve new user problem. The clustering technique is used to group items and make prediction for item to solve new item problem. Experiments have been done to show that combining two techniques such as association rule and clustering gives better recommendations than using a single technique such as association rule only.

The paper is organized as follows. Section 2 discusses related works. The architecture and the concepts of the proposed system for solving the cold-start problem are presented in Section 3. Section 4 outlines the implementation of the system. Section 5 outlines the experimental setup. Section 6 discusses the results. Finally, Section 7 concludes the paper.

2. RELATED WORKS

In this section review of some of the works related to the proposed approach is done.

Much work has been done in the area of recommender systems. Qing Li and Byeong Man Kim explain how the clustering techniques can be applied to the item-based collaborative filtering framework to solve the cold start problem [19]. The work done by Gavin Shaw, YueXu and ShlomoGeva explains how to expand a user profile from a dataset with the help of association rules [20]. The collaborative filtering, content based and hybrid approaches and the issues in recommender systems are clearly explained in the survey done by Adomavicius and G. Tuzhilin [1]. Schein and Popescul et al... , proposed a hybrid recommender system, the aspect model, to recommend items that is not yet recommended [2]. The relationship between ontology's and recommender system and how to exploit this synergy to solve cold start problem is given by Middleton, Alani et al... [3]. Ziegler, C.N. Lausen and G. Schmidt proposes a method in which the taxonomic background knowledge is used for computing personalized recommendations in particular domain [4]. In [5] Leung, et al... Chung implements a hybrid recommendation algorithm which makes use of Cross-Level Association Rules (CLARE) to integrate content information about domain items into collaborative filters. Pasquier, N. Taouil, et al... explains how to remove the redundant association rules without reducing the information in [6]. In [7], [8], [13] by Shaw, Xu and Geva they implemented a method which allows the removal of hierarchically redundant approximate basis rules from multi-level datasets through the use of the dataset's hierarchy or taxonomy. The new algorithmic elements that increase the accuracy of collaborative filtering is discussed by Herlocker, Konstan et al... in [9]. Q. Li and B. Kim described a new filtering approach that combines the content-based filter and collaborative filter to achieve a good performance [10]. The basics of recommenders and the e-commerce sites which use the recommenders are described in [11] by Schafer, Konstan, and Riedl. Al Mamunur Rashid et al... proposed an online simulation framework to address cold start problem in their paper [12]. Another interesting work done by Sunita B Aher, and Lobo proposed various combinations of algorithms in recommending the courses to students in E-learning [14]. A combination of clustering and association rules was used to improve recommendation for digital library in the study conducted by HuiLia and XinyueLiub [15]. In [16] Herlocker, Konstan et al... explain the key decisions in evaluating the collaborative filtering recommender system. [17], by AmitSinghal give a brief overview of the key advances in information retrieval field. The basic concepts of collaborative filtering and its limitations and a clustering based algorithm for a large dataset is explained by Badrul M. Sarwar et al... in [18]. The study of how to evaluate the recommendation accuracy for top-k recommendations is outlined by HaraldSteck in [21].

3. ARCHITECTURE OF THE PROPOSED SYSTEM

In this section an outline of the proposed approach for solving the cold-start problem in recommender systems is drawn.

The approach is to combine two existing approaches in a sequential manner. First, association rule technique is applied to expand the user profile as suggested by Gavin Shaw, YueXu and ShlomoGeva. With the help of this expanded user profile, clustering technique is applied for recommendation, focusing on new item recommendation, with reference to the work done by Qing Li and Byeong Man Kim.

The functional architecture of the proposed work is as shown in Fig 1.

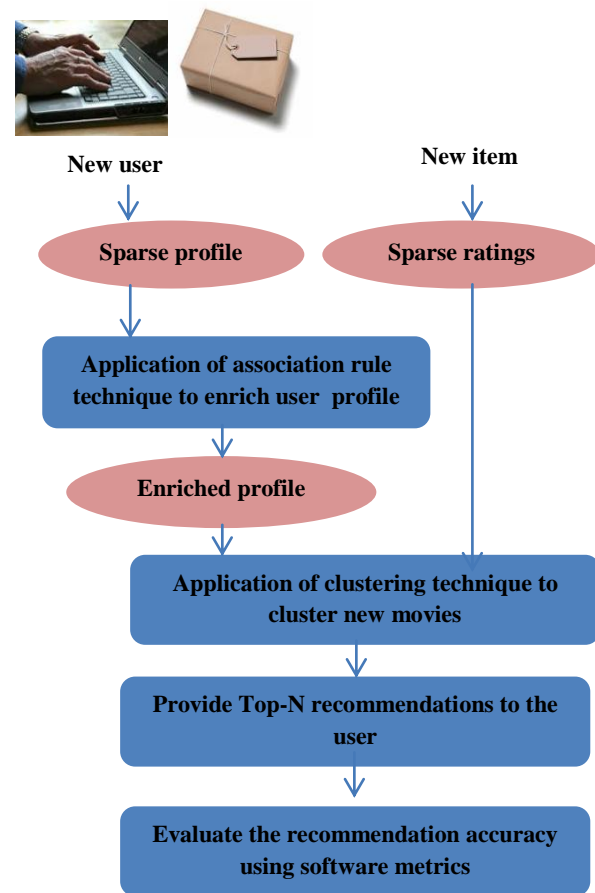


Fig 1: Functional architecture of the proposed recommender system

To implement the proposed system, three-tier architecture is used. Three-tier is a client-server architecture in which the user interface, functional process logic, computer data storage and data access are developed and maintained as independent modules, most often on separate platforms.

In the proposed system, the following tiers are defined.

- 1) *Client*: - This is the top most layer and defines the user interface module. The client tier is the user's window to the recommender application.
- 2) *Recommender System (Rec system)*: - This is the middle tier of the application which implements the logic of the recommender system. The two components of the recommender system are.
 - a) *User Profiler*: - which generates user profile from the information provided by the user and stores it in recommender database.
 - b) *Recommendation generator*: - which generates recommendations for the user. It uses the data from the recommender database.
- 3) *Data source*: - This is the bottom tier of the system which contains the data sources of the recommender

system. The file storage system is used to store the dataset files in .arff and .dat format.

The system architecture of the proposed work is outlined in Fig 2.

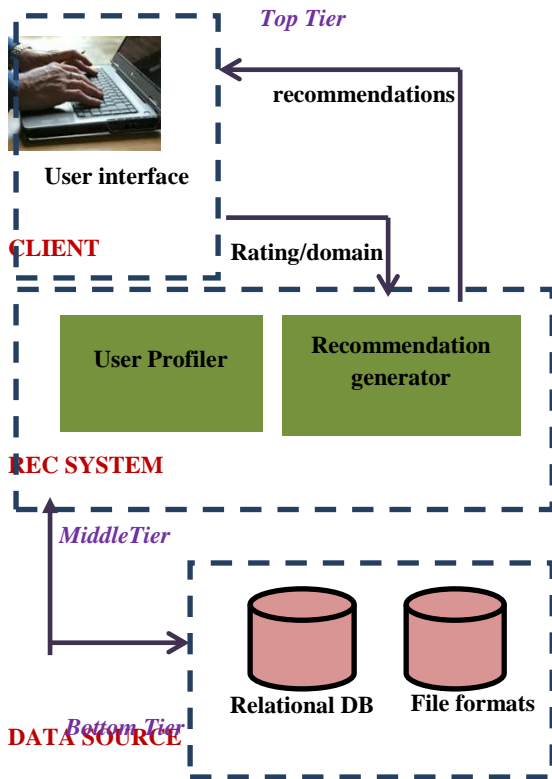


Fig 2: System architecture of the proposed recommender system

Now let us go through the mentioned techniques in some detail.

3.1 Association rule technique

Recommendation system shows best behavior when the user profiles are extensive and dataset has high information density. Expanding user profile means it contains more ratings. From the existing taxonomy driven user profile (P) construct a transactional dataset [20]. A transaction refers to the topics that the user is interested in. Using this transactional dataset, mine the frequent patterns. The frequent patterns are mined using Apriori algorithm. Association rules between topics that interest users can be derived from the patterns. These rules allow us to discover those topics that frequently appear together. Now consider this rule set and the user profiles (P). For each user profile $p(u_x)$, extract all topics (t) within and list of all combination possible from group of topics is generated. Each combination represents a possible antecedent of an association rule. Take each combination and search the set of association rules for any rules that have a matching antecedent. If matching rule exists, take topics in its consequent and add them to the profile. An example of this process is shown in Fig 3 and Fig 4 below. This approach thus helps us in resolving the new user problem to an extent by adding some more domains/genre of interest to user profiles.

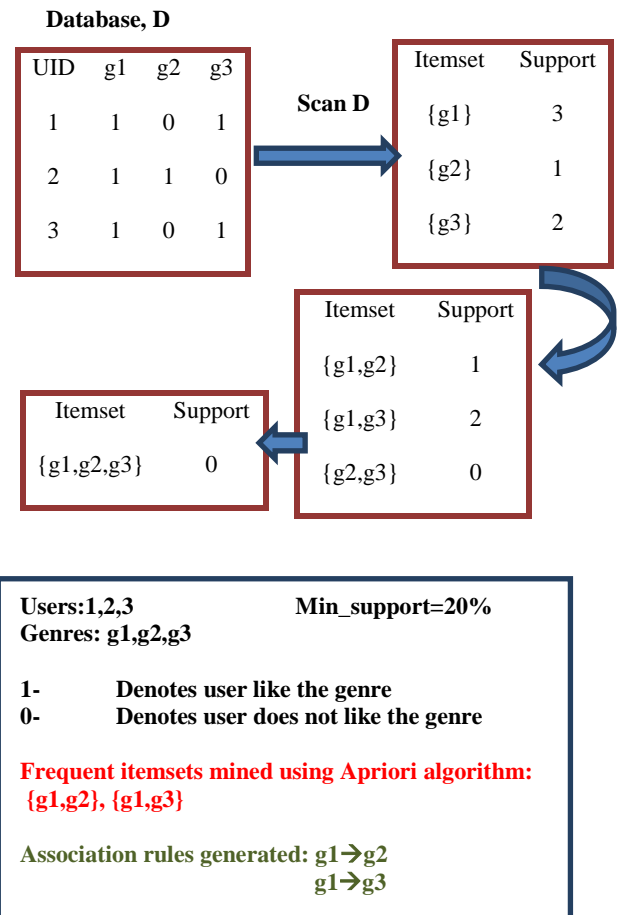


Fig 3: Generation of association rules using Apriori algorithm

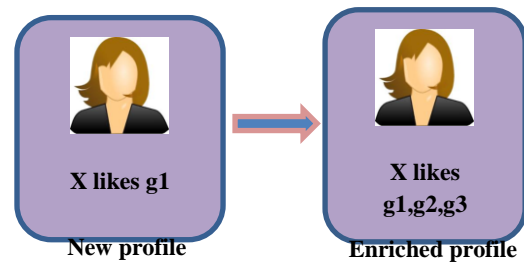


Fig 4: Profile enrichment using association rules

3.2 Clustering technique

Clustering technique is one of the most important techniques which have got a wide variety of applications. In this approach clustering algorithm is applied to group items as explained in [19]. Here in this work, k-means algorithm is used to cluster the new movie which is added to the recommender system. For the similarity calculation between two movies, Jaccard similarity is used since focus is on the non-numeric attribute of the movie, genre/domain.

$$\text{Jaccard similarity, } S_{i,j} = \frac{p}{p + q + r(1)}$$

where p = number of variables that are positive for both movies i and j , q = number of variables positive (1) for i^{th} movie and negative (0) for j^{th} movie and r = number of variables negative (0) for i^{th} movie and positive (1) for j^{th}

movie. After assigning the new movie to a cluster, predictions for the new movie is calculated by taking the average of the ratings of the other movies in the cluster. The example for the clustering process is shown in Fig5.

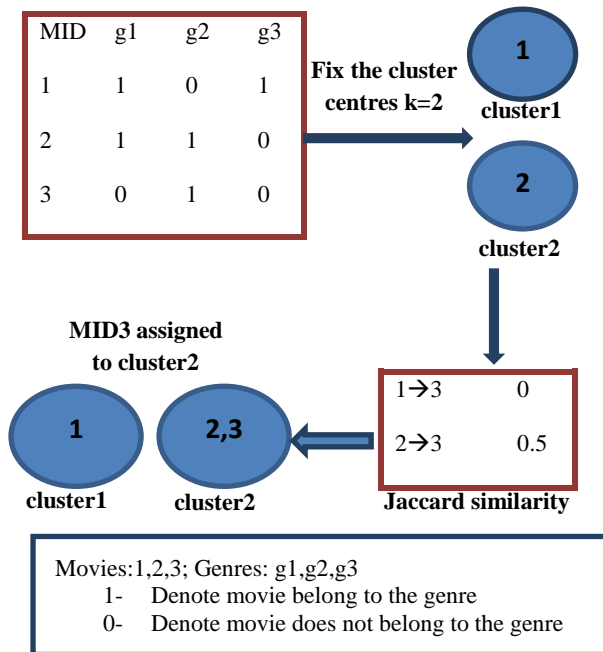


Fig 5: Clustering a new movie added to the system

3.3 Combination of association rule and clustering technique

The above techniques are combined in a sequential manner in the proposed work. For a new user, expansion of the user profiles is done using association rule technique. The result is the enriched user profile which now contains more domains/genres of interest. Now recommendation of a wide number of movies can be provided which come under these domains even if the user is new to the system. Clustering technique is applied to the new items so as to assign them to a cluster. Predictions for the new movie are calculated based on the average of the user ratings for the movies in the same cluster to which the new movie is added. These new movies can be randomly recommended to the user. The user is provided with Top-N recommendations which consist of both the new movies and the already rated movies.

4. IMPLEMENTATION

Our methodology for the implementation of this proposal is as follows. The entire work is divided into six stages.

4.1 Data collection and preprocessing

Dataset for the experiment is the MovieLens dataset collected from the GroupLens site. The dataset size is 10M (users: 2113, movies: 10197, genres: 20, ratings: 855598). This dataset files are in .dat format. The files are loaded into the MS Access database and then pre-processed using WEKA. The real-world data that is to be analysed by data mining techniques are incomplete, noisy and inconsistent. The Weka generates .arff file after pre-processing the data. These .arff files are loaded into the database.

4.2 Creation of taxonomy based user profiles

Now user profiles have to be created based on the information obtained from the users. The information is about the movie domains/genres in which the user is interested in. Taxonomy tree for movie is created manually. The taxonomy based user profile is created as explained in [4]. Using this taxonomy driven user profile (P) we can construct a transactional dataset and mine the frequent patterns. Here the frequent patterns imply the domains/genres which appear together.

4.3 Application of association rule technique to enrich the user profiles

From the taxonomy based user profiles, construct a transactional dataset and mine the frequent patterns using Apriori algorithm. The association rules between domains/genres that interest users can be derived from the patterns and these rules allow us to discover the domains/genres that frequently appear together. List of all combination possible from group of domains/genres is generated. Each combination represents a possible antecedent of an association rule. For each combination, search the set of association rules for any rules that have a matching antecedent. If a matching rule is found, take domains/genres in its consequent and add them to the profile. Now the user profile will have more number of domains/genres of user interests. So the movies which come under these genres can be recommended to the users even if he is a new user.

4.4 Application of clustering technique to improve item recommendation

Now apply the clustering technique as explained above when a new movie is added to the recommender system. The k-means algorithm is used to assign the new movie to a cluster. Since the movie attribute, genre/domain, is non-numeric we use Jaccard similarity to find the similarity between two movies. Based on this similarity calculation the new movie is assigned to a cluster to which it is most similar. The prediction for the new movie is the average of the user ratings of the movies in the cluster to which the new movie is assigned.

$$P_{u,k} = \sum_{i=1}^n R_{i,k}(2)$$

where $P_{u,k}$ is the prediction for the user u on item k, n is the total number of neighbors of user u in the cluster and $R_{i,k}$ is the rating of user I on item k.

Finally, a random number of new movies can be selected and recommended for the users. From the available recommendations, Top-N are taken and given to the users.

4.5 Project the Top-N recommendations to the user

Recommendations generated by the recommender application are to be displayed to the user. The recommendations are displayed through the user interface of the application. In the implemented recommender system, the movies which are not rated by him but come under his domain/genre of interests are given as recommendations. Also randomly selected new movies are also recommended. Thus a total of 10 movies (out of which 5 are new movies) are recommended to the user.

4.6 Evaluating the recommendation quality using metrics

The quality of the recommendations can be checked using the software evaluation metrics such as precision, recall and F1 measure [16]. Here the evaluation of the recommendation accuracy is calculated using the precision metrics. Since only the Top-k recommendations are displayed, the precision of the recommendations can be calculated as

$$\text{Precision} = N_u(k)/k(3)$$

where $N_u(k)$ is the number of items that are relevant to user u and which made it to top-k items based on recommender system predictions [21].

Relevancy is a term associated with the human perception. A movie which is relevant to one user may not be relevant to another user. So in the proposed system, experiments are done with the recommended movies with predictions >3 , to be considered as relevant. The precision is calculated for the recommendations generated by a) Association rule only b) Combination of association rule and clustering

The work is based on the suggestion that the use of taxonomy driven profile will improve the recommendations since the system will be able to cover more topics in which the user is interested in. The enriched profile thus generated will be an added advantage in clustering the new items and produce quality recommendations for the users. So the experiment proves that the proposed approach will be better than applying the association rule and clustering techniques separately.

5. EXPERIMENTAL SETUP

The ground work for the implementation is the data collection and pre-processing. The implemented system uses the MovieLens dataset collected from the GroupLens site. The dataset size is 10M (users: 2113, movies: 10197, genres: 20, ratings: 855598). This dataset files are in .dat format. The files are loaded into the MS Access database and then pre-processed using WEKA. The real-world data that is to be analyzed by data mining techniques are incomplete, noisy and inconsistent. Pre-processing is needed to enhance the mining process and improve data quality. The Weka generates .arff file after pre-processing the data. These .arff files are loaded into the database. For experimental feasibility, only a subset of the dataset is considered because of the computational constraints.

The next step is the creation of taxonomy based user profiles. Since the dataset is based on movies, a manual taxonomy tree has been developed. The taxonomy tree was developed based on the genre information defined in the dataset.

Apriori algorithm is implemented so as to generate the frequent itemsets. Only those profiles with less than four domains/genres of interests are considered to be cold profiles and thus selected for the enrichment process. From the generated frequent itemsets, association rules are formed. For the enrichment of user profiles, only five association rules are taken into consideration.

A very large number of association rules may be generated but it is not feasible to experiment with all the rules. Clustering algorithm is implemented by fixing the cluster centers, $k=5$. For the ideal clustering of the movies the algorithm is iterated for two times. Then recommendations are displayed to the user which is the Top-10 recommendations, out of which 5 are new item/movie recommendations.

Finally, the accuracy of the recommendations generated is calculated for the two cases

- 1) *Association rule only*: In this case the precision of the recommendations is given by

$$\text{Precision} = (\text{no of relevant movies in top} - 5)/5 \quad (4)$$

- 2) *Combination of association rule and clustering*: In this case the precision of the recommendations is given by

$$\text{Precision} = (\text{no of relevant movies in top} - 10)/10 \quad (5)$$

6. RESULT ANALYSIS

On analysing the recommender system performance, the following results were obtained.

6.1 Comparison of recommendations generated by association rule only and recommendations generated by association rule and clustering.

In this work, the assumption was that out of the movies given as recommendations only those which are having a prediction rating of more than 3 are relevant to the user. The precision values were obtained for 100 number of users. The accuracy calculation screen of the implemented system is shown in Fig 6. It contains the number of users who used the recommender service, the number of users registered in the system and the precision values for both techniques.

Also the history of the precision values with number of users for both the techniques is displayed in a table as shown in Fig 7. From the history table we plotted the graph with precision versus number of users (for both techniques). The graph is shown in Fig 8.

In the graph, the red line shows the precision of the system when association rule and clustering is applied and the green line shows the precision of the system when association rule alone is applied. The result clearly proves that the combination technique is better than applying association rule alone.

An accuracy improvement of 36% was obtained on using the combination technique over association rule technique.

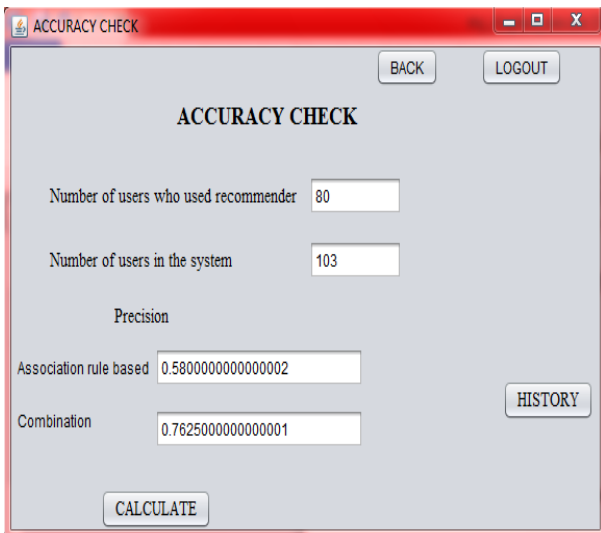


Fig 6: Snapshot of accuracy calculation screen

No of users	Asso Precision	Combo Precision
9	0.19999999999999998	0.4
10	0.19999999999999998	0.4090909090909091
11	0.29411764705882354	0.5117647058823529
17	0.30526315789473685	0.5210526315789473
19	0.5263157894736841	0.6710526315789472
20	0.5428571428571428	0.6666666666666666
21	0.54399999999999999	0.68399999999999999
25	0.54399999999999999	0.68399999999999999
30	0.52666666666666666	0.69
35	0.5485714285714286	0.7114285714285713
36	0.56111111111111112	0.71944444444444443
37	0.5729729729729731	0.7270270270270269
39	0.5743589743589745	0.7307692307692307
40	0.58500000000000002	0.73749999999999999
43	0.5720930232558141	0.7348837209302325
45	0.5777777777777781	0.74
46	0.5739130434782611	0.7391304347826085
48	0.57083333333333336	0.7395833333333333
49	0.5673469387755105	0.7387755102040813
50	0.56400000000000004	0.73799999999999995
52	0.5538461538461541	0.7346153846153842

Fig 7: Snapshot of precision history table

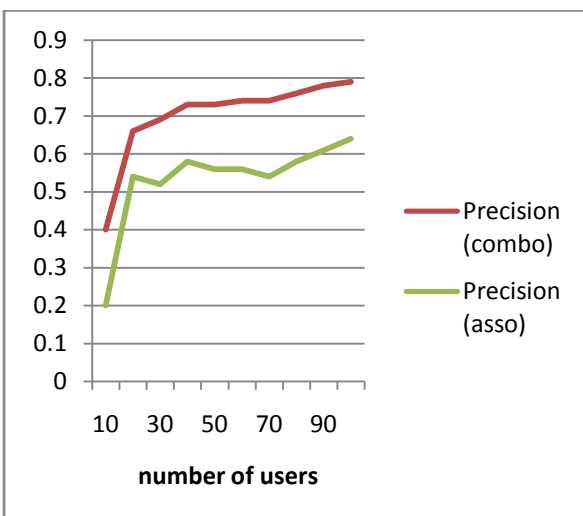


Fig 8: Comparison of precision for association rule - clustering combo versus association rule only

6.2 Comparison with existing systems

The base paper [20] is about using association rule technique to improve recommendation accuracy. In the proposed approach [22], the idea was to add the clustering technique to this existing method to experiment with the recommendation accuracy. In their work Gavin Shaw, YueXu and ShlomoGeva [20] applied association rule technique for recommendation. They used Bookcrossing dataset for their experiments and achieved an improvement of 31.5% over baseline (without using association rule technique) in terms of recommendation accuracy. In this implemented system, experiments were done to generate recommendations using association rule technique and also recommendations were generated using by combining association rule and clustering technique. MovieLens dataset was used for the experiment. The results showed an improvement of 36% when combination of association rule and clustering is used rather than using association rule only (baseline). The comparison of recommendation accuracy improvement in both the methods is shown in Fig 9.

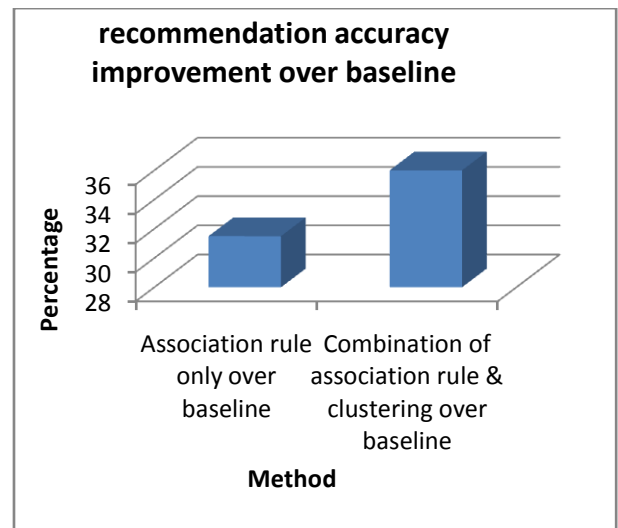


Fig 9: Comparison of recommendation accuracy improvement

7. CONCLUSION

In this paper, a method to solve the cold start problem in recommender system is proposed. The system is implemented and experimentation is done with the data available. The dataset used is the MovieLens dataset collected from the GroupLens site. The dataset size is 10M (users: 2113, movies: 10197, genres: 20, ratings: 855598). For the experiment, only a subset of the dataset is used because of the computational constraints. Using the results a comparative study of the quality of recommendations when association rule technique is applied and when combination of association rule and clustering technique is applied. From the experimental study of the implemented recommender system, it was found that quality recommendations can be improved if combination of association rule and clustering technique is used rather than applying association rule alone. An accuracy improvement of 36% was achieved by using the combination technique over the association rule technique. This work can be enhanced in future by experimenting with a combination of association rule, clustering and classification combination to find out whether recommendation accuracy can be improved further.

8. ACKNOWLEDGEMENT

We are very grateful to all people who directly or indirectly helped us in the discussions related to this work. We are also thankful to the reviewers who gave valuable suggestions towards the improvement of this paper.

9. REFERENCES

- [1] Adomavicius, G., Tuzhilin, A. 2005. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering* 17, 734–749.
- [2] Schein, A.I., Popescul, A., Ungar, L.H., Pennock, M. 2002. Methods and Metrics for Cold-Start Recommendations. In: 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'02). Pp.253–260. Tampere, Finland.
- [3] Middleton, S.E., Alani, H., Shadbolt, N.R., Roure, and D.C.D. 2002. Exploiting Synergy between Ontologies and Recommender Systems. In: The Semantic Web Workshop, World Wide Web Conference (WWW'02). pp. 41–50. Hawaii, USA.
- [4] Ziegler, C.N., Lausen, G., Schmidt-Thieme, L. 2004. Taxonomy-driven Computation of Product Recommendations. In: International Conference on Information and Knowledge Management (CIKM'04). pp. 406–415. Washington D.C., USA.
- [5] Leung, C.W., Chan, S.C., Chung, F. 2007. Applying Cross-level Association Rule Mining to Cold-Start Recommendations. In: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Workshops. pp. 133–136. Silicon Valley, California, USA. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.
- [6] Pasquier, N., Taouil, R., Bastide, Y., Stumme, G. 2005. Generating a Condensed Representation for Association Rules. *Journal of Intelligent Information Systems* 24, 29–60.
- [7] Shaw, G., Xu, Y., Geva, S. 2008. Eliminating Association Rules in Multi-level Datasets. In: In 4th International Conference on Data Mining (DMIN'08). pp. 313–319. Las Vegas, USA.
- [8] Shaw, G., Xu, Y., Geva, S. 2008. Extracting Non-Redundant Approximate Rules from Multi-Level Datasets. In: In 20th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'08). pp. 333–340. Dayton, Ohio, USA.
- [9] J. Herlocker, J. Konstan, B. A., and J. Riedl. 1999. An algorithmic framework for performing collaborative filtering. In *Proc. ACM-SIGIR Conf.*, pages 230–237.
- [10] Q. Li and B. Kim. 2003. An approach for combining content-based and collaborative filters. In *Proc. of IRAL2003*.
- [11] J. B. Schafer, J. Konstan, and J. Riedl. 1999. Recommender systems in e-commerce. In *Proc. 1st ACM Conf. on Electronic Commerce (EC'99)*.
- [12] Al Mamunur Rashid, George Karypis, and John Riedl. Learning Preferences of New Users in Recommender Systems: An Information Theoretic Approach (Department of Computer Science & Engineering, University of Minnesota, Minneapolis, MN-55455)
- [13] Yue Xu, Gavin Shaw, Yue feng Li. 2009. Concise Representations for Association Rules in Multi-level datasets (ISSN: 1004-3756 (Paper) 1861- 9576 (Online) DOI: 10.1007/s11518-009-5098-x CN11-2983/N, © Systems Engineering Society of China & Springer-Verlag 2009)
- [14] Sunita B Aher, Lobo L.M.R.J. 2012. Best Combination of Machine Learning Algorithms for Course Recommendation System in E-learning (International Journal Of Computer Applications (0975 – 8887) Volume 41– No.6.
- [15] Hui Lia, Xinyue Liub. 2012. A Personalized Recommendation System Combining User Clustering and Association Rules with Multiple Minimum Supports in 2nd International Conference on Future Computers in Education Lecture Notes in Information Technology, Vols.23-24.
- [16] Jonathan L. Herlocker, Joseph A Konstan, Loren G Terveen and John T Riedl. 2004. Evaluating Collaborative Filtering Recommender Systems in ACM 1046-8188/04/0100-0005 \$5.00 ACM Transactions on Information Systems, Vol. 22, No. 1, Pages 5–53.
- [17] Amit Singhal. 2001. Modern Information Retrieval: A Brief Overview (Copyright 2001 IEEE. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering)
- [18] Badrul M. Sarwar, George Karypis, Joseph Konstan, John Riedl. Recommender Systems for Large-scale E-Commerce: Scalable Neighborhood Formation Using Clustering. (GroupLens Research Group / Army HPC Research Center)
- [19] Qing Li, Byeong Man Kim. 2003. Clustering Approach for Hybrid Recommender System, Proceedings of the IEEE/WIC International Conference on Web Intelligence (WI'03) © 2003 IEEE
- [20] Shaw, Gavin, Xu, Yue, & Geva, Shlomo. 2010. Using Association Rules to Solve the Cold-Start Problem in Recommender Systems, QUT Digital Repository: <http://eprints.qut.edu.au/40176>, © Copyright 2010 Springer
- [21] Harald Steck. 2010. Training and Testing of Recommender Systems on Data missing not at random, KDD '10 - July 2010, Copyright 2010 ACM 978-1-4503-0055-1/10/07.
- [22] Sobhanam Hridya, Mariappan, A.K. 2013. Addressing cold start problem in recommender systems using association rules and clustering technique. International Conference on Computer Communication and Informatics (ICCCI- 2013), Coimbatore, India. Print ISBN: 978-1-4673-2906-4.
- [23] J. Han and M. Kamber. 2000. Data mining: Concepts and Techniques. Morgan-Kaufman, New York.