



An Approach for Travel Package Recommendation using Clustering and Typicality based Approach

Parnika Patil¹, Vaishali Kolhe²

PG Student, Dept. of Computer, DYPCOE Akurdi, Savitribai Phule Pune University, Pune, India¹

Assistant Professor, Dept. of Computer, DYPCOE Akurdi, Savitribai Phule Pune University, Pune, India²

ABSTRACT: Recommendation systems are used to predict the 'rating' or 'preference' that user would give to an item and are applied in a variety of applications like music, movies, news, research articles, books, social tags, search queries and products in general. This paper focuses on application in tourism. However traditional recommendation system suffers from problems such as data sparsity, recommendation accuracy. A new approach 'Clustering and Typicality based Collaborative Filtering' has been detailed out herewith this paper, which includes preprocessing methods i.e. clustering of items and measuring user typicality degree in user groups. After preprocessing the remaining recommendation process is done based on user typicality degree instead of corated items of user or common users of items, as in traditional CF. This method helps to reduce data sparsity and improve the accuracy of prediction.

KEYWORDS: Recommendation System, Travel Package, Typicality, Fuzzy C-means clustering.

I. INTRODUCTION

The Recommendation systems are software tools providing suggestions for items for a user. The suggestions provided are aimed at supporting their users in various decision making processes, such as where to plan a tour, what items to buy, for which season, what music to listen, or what news to read. Recommender systems have proven to be valuable means for online users to cope with the information overload and have become one of the most popular and powerful tools in electronic commerce field. Various techniques for recommendation system have been proposed like content-based, user-based and item based collaborative filtering and hybrid recommendation system. Many of them have also been deployed successfully in commercial environments. Added to it there are many evolutionary methods that could be incorporated to achieve better results in terms of handling various challenges of recommendation system like data sparsity, cold start problem, scalability and accuracy issues and accuracy in prediction.

Recommendation system is very useful for both customers and providers. For the customers it will help to narrow down the set of choices, explore the set of options, find the things that are more interesting to user and discovers the new things [2]. In case of provider, it will help to increase trust and customer loyalty, increase customer conversation, sales, click through rates. It will provide good opportunity for promotion and obtain more knowledge about the customer [7].

A distinct feature of the clustering with typicality-based CF recommendation is that it selects the "neighbors" of users by calculating typicality degrees in user groups, which differentiates it from previous methods. The mechanism of clustering with typicality-based CF recommendation is as follows: First, cluster all items into several item groups using fuzzy c means clustering method. Second, form a user group i.e., a set of users who like items of a particular item group, corresponding to each item group, with all users having different typicality degrees in each of the user groups. Third, we build a user-typicality matrix and measure users' similarities based on users' typicality degrees in all user groups so as to select a set of "neighbors" of each user. Then, we predict the unknown rating of a user on an item based on the ratings of the "neighbors" of at user on the item.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

The reminder of this review is structured as follows. The section II comments the approaches for recommendation system, focusing on content-based, collaborative and hybrid approaches and related work. The section III exposes the proposed clustering and typicality based approach in detail. In section IV we evaluate the proposed system using dataset. We conclude the paper in section V.

II. RELATED WORK

A. Approaches for Recommendation System :

Recommender systems have been generally classified, according to the way in which they analyze the information of the user and filter the list of items, into content-based, collaborative and hybrid systems [6][10]

- **Content based Recommendation System (CB) –**

Content-based systems calculate a degree of similarity between the items and the users to be recommended. The inspiration of this kind of recommendation methods comes from the fact that people have their subjective evaluations on some items in the past and in the future will have the similar evaluations on other similar items. This process is carried out by comparing the item features with respect to the user's preferences. So, it is assumed that both alternatives and users share a common representation (e.g., they are composed of the same set of keywords or attributes). The output of the comparison process is an overall performance score, which indicates the degree of matching between the user's profile and each alternative. The higher score indicates, higher performance of the alternative for a given user. Sometimes rating history of the user is considered. In this approach, the recommendation system relies on having an accurate knowledge of the user's preferences to be able to select the appropriate items [5].

- **Collaborative Filtering Recommendation System –**

Collaborative filtering systems make recommendations based on groups of users with similar preferences. The similarity between users is computed by comparing the ratings that they give to some of the items. When the system finds out who are the people that share similar interests with the active user, then the items that those people liked are recommended to this user. In this approach, some feedback about the provided recommendations is necessary, so as to know which items the user has liked or disliked (e.g. which places he has enjoyed visiting) [6].

- **Collaborative Filtering Recommendation System-**

Collaborative filtering systems make recommendations based on groups of users with similar preferences. The similarity between users is computed by comparing the ratings that they give to some of the items. When the system finds out who are the people that share similar interests with the active user, then the items that those people liked are recommended to this user. In this approach, some feedback about the provided recommendations is necessary, so as to know which items the user has liked or disliked (e.g. which places he has enjoyed visiting) [6].

For the reason that CF methods do not require well-structured item descriptions, they are often implemented than CB methods and many collaborative systems are developed in academia and industry. There are two types of CF approach namely —item based and user-based. The main idea of user-based CF approach is to give recommendation of an item for a user based on the opinions of other like-minded users on that item. The user-based CF approach initially finds out a set of nearest “neighbours” (similar users) for each user, who share similar interests or favourites. Finally, based on the ratings given by the user's “neighbours” on the item, the rating of a user on an unrated item is predicted. The item-based CF approach provides a user with the recommendation of an item based on the other items with high correlations. The item-based CF approach first finds out a set of nearest “neighbours” (similar items) for each item. The item based CF try to predict a user's rating on an item based on the ratings given by the user on the neighbours of the target item.

For both item-based CF and user-based CF, to find similarity of measurement between items or users is a significant step [1]. Pearson correlation coefficients, cosine-based similarity, vector space similarity, distance based similarity and so on are widely used as similarity measurement in CF methods.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

• Hybrid Recommendation System –

Some recommender systems use a hybrid approach by combining content based and collaborative methods, to avoid some limitations of content-based and collaborative systems. A hybrid approach initially implements CB and collaborative methods separately and then combines their predictions by a linear combination of ratings or a voting scheme or other metrics [12]. For hybrid recommender systems it is also possible to combine user-based CF and item-based CF.

Yi Cai et al [1] proposed Typicality-Based Collaborative Filtering Recommendation approach which finds “neighbors” of users based on user typicality degrees in user groups. It will minimize the data sparsity and improves accuracy than other traditional approaches. Also it will takes less amount of time.

Joan Borrs et al.[3] focus on recommender system application in tourism. The paper provides a detailed and up-to-date survey of the field, considering the different kinds of interfaces, the diversity of recommendation algorithms, the functionalities offered by these systems and their use of Artificial Intelligence techniques. The survey also provides some guidelines for the construction of tourism recommenders and outlines the most promising areas of work in the field for the next years.

Yi Cai et al.[4] proposed Recommendation Based On Object Typicality. Current recommendation methods are mainly classed into content based, collaborative filtering and hybrid methods. These methods are based on similarity measurements among items or users. In this paper, investigated recommendation systems from a new perspective based on object typicality and propose a novel typicality based recommendation approach. Experiments show that this method outperforms on recommendation quality.

Qi Liu et al.[2] proposed A Cocktail Approach for Travel Package Recommendation system. This paper provides a study of exploiting online travel information for personalized travel package recommendation. A critical challenge along this line is to address the unique characteristics of travel data, which distinguish travel packages from traditional items for recommendation. It will extend the TAST model to the tourist-relation-area-season topic (TRAST) model for capturing the latent relationships among the tourists in each travel group. Finally, evaluate the TAST model, the TRAST model, and the cocktail recommendation approach on the real-world travel package data.

Mutasem K. Alsmadi [20] proposed A hybrid firefly algorithm with fuzzy c means algorithm. In this paper fuzzy C means algorithm get explored and compared it with other clustering algorithm i.e. k-means and concluded that fuzzy C means is better than K-means.

Subhagata Chattopadhyay focused on a comparative study of fuzzy c-means clustering. The performances of these algorithms have been compared with different datasets, in terms of the quality of the clusters obtained and their computational time.

B. Challenges with Recommendation System

Each recommendation system type has its own strengths and weaknesses. Recommendation system has some limitations which will extremely affect on recommendations' accuracy. Current CF methods suffer from problems such as data sparsity, recommendation inaccuracy and big error in predictions. The problems are as follows [3]-

1. Data Sparsity

The consumer-product interaction matrix by a $|C| \times |R|$ matrix $R = (r_{i,j})$ such that,

$$f(x) = \begin{cases} k, k = 1, 2 \dots 5 & \text{if user } i \text{ rated item } j. \\ 0, & \text{otherwise.} \end{cases}$$

In many large-scale applications, both the number of consumers and the number items of are large. In such cases, even when many events have been recorded, the consumer-product interaction matrix can still be extremely sparse, that is, there are very few elements in R whose value is not 0. This problem, generally referred as the sparsity problem, has a major negative impact on the effectiveness of a collaborative filtering approach. Due to

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

sparsity, it is highly probable that the similarity (or correlation) between two given users is zero, rendering collaborative filtering useless. Even for positively correlated pairs of users, such correlation measures may not be reliable.

2. Cold Start Problem

The situation in which a new user or item has just entered the system; cold start problem refers such problems [4]. Collaborative filtering is not able to generate useful recommendations for the new user because of the lack of sufficient previous ratings or purchases. Similarly, when a new item enters the system, it is unlikely that collaborative filtering systems will recommend it to many users as very few users have yet rated or purchased this item. Conceptually, the cold-start problem can be viewed as a special instance of the sparsity problem, where most elements in certain columns or rows of the consumer product interaction matrix A are 0.

3. Scalability

Scalability of recommendation system is generally understood as the ability to provide good quality recommendations independently of the size of dataset and its primary dimensions (number of users and items), its growth, and the dynamic of the growth [19].

4. The 'Long Tail'

Many recommender systems ignore newly introduced or unpopular items having only few ratings and focus only on items having enough ratings to be of real use in the recommendation algorithms. Alternatively, such unpopular or newly introduced items can remain in the system. It is one of the forms of cold star problem [18].

III. PROPOSED ALGORITHM

The main objective of proposed system is to cluster the items and then find "neighbors" of users based on user typicality degree in user groups (instead of the corated items of users, or common users of items, as in traditional CF) and predict the ratings.

In recommendation system the input is given as dataset which contains users, packages and ratings given by users for any packages. By using such inputs the expected outputs are recommended packages for a user and predicted ratings.

A. System Architecture:

The figure 1 shows the system architecture. The pre-processing is done first then it is followed by recommendation process.

1. Preprocessing Phase

In a recommendation system, there are set of users U and a set of items I. The preprocessing phase consists of 3 parts-

- **Item Grouping using Clustering** – Items can be clustered into several item groups (O). This grouping is done using Fuzzy clustering algorithm

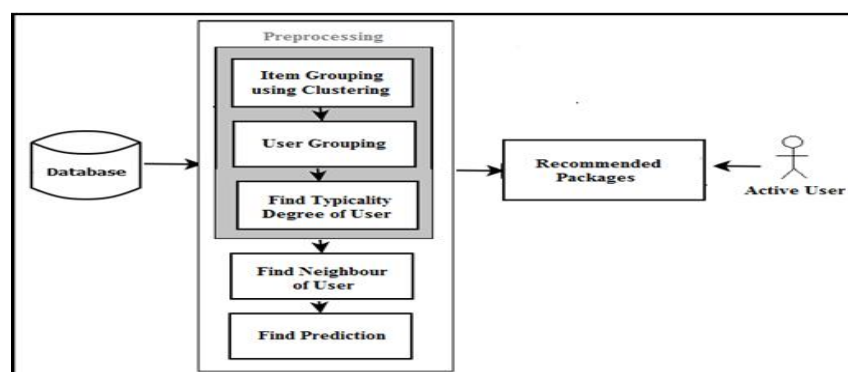


Figure 1: System Architecture



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

- **User Grouping** – Users who share similar interests on an item group could form a community, called as user group (U) and user grouping is done based on item grouping.
- **Typicality Measurement** – The most important step is to calculate typicality degree of user. Users in different user groups have different typicality degrees. The further process of recommendation system is based on this typicality measure [1].

2. Recommendation Process

The next processing is same as traditional recommendation system. This process consists of following steps –

- **Neighbor Selection** – The neighbor selection is a very important step before prediction. If the selected neighbors are not sufficiently similar to the active user then prediction ratings of an active user on items will be inaccurate. The neighbor selection is calculated using distance-based similarity measure.
- **Prediction** - Once the set of “Neighbors” of users is obtained then it is easy to predict the rating of an active user U_i on an item O_j .
- **Package Recommendation** – This is the final step of process in which the packages are recommended to an active user based on its neighbor user’s preference (calculated using typicality degree) that has similar kind of interest.

B. Algorithm

The proposed system implements following algorithm

Algorithm 1: Clustering and typicality based approach for recommendation

Input : Dataset having users, packages and ratings

Output : Recommended packages, Predicted ratings

1. Preprocessing :
 - a. Package grouping is done using fuzzy c-means clustering [Algorithm 2]
 - b. User grouping- For each package group form a group of users who like those package.
 - c. Calculate typicality degree of each user in that group. It can be calculated by considering two user properties-
 - i. Users having rated items in the corresponding package group to the highest degrees.
 - ii. Users having frequently rated packages in the corresponding package group.
2. Recommendation process
 - a. Neighbor selection is calculated using distance based similarity measure based on typicality degree.
 - b. Prediction is determined using set of neighbors of user and it will return weighted sum of all ratings given by neighbors of user.
 - c. Based on neighbors of user’s preference, system will return recommended packages.

FCM algorithm was selected as an alternative for the typical K-means algorithm to allow each element in the dataset to belong to more than one cluster. Despite of this improvement, the K-means algorithm still suffering from some drawbacks such as (low convergence rate and getting trapped in local minima) [20]. The pseudo-code of the FCM algorithm is described as the following:

Algorithm 2: Fuzzy C Means (FCM)

1. Initiates with c random initial cluster centers for each iteration.
 2. Calculate the membership matrix of each data point in each cluster.
 3. Cluster centers are recalculated for each iteration
 4. Repeat steps 2 and 3 until no further change in the cluster centers the FCM algorithm will be terminated
-

C. Mathematical Model

The mathematical model for clustering and typicality based collaborative filtering recommendation system is as follows -

- **I - Set of Inputs** - The input to the system is a dataset having users, packages and ratings.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

- Calculate set of Item groups
 - The item (i.e. package) groups can be calculated using algorithm 2- Fuzzy C Means Clustering. It has ability to minimize the objective function given below –

$$J_m = \sum_{j=1}^c \sum_{i=1}^n u_{ij}^m \|x_i - v_j\| \quad (1)$$

- The process of FCM algorithm initiates with c random initial cluster centers for each iteration. FCM algorithm used the following Equation 4 to find the fuzzy membership for each data point in each cluster

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left[\frac{\|x_i - v_j\|}{\|x_i - v_k\|} \right]^{\frac{2}{m-1}}} \quad (2)$$

- The cluster centers are recalculated based on the membership values using the following Equation 5:

$$v_j = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m} \quad (3)$$

When the value of the cluster centers is constant the FCM algorithm will be terminated.

- U – Calculate set of User groups

$$g_i = \{U_1^{v_{i,1}}, U_2^{v_{i,2}}, \dots, U_m^{v_{i,m}}\} \quad (4)$$

Where, U_m = Users and $V_{i,m}$ = Typicality degree of user U_m in user group g_i

- N – Neighbor Selection - Neighbors can be calculated as given below-

$$\vec{N}_j = \{U_i \mid \text{Sim}(U_i, U_j) \geq \gamma\} \quad (5)$$

Where, $\text{Sim}(U_i, U_j)$ – distance similarity measure and γ – Threshold value

- M – Calculate Sparsity Measure

The sparsity measure can be calculated by using distance similarity measure [17].

$$\text{Sim}(U_i, U_j) = \exp\left(-\sqrt{\sum_{y=1}^n |v_{i,y} - v_{j,y}|^2}\right) \quad (6)$$

Where n = number of user groups, $\sqrt{\sum_{y=1}^n |v_{i,y} - v_{j,y}|^2}$ = Euclidean distance Between U_i and U_j

- P – Calculate Prediction

The prediction can be calculated as

$$R(U_i, U_j) = \frac{\sum_{u_x \in \vec{N}_i} R(U_x, O_j) \cdot \text{Sim}(U_x, O_j)}{\sum_{u_x \in \vec{N}_i} \text{Sim}(U_x, O_j)} \quad (7)$$

Where, U_x is user in the set of “neighbors” of U_i and $R(U_i, U_j)$ indicates Rating of user U_x and U_i .

- O – Output of the system - The will output the recommended packages, predicted ratings.

IV. RESULTS AND DISCUSSION

A. *Data set* - To evaluate this recommendation method, we use the dataset that contains 150 travel packages and 5000 user’s ratings for those packages. The ratings follow the 1 to 5 numerical scales.

B. *Results* - The metric used to evaluate recommendation system is Mean Absolute Error (MAE) which is defined as the average absolute difference between predicted ratings and actual ratings. A lower MAE value means that the recommendation method can predict user’s ratings more accurately. Thus, for MAE values of a recommendation method, the smaller the better.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

Table I: Results for MAE with Different Train/Test Ratios for $\gamma = 0.6$

Method	$X = 0.3$	$X = 0.6$	$X = 0.9$	AVG
TBCF	0.7757	0.7481	0.7349	0.7529
CWTB	0.7643	0.7393	0.7217	0.7417

To test MAE, we set γ to 0.6. A variable X named train/test ratio, to denote the percentage of data used as the training and test sets. A value of X , 0.9 means that 90 percent of data are used as the training set and the other 10 percent of data are used as the test set. We set $X = 0.3, 0.6$ and 0.9 to and take the average MAE. As the table shows proposed method CWTB has lower MAE than TBCF that means it can predict more accurate user ratings.

Another measure used is Coverage. It measures the percentage of items for which a recommender system is capable of making predictions. Larger the coverage values are better for recommendation that means it can predict more ratings for users on unrated items. For example, if recommendation system can predict 8500 out of 10,000 ratings then the coverage is 85%.

Table II: Results for Coverage with Different Train/Test Ratios for $\gamma = 0.6$

Method	$X = 0.3$	$X = 0.6$	$X = 0.9$	AVG
TBCF	0.9874	0.9895	0.9896	0.9888
CWTB	0.9877	0.9897	0.9898	0.9890

As table II shows CWTB can predict more unknown ratings of users because it has slightly more coverage than TBCF. But this coverage is better than item based and user based CF. Following fig.2 and fig.3 shows graph plotted for MAE and Coverage respectively according to values of Table I and II. Using fuzzy C means clustering for preprocessing will help to get more accurate neighbors and will have a chance to get more coverage. Also time cost which further results can be examined and improved.

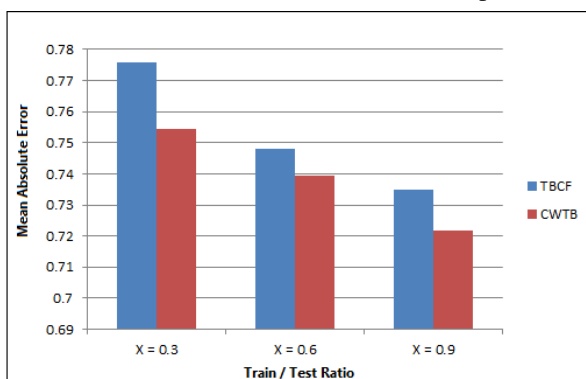


Figure 2: Graph for MAE

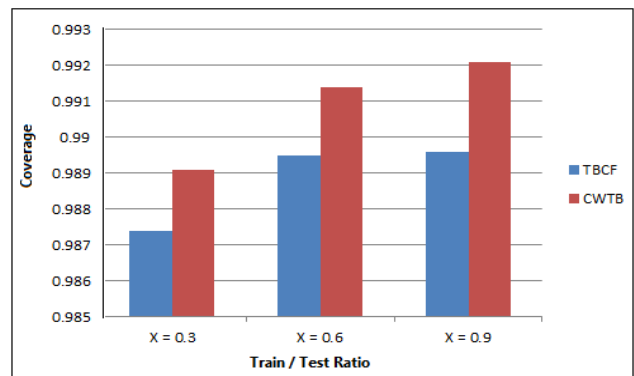


Figure 3: Graph for Coverage

V. CONCLUSION AND FUTURE WORK

The Recommendation system is very useful system for customer as well as for provider. This system has some challenges like data sparsity, scalability and accuracy. The proposed perspective of collaborative filtering recommendation method named 'clustering and typicality based approach' helps to overcome these challenges. It will find the "neighbour" of user depending upon typicality degree of user in the item group. It will address the problem of data sparsity. This outperforms many CF recommendation methods on recommendation accuracy (in terms of MAE) with an improvement and has more coverage and it will predict more unknown ratings.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

REFERENCES

1. Y. Cai, H.-F. Leung, Q. Li, H. Min, J. Tang, and J. Li, "Typicality based collaborative filtering recommendation" IEEE Transactions on Knowledge and Data Engineering, vol. 26, pp. 766-779, March 2014.
2. Q. Liu, E. Chen, H. Xiong, Y. Ge, Z. Li, and X. Wu, "A cocktail approach for travel package recommendation," IEEE Transactions on Knowledge and Data Engineering, vol. 26, pp. 278-293, Feb 2014.
3. J. Borrs, A. Moreno, and A. Valls, "Intelligent tourism recommender systems: A survey," Expert Systems with Applications, vol. 41, no. 16, pp. 370 -389, 2014.
4. Y. Cai, R. Y. Lau, S. S. Liao, C. Li, H.-F. Leung, and L. C. Ma, "Object typicality for effective web of things recommendations," Decision Support Systems, vol. 63, no. 0, pp. 52- 63, 2014.
5. Sagarika Bakshi, Alok Kumar Jagadev, Satchidananda Dehuri, and Gi-Nam Wang. "Enhancing scalability and accuracy of recommendation systems using unsupervised learning and particle swarm optimization", Applied Soft Computing, 15(0):21- 29, 2014.
6. Pawan Kumar Tiwaria, Deo Prakash Vidyarthib. "Evolutionary Computation and Its Applications: A Survey". Information Engineering Letters, ISSN: 2160-4114 Volume 2, Number 1, March, 2012.
7. Shweta Singh¹, Shivangi Goyal², Shubhra Goyal³ and Divya Bhatt⁴ "A Comparative Study of a Class of Nature Inspired Algorithms", Computing For Nation Development, March 10 – 11, 2011.
8. Lee, C. S., Chang, Y. C., Wang, M. H. "Ontological recommendation multi agent for Tainan city travel", Expert Systems with Applications, 36, 6740–6753, 2009.
9. Meehan, K., Lunney, T., Curran, K., & McCaughey, "A.Context-aware intelligent recommendation system for tourism" In Proceedings of the 11th IEEE international conference on pervasive computing and communications (pp.328–331), 2013.
10. Kurata, Y., & Hara, T. "CT-planner4: Toward a more user-friendly interactive day-tour planner", Springer International Publishing, In Information and communication technologies in tourism (pp. 73–86), 2013.
11. Martínez-Santiago, F., Ariza-López, F., Montejo-Ráez, A., & Ureña-López, "A.GeOasis: A knowledge-based geo-referenced tourist assistant", Expert Systems with Applications, 39(14), 11737–11745, 2012.
12. Gavalas, D., & Kenteris, M." A web-based pervasive recommendation system for mobile tourist guides".Personal and Ubiquitous Computing, 15, 759–770, 2011
13. Lucas, J., Luz, N., Moreno, M., Anacleto, R., Figueiredo, A., & Martins, C. " A hybrid recommendation approach for a tourism system" Expert Systems with Applications, 40(9), 3532–3550, 2013
14. Pinho, J., da Silva, B., Moreno, M., de Almeida, A. M., & Lopes, C. "Applying recommender methodologies in Tourism sector", Springer Highlights in practical applications of agents and multiagent systems. In Advances in Intelligent and Soft Computing (pp. 101–108), 2011.
15. Z. Huang, H. Chen, and D. Zeng, "Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering," ACM Trans. Inf. Syst., vol. 22, pp. 116, 142, Jan. 2004.
16. G. Adomavicius and A. Tuzhilin. "Towards the next generation of recommender systems: a survey of the state-of-the-art and possible extensions", IEEE Trans. on Data and Knowledge Engineering, 17(6):734-749, 2005.
17. Deepa Anand and Kamal K. Bharadwaj. "Utilizing various sparsity measures for enhancing accuracy of collaborative recommender systems based on local and global similarities", Expert Systems with Applications, 38(5):5101- 5109, 2011.
18. Shweta Tyagi and Kamal K. Bharadwaj."Enhancing collaborative filtering recommendations by utilizing multi-objective particle swarm optimization embedded association rule mining", Swarm and Evolutionary Computation, 13(0):1- 12, 2013.
19. Jesus Bobadilla, Fernando Ortega, Antonio Hernando, and Javier Alcal. "Improving collaborative filtering recommender system results and performance using genetic algorithms", Knowledge-Based Systems, 24(8):1310- 1316, 2011.
20. Mutasem K. Alsmadi. "A hybrid firefly algorithm with fuzzy c means algorithm for mri brain segmentation", American Journal of Applied Science, 2014.

BIOGRAPHY

Parnika Patil received the BE in Computer Engineering from J.J.M.C.O.E which is under Shivaji University, in 2008, and currently pursuing Master of Engineering degree in Computer Engineering from D. Y. Patil College of Engineering, Akurdi, Pune. Her research interests include Data Mining and Soft Computing.

V. L. Kolhe is an Assistant Professor in the Computer Engineering Department, D. Y. Patil College of Engineering, Akurdi, Pune. She received the BE degree in Computer Technology from Nagpur University, and ME degree in Computer Engineering from Savitribai Phule Pune University.