International Journal of Advanced Computer Research (ISSN (print):2249-7277 ISSN (online):2277-7970)
Volume-3 Number-2 Issue-10 June-2013

Reduction of Data Sparsity in Collaborative Filtering based on Fuzzy Inference Rules

Atisha Sachan[1], Vineet Richhariya[2]

Abstract

Collaborative filtering Recommender system plays a very demanding and significance role in this era of internet information and of course e commerce age. Collaborative filtering predicts user preferences from past user behaviour or user-item relationships. Though it has many advantages it also has some limitations such as sparsity, scalability, accuracy, cold start problem etc. In this paper we proposed a method that helps in reducing sparsity to enhance recommendation accuracy. We developed fuzzy inference rules which is easily to implement and also gives better result. A comparison experiment is also performing with two previous methods, Traditional Collaborative Filtering (TCF) and Hybrid User Model Technique (HUMCF).

Keywords

Collaborative Filtering, Sparsity, Accuracy, Fuzzy Inference Rule, MovieLens.

1. Introduction

Enormous expansion of internet services such as e commerce sites results to recommender systems. Recommender Systems have been developed to automate the recommendation process. Recommendation systems help users find and select items (e.g., books, movies, restaurants) from the huge number available on the web or in other electronic information sources . Only a large database is given which is related to item and also description of users' preferences and they provide a selected small set of items list which they want or prefer [1]. There are four types of filtering technology which is used in Recommender System-demographic[2], content[3], collaborative[4][5][6] and combinational filtering[7][8][9].Recommender systems are an important part of the information and in e-commerce. They represent a powerful technique for enabling users to filter through large information from the database.

Atisha Sachan, M TECH(CSE), LNCT, Bhopal, India.
Vineet Richhariya, HOD(CSE), LNCT, Bhopal, India.

Nearly two decades of research on collaborative filtering have led to a varied set of algorithms and a rich collection of tools for evaluating their performance. Research in this field is increasingly moving in the direction of a richer understanding to see how recommender technology may be embedded in specific domains[10].

Collaborative filtering is most widely used method learns preferences of user having same interests with the same items. It is of two types, memory based and model based. Filtering based on collaborative aiming on collecting and analyzing a large amount of information on users' behaviour, their activity or like or dislike about an item and predicting what user will like based on similarity to other users. Despite its widespread adoption, collaborative filtering suffers from various major limitations including sparsity, scalability, cold start, accuracy etc [11].

In this paper we focus on the sparsity problem. The sparsity problem originates when available data are insufficient for identifying similar users or we can say neighbours and it is a major issue since it limits the quality of recommendations and the applicability of collaborative filtering in general. The main objective of our work is to develop an effective approach that provides high quality recommendations even when sufficient data are unavailable [12]. We use fuzzy inference rule to reduce the sparsity and k nearest neighbour algorithm to measure the similarity between users.

The paper is arranged as follows. After introduction we present a literature review on collaborative filtering and the sparsity problem in detail. Another section introduced the proposed method for reducing sparsity in collaborative filtering recommender system. Then we present experimental results and last section concludes summary and future work.

2. Collaborative Filtering and Sparsity Problem

Collaborative filtering techniques have been successful in enabling the prediction of user preferences in the recommendation systems (Hill et al., 1995, Shardanand & Maes, 1995). There are three major processes in the recommendation systems: data item collections and representations, similarity measure, and recommendation computations. Collaborative filtering finds the relationships among the new individual and the existing data in order to further determine the similarity and provide recommendations. For example, people that like and dislike movies in the same categories would be considered as the ones with similar behaviour (Chee et al., 2001). The concept of the nearest-neighbour algorithm has been included in the implementation of the recommendation systems (Resnick et al., 1994). The designs of pioneer recommendation systems focus on entertainment fields (Resnick et al., 1994, Hill et al., 1995, Shardanand & Maes, 1995, Dahlen et al., 1998). The challenge of conventional collaborative filtering algorithms is the scalability issue (Sarwar et al., 2000a). Conventional algorithms explore the relationships among system users in large datasets.

There are some potential problems with the Collaborative Filtering RS. First one is the scalability, which defines as how quickly a recommender system can generate recommendation; second one is sparsity third one cold start Problem. These problems became a big hurdle to achieve better accuracy. We discuss all of them below [13]:

Scalability

In this e commerce or internet environment, there are millions of users and products. Thus a large amount of computation power is often required to calculate recommendation. For example, with tens of millions of customers (*C*) and millions of distinct catalogue items (*I*), a *CF* algorithm with the complexity of $O(n)$ is already too large. As well as, many systems need to react immediately to online requirements and to make quick recommendations for all users regardless of their purchasing and ratings history that demands a high scalability of a *CF* system [14].

Data Sparsity

The number of items sold or buy on major e commerce sites is extremely large in quantity. Even the most active users will only have rated a small subset of the overall database which results that even most popular items have very few ratings.

Cold Start Problem

The cold start problem mainly occurs when a new user or item has been added into the database system, it is difficult to find similar users because there is not enough information (sometimes the cold start problem is also known as the new user problem or new item problem [15, 16]). New items cannot be recommended until some users' rate it, and so because of the lack of their rating or purchase history new users are unlikely given good recommendations.

3. Literature Review

We study various recent research papers also. Some of the papers are discussed here about work in collaborative filtering and on sparsity issue.

A. By Guangping Zhuo, Jingyu Sun and Xueli Yu [17] **"A Framework for Multi-Type Recommendations"** deals in the field of web minning concern on some drawbacks in collaborative filtering and also on multi type recommendation. Collaborative filtering (CF) is an effective method of recommender systems (RS) has been widely used in online stores. It suffers some weaknesses: problems with new users (cold start), data sparseness, and difficulty in spotting "malicious" or "unreliable" users and so on. So in order to make it adaptive new Web applications, such as urban computing, visit schedule planning and so on, introduced a new recommendation framework, which combines CF and case-based reasoning (CBR) to improve performance of RS. Based on this framework, the authors have developed a semantic search demo system MyVisit, which shows that our proposed framework is an effective recommendation model. Two key algorithms, **MIFA** and **RAA**, are used. Additionally, authors have validated them using an application instance, which is a demo system for recommending multi type recommendations combining CF and CBR. Advantage of this method is that it involves a few of cases in the online and adjusts the rating of main items through associative other type items in order to find fit recommendations.

B. By Yechun Jiang, Jianxun Liu, Mingdong Tang and Yechun Jiang, Jianxun Liu, Mingdong Tang [18**] "An Effective Web Service Recommendation Method based on Personalized Collaborative Filtering"**. The paper describing an effective personalized collaborative filtering method for Web service recommendation. A key component of Web service recommendation techniques is computation of similarity measurement of Web services. Different from the Pearson Correlation Coefficient (PCC) similarity measurement, they take into account the personalized influence of services when computing similarity measurement between users and personalized influence of services. Based on the similarity measurement model of Web services, develop an effective Personalized Hybrid Collaborative Filtering (PHCF) technique by integrating personalized user-based algorithm and personalized item-based algorithm. Also conduct series of experiments based on real Web service QoS dataset WSRec [20] which contains more than 1.5 millions test results of 150 service users in different countries on 100 publicly available Web services located all over the world.

C. By Qian Wang, Xianhu Yuan, Min Sun [19**] "Collaborative Filtering Recommendation Algorithm based on Hybrid User Model".** The paper proposes a hybrid user model to remove some of its drawbacks. The recommender system based on this model not only holds the advantage of recommendation accuracy in memory-based method, but also has the scalability as good as model-based method. The user model is constructed based on item combination feature and demographic information, and it focuses on searching for set of neighbouring users shared with same interest, which helps to improve system scalability. To enhance recommendation accuracy, each feature in user model is given a different weight when computing the similarity between users. Genetic algorithm is adopted to learn the weight values of features. Methodology proposed improves recommendation accuracy and scalability to a certain extent. It constructs a concise and representative hybrid user model, and combines and integrates item ratings, item detailed description and demographic information together, which raises the density of data and improves the problem of sparse data. Besides, genetic algorithm is adopted to learn a best feature weight vector in computation of the nearest neighbour set, which helps to get a more accurate similarity.

D. By Chuangguang Huang and Jian Yin [20] **"Effective Association Clusters Filtering to Cold-Start Recommendations".** The paper focuses on cold-start problem. The popular technique of RS is Collaborative Filtering (CF). While in real online RS, CF can't practically solve cold-start problem for the sparsity ratings dataset. The paper proposed a novel efficiently association clusters filtering (ACF) algorithm. Considering hybrid approaches, using clustering and also filtering to relieve cold-start problem. ACF algorithm establishes clusters models based on the ratings matrix. We assume the users in the same cluster; they will have the same interests. On the other hand, different users in different clusters present they will have less common interests. The more users ratings for some item in the cluster, can delegate the opinion of the cluster.

E. By Mustansar Ali Ghazanfar and Adam Prugel-Bennett [21] **"A Scalable, Accurate Hybrid Recommender System".** The paper proposes a unique cascading hybrid recommendation approach by combining the rating, feature, and demographic information about items. They empirically show that their approach outperforms the state of the art recommender system algorithms, and eliminates recorded problems with recommender systems. Since there are three main types of recommender systems: collaborative filtering, content-based filtering, and demographic recommender systems. They combine all these filtering to form a hybrid recommender system.

F. By Liang He and Faqing Wu [22] **"A Time-context-based Collaborative Filtering Algorithm".** The paper incorporates the time-context, one of the most important contexts, into the traditional collaborative filtering algorithm and proposes a Time context-Based Collaborative Filtering (TBCF) Algorithm to improve the performance for traditional collaborative filtering algorithm. Experiments evaluating this approach are carried out on real dataset taken from movie recommendation system provided by MovieLens web site. The result shows the proposed approach can improve predication accuracy and recall ratio compared with existing methods. The time context is a very important factor in recommendation system. And the paper introduced time interval into the traditional user-based collaborative filtering algorithm. The strategies proposed improved both the prediction accuracy and recall ratio of standard user-based collaborative filtering methods.

G. By Ling Yun, Wang Xun and Gu Huamao [23**] "A Hybrid Information Filtering Algorithm Based on Distributed Web log Mining"**. For distributed large commercial mirror sites, the paper presents a hybrid information filtering algorithm based on distributed web log mining. Based on multiagent technology, the algorithm pre-processes the web logs of mirror sites, in which the web page's manual rating is replaced by user browsing preference, and then user access matrix is constructed and standardized. The paper proposes a distributed web log mining based hybrid filtering algorithm. To solve the problem that users are reluctant to rate web pages, this paper establishes the user access matrix on the basis of web log mining to gather fundamental data for both filtering. For the sparseness of user rating data of collaborative filtering, a collaborative filtering algorithm is proposed based on web page rating prediction, which effectively overcomes the drawbacks of traditional similarity measuring methods under circumstances of data sparseness and improves the accuracy of target user's calculation of the nearest neighbour. With the optimal weight, this model further improves the recommendation quality.

H. By Ibrahim A. Almosallam and Yi Shang [24] **"A New Adaptive Framework for Collaborative Filtering Prediction".** The paper focused on memory-based collaborative filtering (CF). Existing CF techniques work well on dense data but poorly on sparse data. To address this weakness, the paper proposed to use z-scores instead of explicit ratings and introduce a mechanism that adaptively combines global statistics with item-based values based on data density level. They present a new adaptive framework that encapsulates various CF algorithms and the relationships among them. An adaptive CF predictor is developed that can self adapt from user-based to item-based to hybrid methods based on the amount of available ratings. The experimental results show that the new predictor consistently obtained more accurate predictions than existing CF methods, with the most significant improvement on sparse data sets. When applied to the Netflix Challenge data set, our method performed better than existing CF and singular value decomposition (SVD) methods and achieved 4.67% improvement over Netflix's system.

I. By Cane Wing-ki Leung, Stephen Chi-fai Chan and Fu-lai Chung [25] **"Applying Cross-Level Association Rule Mining to Cold-Start Recommendations".** The paper proposed a novel hybrid recommendation algorithm for addressing the well-known cold-start problem in Collaborative Filtering (CF). The algorithm makes use of Cross-Level Association RulEs (CLARE) to integrate content information about domain items into collaborative filters. They first introduce a preference model comprising user-item and item-item relationships, and described the CLARE algorithm for generating cold-start recommendations. When no recommendations can be generated for an item from ratings data, CLARE takes into consideration the attributes of the item for generating cold-start recommendations. Experimental results validated the ability of CLARE to recommend cold-start items and to improve significantly the number of recommendable items in a system. They experimented with only one type of attribute (cast) for mining CARs as an initial effort. They studied the behaviour of CLARE using more attribute types with varying characteristics, and obtained improved recommendation quality and coverage.

J. By Leo Iaquinta, Anna Lisa Gentile, Pasquale Lops, Marco de Gemmis and Giovanni Semeraro [26] **"A Hybrid Content-Collaborative Recommender System integrated into an Electronic Performance Support System".** The paper proposed the adoption in an EPSS of a novel hybrid recommender that implements a neighbourhood formation process based on the idea of grouping users by computing similarities between their semantic user profiles instead of their rating style. This hybrid recommender overcomes Sparsity Problem and Lack of Transparency Problem.

K. By Manos Papagelis, Dimitris Plexousakis, Themistoklis Kutsuras **"Alleviating the Sparsity Problem of Collaborative Filtering Using Trust Inferences"**. In this research, author's main objective was to describe a method that is able to provide high-quality recommendations even when information available is insufficient. To deal with the sparsity problem they proposed a method that is based on trust inferences. Trust inferences are transitive associations between users that participate in the underlying social network. Employment of this model provides additional information to Collaborative Filtering algorithm and remarkably relaxes the sparsity and the cold-start problems. Furthermore, the model considers the subjective notion of trust and reflects the way in which it is raised in real world social networks. The methodology described is general and may probably be easily adopted to alleviate the sparsity problem in other application areas, especially where underlying social networks can be identified.

4. Proposed Methodology

Collaborative filtering play an important role in web based recommendation systems. The rating of product decided by user choice, user select a feature of product and categorised in two sections like and

dislike. They like and dislike categories creates a user rating matrix. For the user rating of the product categorises into three sections. Firstly, user like, secondly user dislike and thirdly common users' like and dislike both and finally measures the loss at attribute at feature due to sparsity. The rule of inference plays a central role at product rating and measures the correlation of recommendation matrix.
The decision rules are:

1. If u_k like the item i_k and the interest categories of u_k is involved in the interest categories of u_a and i_k belongs to same interest categories.
2. If u_k dislikes the item i_k and the interest of categories of u_a belongs to interest of u_a on i_k.
3. If u_k likes the item i_k and user u_a does not like that item, but the users' likes the i_x for common interest category.
4. If u_k and u_a doesn't like the product category i_k .
5. If u_k and u_a rated item i_k and i_x but i_k loss the feature rating matrix value.

The proposed formulae for prediction of user rating:

$$\mathbf{R}= \frac{\sum_{I_x \in I_{sa}(I_t)} w(I_x,I_t).r_x.ratio_{x,t}+\sum_{u_k \in u_{sa}(u_a)} w(U_x,U_a).r_{k,t}.ratio_{k,a}}{\sum_{I_x \in I_{sa}(I_t)} w(I_x,I_t) + \sum_{u_k \in u_{sa}(u_a)} w(I_x,I_t)}$$

Where,
$w(I_x,I_t)$ = Weight rating of product,
I_{sa} = Similarity of two items,
u_a, u_k = Two users,
r_x = Rank of x.

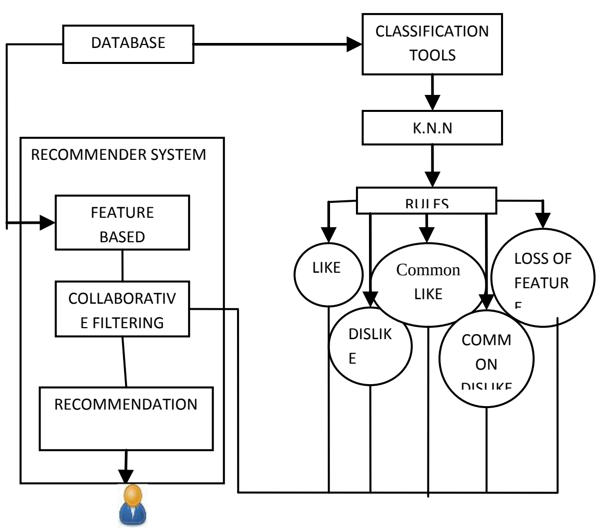

Figure 1: Model Diagram of Proposed Method

5. Experimental Results

Experimental Dataset

We take MovieLens Dataset as work experiment, which contains 100,000 rating records that 943 users graded for 1682 movies. MovieLens Dataset [27] was collected by GroupLens Research Project. The rating value ranges from 1 to 5.

Measure Standard

The most commonly used statistical accuracy metric is the Mean Absolute Error (MAE). Mean Absolute Error measures the deviation of predictions generated by the Recommender System from the true rating values, as they were specified by the user. The MAE is measured only for those items, for which user u_k has expressed his opinion. The Mean Absolute Error can be defined as:

$$MAE= \frac{\sum_{i=1}^{n} | x_i - y_i |}{n}$$

Here n represents the number of items assuming the predicted rating set of target user is(x_1, x_2, x_3 …x_n). And the real rating set is (y_1,y_2,…y_n). The smaller MAE value is, the higher accuracy recommendation is with.

Performance Analysis

We study and compare the new proposed algorithm to the traditional algorithms and hybrid user based algorithms and takes experiment as the number of nearest neighbours 5, 10, 15, 20, 30. The experimental results shown in below table and the graph shows the comparison graph between TCF (Traditional Collaborative Filtering), HUMCF (Hybrid User Model CF) and the FICF (Fuzzy Inference CF)

Table 1: Comparison Chart

No of Neighbours	Proposed FICF	HUMCF	TCF
5	3.4	4.24	7.22
10	3.4	4.30	7.27
15	3.5	4.37	7.32
20	3.54	4.43	7.37
25	3.58	4.5	7.42
30	3.63	4.55	7.47
40	3.73	4.68	7.56
50	3.83	4.80	7.66
60	3.93	4.93	7.76
70	4.02	5.06	7.85

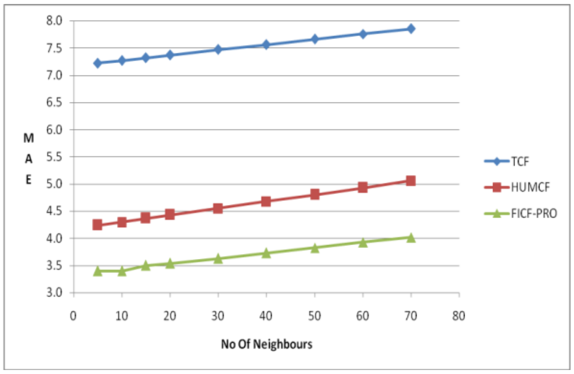

Figure 2: Comparison of Recommendation Accuracy

6. Conclusion and Future Work

Methodology proposed in the paper improves recommendation accuracy to a great extent. When user rating matrix is generated there is also some loss of feature value which results in data sparsity. To regain the values and to overcome this problem we developed 5 fuzzy inference rules to grouping the users according to rules so that filtering becomes easy i e easy to predict the user item preference. We also compare the method from two other methods, TCF and HUMCF. In future we enhance this method more to get more accurate results.

References

[1] Cynthia A. Thompson, Mehmet H. Goker, Pat Langley,"A Personalized System for Conversational Recommendations", Journal of Artificial Intelligence Research 21 (2004) 393-428.

[2] Krulwich, B. (1997),"Lifestyle Finder: Intelligent user profiling using large-scale demographic data", Artificial Intelligence Magazine, 18(2), 37–45.

[3] Lang, K. (1995). "NewsWeeder: Learning to filter netnews". In Proceedings of the 12th international conference on machine learning, Tahoe City, CA.

[4] Resnick, P., Iakovou, N., Sushak, M., Bergstrom, P., & Riedl, J. (1994). " GroupLens: An open architecture for collaborative filtering of Netnews". In Proceedings of ACM conference on computer supported cooperative work (pp. 175–186). NC: Chapel Hill.

[5] Shardanand, U., & Maes, P. (1995). "Social information filtering: Algorithms for automating 'Word of Mouth'". In Proceedings of the conference on human factors in computing systems (CHI'95), Denver(pp. 210–217).

[6] Boucher-Ryan, P. D., & Bridge, D. (2006). "Collaborative recommending using formal concept analysis". Knowledge-based Systems, 19, 309–315.

[7] Balabanovic,M., & Shoham, Y. (1997). "Fab: Content-based, collaborative recommendation". Communications of the ACM, 40(3), 66– 72.

[8] Shahabi, C., Banaei-Kashani, F., Chen, Y., & McLeod, D. (2001). "Yoda: An accurate and scalable web-based recommendation systems". In Proceedings of the sixth international conference on cooperative information systems (CoopIS 2001), Trento, Italy.

[9] Sutheera Puntheeranurak and Hidekazu Tsuji: "An Improved Hybrid Recommender System Using Multi-Based Clustering Method", IEEJ Trans. EIS, Vol. 129, No. 1, pp.125-132 (2009).

[10] By Michael D. Ekstrand, John T. Riedl and Joseph A. Konstan, "Collaborative Filtering Recommender Systems", Foundations and TrendsR_ in Human–Computer Interaction Vol. 4, No. 2 (2010) 81–173(2011).

[11] Sarwar, B., Karypis, G., Konstan, J., and Reidl, J. "Analysis of recommendation algorithms for e-commerce". In Proceedings of the ACM Conference on Electronic Commerce (2000), 158-167.

[12] Manos Papagelis, Dimitris Plexousakis and Themistoklis Kutsuras "Alleviating the Sparsity Problem of Collaborative Filtering Using Trust Inferences". iTrust'05 Proceeding of 3rd International on Trust Management. ACM DL(2005).

[13] Xiaoyuan Su and Taghi M. Khoshgoftaar "A Survey of Collaborative Filtering Techniques", Hindawi Publishing Corporation, Advances in Artificial Intelligence, Volume 2009.

[14] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," *IEEE Internet Computing*, vol. 7, no. 1, pp. 76–80, 2003.

[15] G. Adomavicius and A. Tuzhilin, "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*, vol. 17, no. 6, pp. 734–749, 2005.

[16] K. Yu, A. Schwaighofer, V. Tresp, X. Xu, and H.-P. Kriegel, "Probabilistic memory-based collaborative filtering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 1, pp. 56–69, 2004.

[17] Guangping Zhuo, Jingyu Sun and Xueli Yu "A Framework for Multi-Type Recommendations", Eighth International Conference on Fuzzy Systems and Knowledge Discovery, 2007.

[18] 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.
[19] Qian Wang, Xianhu Yuan, Min Sun "Collaborative Filtering Recommendation Algorithm based on Hybrid User Model", FSKD, 2010.
[20] Chuangguang Huang and Jian Yin "Effective Association Clusters Filtering to Cold-Start Recommendations", 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery.
[21] Mustansar Ali Ghazanfar and Adam Prugel-Bennett," A Scalable, Accurate Hybrid Recommender System", 2010 Third International Conference on Knowledge Discovery and Data Mining.
[22] He, Liang, and Faqing Wu. "A time-context-based collaborative filtering algorithm." In Granular Computing, 2009, GRC'09. IEEE International Conference on, pp. 209-213. IEEE, 2009.
[23] Ling Yun, Wang Xun and Gu Huamao "A Hybrid Information Filtering Algorithm Based on Distributed Web log Mining", Third 2008 International Conference on Convergence and Hybrid Information Technology.
[24] Guangping Zhuo, Jingyu Sun and Xueli Yu "A Framework for Multi-Type Recommendations", Eighth International Conference on Fuzzy Systems and Knowledge Discovery, 2007.
[25] Cane Wing-ki Leung, Stephen Chi-fai Chan and Fu-lai Chung "Applying Cross-Level Association Rule Mining to Cold-Start Recommendations", 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology.
[26] Leo Iaquinta, Anna Lisa Gentile, Pasquale Lops, Marco de Gemmis and Giovanni Semeraro "A Hybrid Content-Collaborative Recommender System Integrated into an Electronic Performance Support System", Seventh International Conference on Hybrid Intelligent Systems, 2007.
[27] MovieLens data, http://www.grouplens.org.

Atisha Sachan pursuing M Tech (Computer Science and Engineering) from Laxmi Narain College Of Technology, Bhopal. BE(IT) from GEC, Gwalior in 2008.