I.J. Education and Management Engineering, **2017, 2, 50-62**
Published Online March 2017 in MECS (http://www.mecs-press.net)
DOI: 10.5815/ijeme.2017.02.06

Available online at http://www.mecs-press.net/ijeme

Research Work Area Recommendation based on Collaborative Filtering

Richa Sharma [a], Sharu Vinayak [b], Rahul Singh [a,b,*]

[a] *Chandigarh University, Mohali, India*
[b] *Chandigarh University, Mohali, India*
[a,b,*] *Chandigarh University, Mohali, India*

Abstract

In this work we present RWARS, a novel recommender system that recommends research work area. So far a number of recommender systems have been developed in the field of e-commerce, e-services, e-library, entertainment, tourism and social networking sites. However, when it comes to the area of education, not much work has been done. So to extend the utility of Recommender systems in the field of education, we have developed RWARS. We have used Cosine similarity and Tanimoto coefficient for developing our system. The aim of this work is to compare the results obtained using each approach to find the most optimal one. Evaluation parameters that have been used are: Mean square error, Root mean square error and Coverage. At present, RWARS is still in its initial phase and its applicability can be further enhanced by converting it into an online system and it surely will prove to be a great boon for young researchers to select the most appropriate research area for them.

Index Terms: Collaborative filtering, Cosine similarity, Tanimoto coefficient, Recommender systems.

1. Introduction

Recommender systems are software tools or techniques which narrow down the choices of products or items available in abundance and suggest the most suitable ones, based on the implicit or explicit user ratings. We can say that a Recommender system is a system of: users, a user interface, a dataset and some recommendation algorithms. We can say that "Recommender systems are- for the user, from the user and by the user" as the recommendations are made for the users and the dataset used for making the recommendations is also taken from the users themselves in the form of implicit or explicit ratings. So users act as the backbone of Recommender systems. The user-interface acts as an intermediate between the user and the system. The user

Corresponding author.
E-mail address:

expresses his preferences, ratings or requirements through the user-interface only and the recommendations are also displayed on the same. The dataset is in the form of user ratings or feedback. Finally, once the user requirements are gathered, some recommendation algorithm is applied. Fig.1 gives an overview of how the framework of recommender systems works.

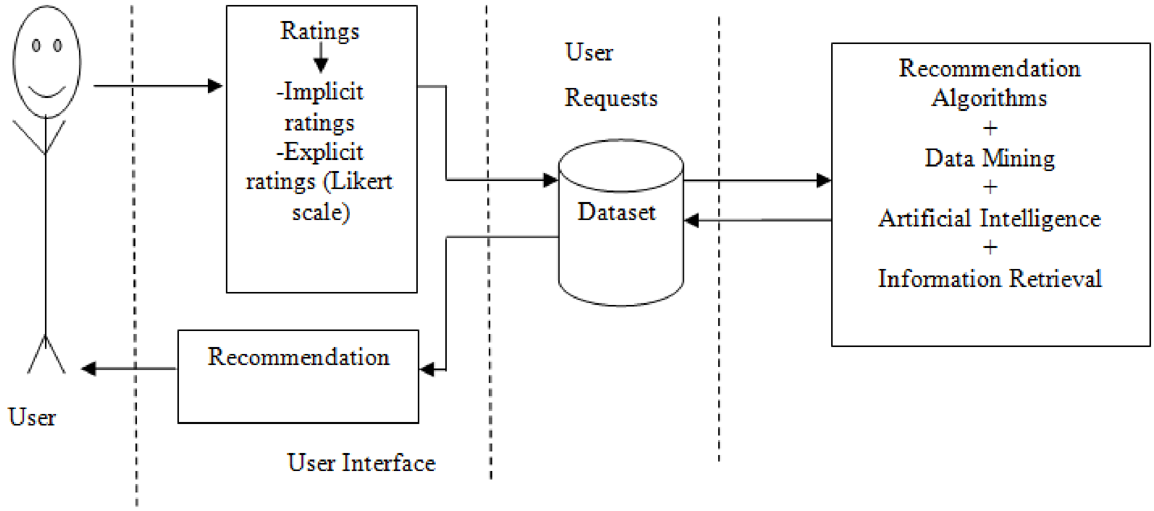

Fig.1. Recommender Framework

There are various Recommendation approaches developed so far: Collaborative Filtering, Content-based Filtering, Knowledge-based systems, Hybrid systems, Demographic recommender systems and Community-based systems. We have recommendations for almost every field ranging from: social networking sites, marriage portals, e-commerce sites, e-library, e-services to movie, music or book recommender system, These systems tend to play a really vital role in internet sites like Amazon, Pandora, Hulu, LinkedIn, Netflix, YouTube, Jester, CiteSeer, Yahoo Matrimonial sites and Social networking sites. These websites use either one of the recommendation approach or a combination of two or more. For example, Amazon uses Item to Item collaborative filtering and content-based filtering, Pandora uses content-based filtering, CiteSeer uses collaborative and content-based filtering, and YouTube uses personalized recommendations while Jester and LinkedIn use collaborative filtering. Over the past two decades, Recommender Systems have evolved tremendously and now have also included the concepts of Artificial intelligence, Information retrieval and Human-computer interaction in their journey[1], and hence have gained more popularity over the time, being more efficient.

Since, Recommender Systems have abridged our effort of going through each and every item for selecting the most suitable one, a Recommender System that can suggest a research work area to a student that he or she can do his research in, would prove to be quite useful. In this paper, we introduce RWARS (Research Work Area Recommender System) that recommends an area for research work. This system finds the similarity between the users in terms of their future objectives, the type of research they want to do, their hobbies, subjects of interests and their programming skills. We have used Cosine similarity and Tanimoto coefficient to find the similarity between the users. Our aim is to find the most suitable approach for our system. As for the dataset, we have collected our own data by conducting a small survey of Post-graduate students who have already worked in some research area. We anticipate that our proposed system will prove to be beneficial.

This paper is organized as follows. Section II evaluates the various works relative to the proposed system. Section III reports the framework of the proposed system. In section IV, various aspects of the system are discussed and the processes in each phase are elaborated. Finally, the paper is concluded in Section V.

2. Related Work

Recommender system has taken over e-commerce, being one of the most popular and impressive tool[1]. The idea that motivated the development of recommender system is to help the users with the selection process of any item, the options in front of them being available in abundance[1]. There are various Recommendation approaches, the most popular being: Collaborative Filtering, Content-based systems, Hybrid systems, Knowledge-based systems, Demography-based systems and Community-based recommender systems[1].

Collaborative filtering ,also known as 'people-to-people co-relation'[2], is based on a simple idea that users having similar interests in the past, will most probably have alike interests in the future also[2]. Collaborative filtering is further classified into user-user and item-item collaborative filtering and can be calculated using: Cosine similarity, Pearson coefficient, Euclidean distance, Tanimoto coefficient and Spearman co-relation[3]. Content-based systems are based on the concept "Show me more of what I have liked"[4]. The purpose is to recommend only those products to users, which are identical to the ones that the user has already shown his interest in, in the past[1].

Knowledge-based systems follow "Tell me what fits my needs"[4]. These systems exploit domain knowledge of a specific product to create recommendations for the user[1]. The system fetches the user requirements, matches it with its knowledge base and makes the relevant recommendations to the user, keeping the preferences in consideration. Hybrid system gives the best of both worlds. As the name says, these systems combine the features of two or more recommendation approaches. The main objective is to combine the features of two systems (recommendation approaches) in such a way that the drawbacks of one approach are overcome by the other[4].

The first recommender system developed was Tapestry[5]. The drive that led to its development was the increasing number of incoming emails, generally pointless ones that were too difficult to maintain. And to overcome this problem, the concept of Collaborative filtering was introduced. Nazpar Yazdanfar *et al.*[6] introduced a Link Recommender system that recommends URLs to twitter users using nearest neighbour approach. They considered hash tags as the topic representatives of the URLs. The authors used Euclidean distance, Jaccard coefficient, Dice coefficient and Cosine similarity measures to find the item-hash tag similarity and item-user similarity. Sobia Zahra *et al*.[7] proposed a k-means clustering based recommendation algorithm to address the scalability issues of recommender systems. The authors implemented a series of novel approaches of centroid selection for k-means clustering along with the traditional approaches, to compare their performance. And for evaluation purpose, mean absolute error and coverage were taken into consideration.

Aansi A. Kothari *et al.*[8] in their work proposed that a Support vector machine learning technique, based on both contextual and non-contextual user preferences, for increasing the accuracy of recommender systems. The objective was to establish a relationship between a user's personal preferences and his profile information. Data from TripAdvisor were taken to conduct the experiments. Results revealed that SVM, when used with collaborative filtering, gave better accuracy and precision. D. De Nart *et al.*[9] in their work proposed a content-based recommender system for scientific libraries. The system used the concept of key phrase extraction to extract the metadata or keywords from the documents. Data from CiteSeer were used by the proposed system, the system being based on network rather than vector. Results revealed that the system was free from cold-start problem and did not require user ratings for generating recommendations.

Pallab Dutta *et al.*[10] proposed a novel approach to spot leaders in a social network site taking trust factor into consideration. The purpose of the proposed system was to study the effects of leaders in a social network and how does they affect the trust of the users. Abhilash Reddy *et al.*[11] developed a personalized travel package recommendation system that made the recommendations to the user, based on his social network profile. The system considered the location of the user for making relevant recommendations. Gurpreet Singh *et. al.*[12] proposed a novel hybrid music recommender system . The system used k-means clustering to find the similarity between the users and similarity between the songs was calculated using PLSA technique.

Tanimoto Coefficient is the measure of the ratio of the intersection of two sets to the union of the sets[13]. In terms of recommender systems, it can be defined as the ratio of the number of items liked by both the user a and b, to the total number of items liked by them individually. This approach is said to be the most useful in making the recommendations in case of extremely sparse or scattered datasets. Mojtaba *et al*.[14] in their work presented a recommender system for recommending learning material to the users. Their work was based on genetic algorithm along with a multi-dimensional information model. The authors used collaborative filtering and content based filtering, first individually and later a hybrid of these two approaches. Kwanghee *et al.*[15] in their work proposed a personalized research paper recommender system that recommends research paper to the user based on his profile. The authors used the concept of keyword extraction to find the similarity between the topics that the user is willing to work in and the research papers available on the internet. Cosine similarity was used to find the similarity between the desired user topic and the ones available on the web. Results showed that this system proved to be quite efficient in making the search for research papers easier.

Pu Wang[16] presented a personalized collaborative filtering approach that was based on the clustering of customers, with the aim to overcome the scalability issue. K-means clustering, Cosine and similarity and nearest neighbour approach were used. The results revealed that along with scalability issue, problems like data sparsity and cold-start problem were also resolved. Diksha Nagpal *et al.*[17] introduced a novel faculty recommender system that recommends faculty members to the students, as per their relevant area of expertise and the ratings given to them by the students. This system used both user-user and item-item collaborative filtering approach. Results of the study revealed that item-item collaborative filtering gave better results than user-user collaborative filtering.

Based on the previous work done, we concluded that some more work can still be done in the field of education by developing a Research Work Area Recommender System. We have article recommender system, faculty recommender system, recommender system for digital libraries and even research paper recommender systems. But selecting what research area to work on is in itself a very tedious task as it requires going through a lot of research works to decide which one interests you the most. With our system, we can not only save time and effort, but we can also make this selection process a lot easier and simpler. The user simply has to answer a short questionnaire and in no time, he or she would be provided with the most suitable recommendations. So we propose RWARS (Research Work Area Recommender System) that would recommend research work areas to the users based on their interests, hobbies, programming skills, subjects of interests and future objectives.

3. Research Work Area Recommendation

As shown in Fig.2, RWARS is divided into three phases: Data gathering, Fetching new user data and Making recommendations.

3.1 Data gathering

For the database of the proposed system, we have first prepared a small survey and then asked the Post-graduate students to fill the survey. The data gathered from the survey is entered into the system manually and is used as the dataset. We have collected the data of 670 post graduate students pursuing their master's degree in computer science.

3.2 Fetching new user data

We have designed a simple User-interface, where a new user has to fill the form as shown in Fig.3. Once the user fills the form and clicks on the submit button, his information is stored in a temporary database. All the algorithms work at the back-end and once this is done; the results are displayed on the user-interface only.

For convenience, we have stored the database in the form of 0s and 1s, .i.e. for the positive responses of the user; we have saved the data as 1 and as 0 in case of negative responses.

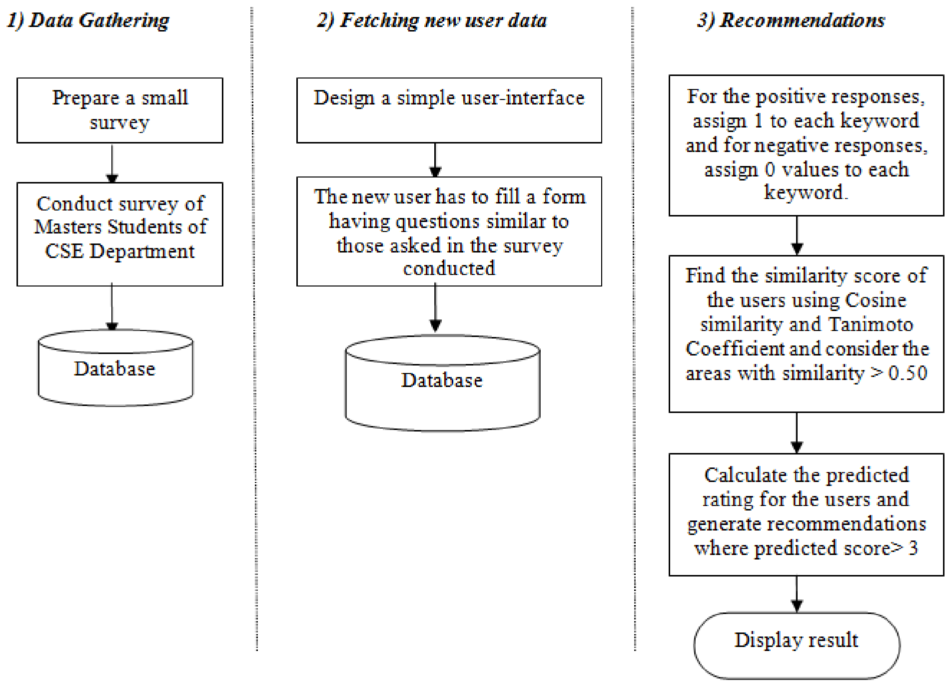

Fig.2. Process of the System

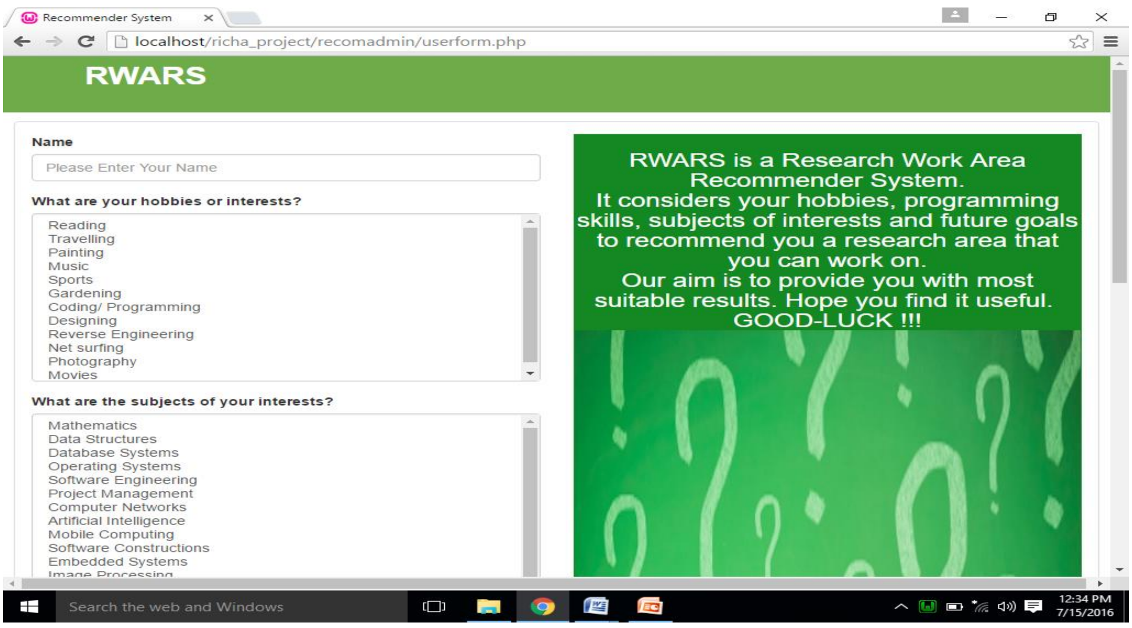

Fig.3. User-Interface

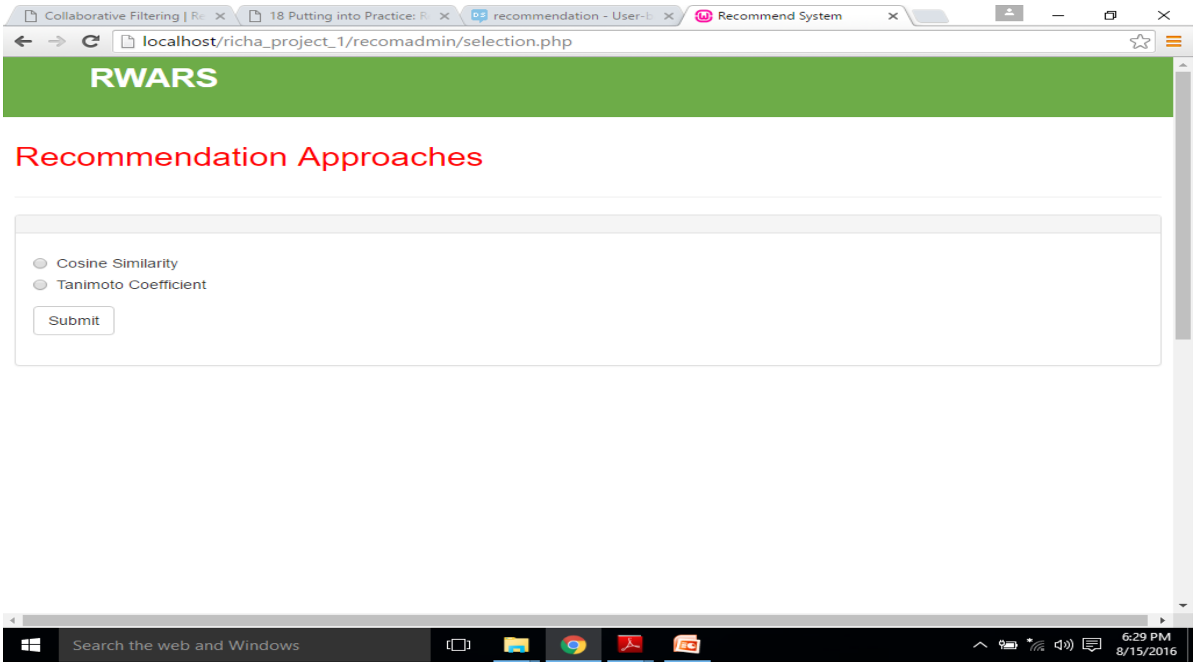

Fig.4. Recommendation Approaches

3.3 Making Recommendations

Collaborative filtering (CF) is used for developing the proposed system. CF further uses two approaches for generating the recommendations: Cosine similarity and Tanimoto coefficient. For RWARS, both these approaches are implemented to see which one provides the better results.

a. *Finding similarity using Cosine similarity-* Cosine similarity between two vectors is a measure of the cosine of the angle between them[18]. It is a vector based approach based on linear algebra. Users are considered as vectors and similarity between them is measured by the cosine distance or dot product of the angles between the rating vectors[2]. It is given by:

$$\text{sim}(a, b) = \frac{r_a \, r_b}{\|r_a\|_2 \|r_b\|_2} \tag{1}$$

In Equation 1,

r_a is the rating given by user a
r_b is the rating given by user
$\|r_a\|_2 \|r_b\|_2$ is the dot product of vector product of the ratings given by user a and b

Equation 1 can also be stated as,

$$\text{sim}(a, b) = \Sigma r_a r_b \div \sqrt{\Sigma (r_a)^2} \sqrt{\Sigma (r_b)^2} \tag{1.1}$$

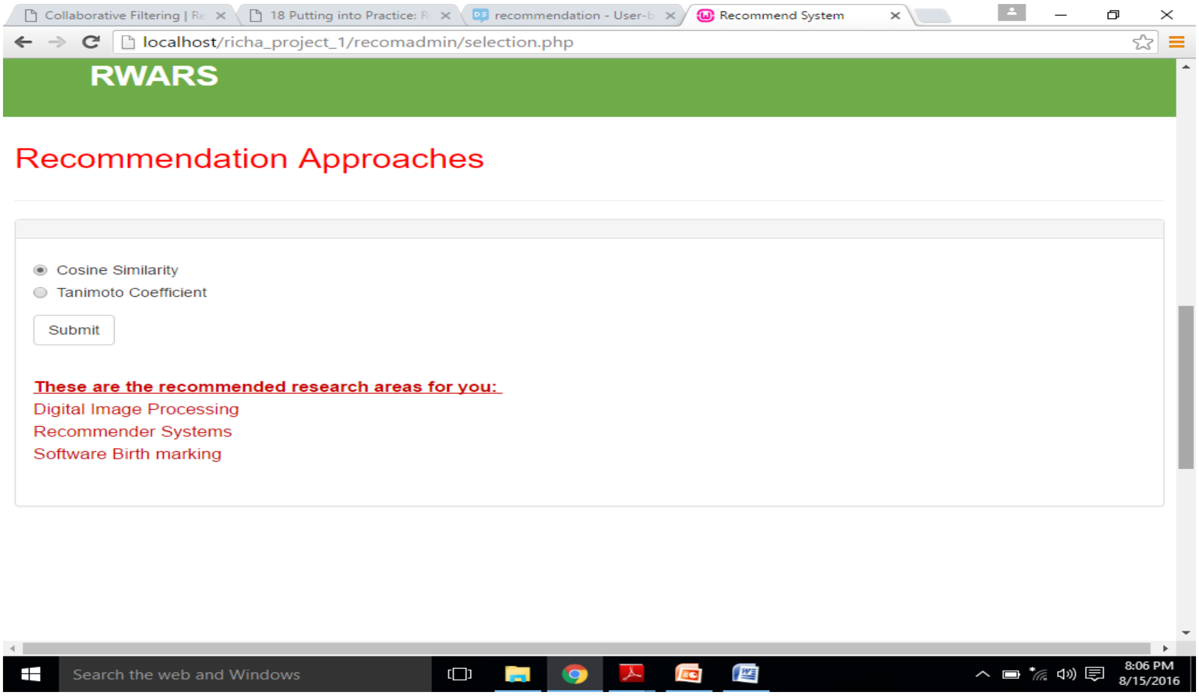

Fig.5. Results obtained using Cosine Similarity

b. *Finding similarity using Tanimoto coefficient*- It is also known as Jaccard similarity coefficient. It is a measure of similarity between finite numbers of sets and is the ration of the intersection of the given sets to the union of the sets[19]. It is given by:

$$S_{AB} = \frac{a \cap b}{a \cup b} = \frac{c}{a+b-c} \quad (2)$$

In Equation 2,

S_{AB} is the similarity between a and b
a gives the bits set to 1 in A
b gives the bits sets to 1 in B
c gives the no. of 1 bits common to both A & B
a +b gives the total no. of bits set to 1 in A & B

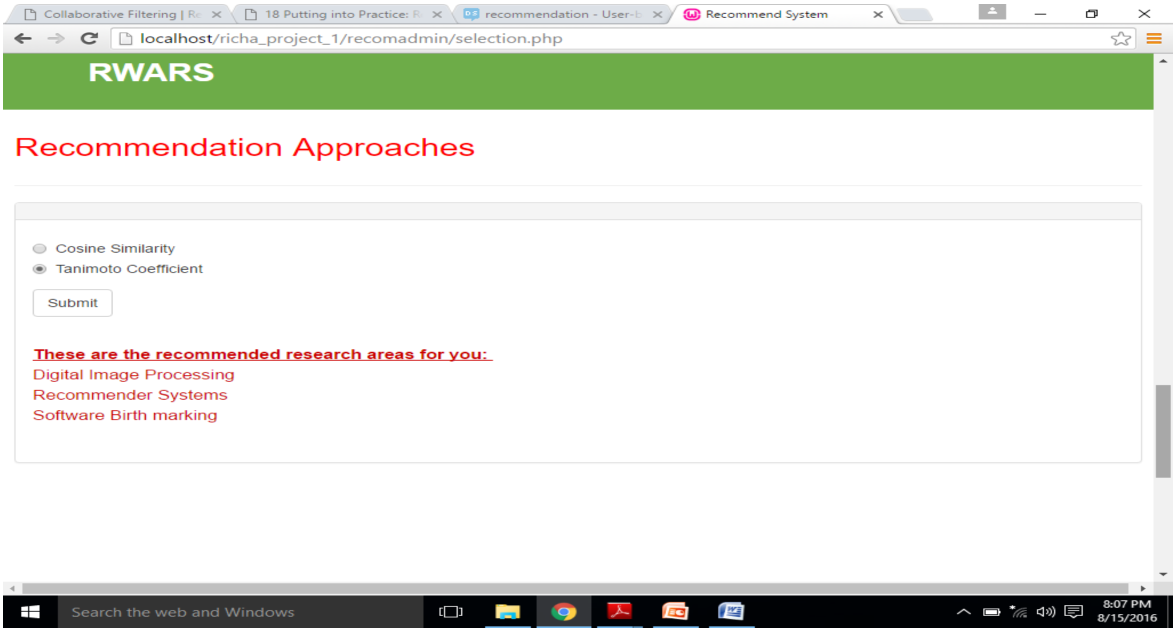

Fig.6. Results Obtained using Tanimoto Coefficient

c. *Calculate Predicted rating*- To generate the recommendations the values of the prediction is used. For finding the predicted score, the areas with similarity > .50 are considered.

$$\text{Pred}_{ab} = \frac{\sum r_b * sim_{a,b}}{\sum sim_{a,b}} \qquad (3)$$

In Equation 3,

$\Sigma\ sim_{a,\,b}$ is the similarity calculated using cosine similarity and tanimoto coefficient
r_b is the rating given by user b to the research area

In this work, the new user is considered as user a and rest all the users of the dataset are considered as user b. Then the cosine similarity algorithm and tanimoto coefficient is applied to calculate the similarity score of the users (a and b). Further, the predictions are generated for a user a by using the ratings of the most similar users or the nearest neighbours. Only the research areas with similarity greater than 0.50 are considered further to find the predicted score or rating for the new user.

Finally, the research areas with predicted rating higher than 3 are recommended to the user. For each user, results obtained using each approach are compared to analyze which approach gives the most optimal results and is suitable RWARS. For evaluation of the recommendation approaches, we have considered the MAE, RMSE and Coverage.

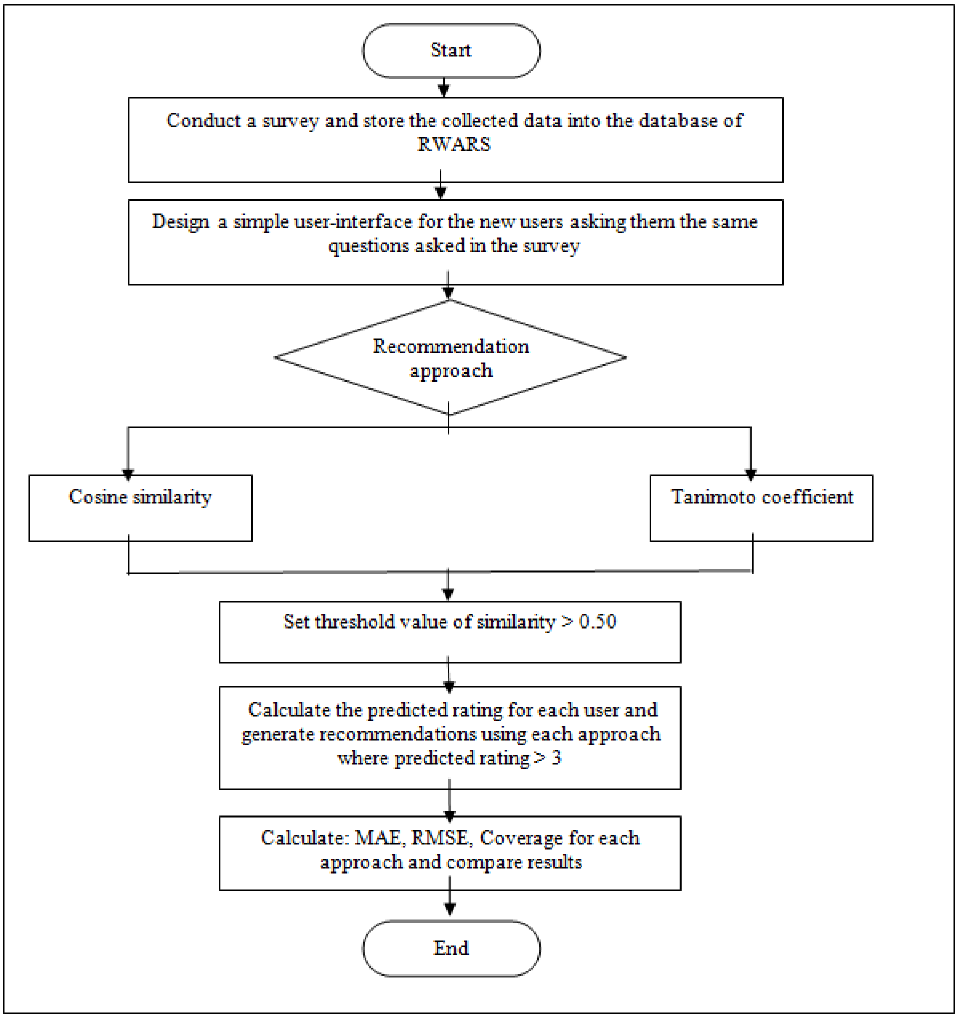

Fig.7. Flowchart of the Proposed Methodology

4. Results Discussion

The proposed system generated a number of recommendations with both collaborative filtering approaches (cosine similarity and tanimoto coefficient). 50 users were asked to use the system and for them some recommendations were generated. For convenience, a threshold value of user similarity > 0.50 is considered .i.e. only the recommendations with user similarity greater than 0.50 are considered valid recommendations while the recommendations with similarity < 0.50 are not considered by the system. Then, the prediction values for those users are calculated and the prediction value is used to generate user recommendations. For each user, three recommendations are generated with prediction rating >3.

Finally, the performance of all the approaches: Cosine similarity and Tanimoto coefficient are evaluated on the basis of: Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Coverage. Table 1 and 2 represent the manually calculated recommendation results which were generated after applying both approaches.

Table 1. Manually Gathered Data

Total users (Master's students of CSE Department)	642
Total number of Research Areas	35

Table 2. Manually Calculated Recommendations using All the Algorithms

Recommendation using Cosine similarity		**Recommendation using Tanimoto coefficient**	
Total recommendations generated	150	**Total recommendations generated**	150
Recommended Research areas	30	**Recommended Research areas**	27

4.1. Evaluation Metrics:

The recommendations generated using each approach is evaluated on the basis of: Mean absolute error (MAE), Root mean square error (RMSE) and Coverage. This evaluation is done using an offline experiment. Also, for the recommendations to be considered as valid, we have set a threshold value of 50% i.e. all the recommendations having similarity value greater than the threshold value are considered as valid recommendations and are further used for finding the prediction score for the user. Finally, the research areas with prediction score higher than 3 are recommended to the user. The comparison of results obtained using each approach is given in table 3.

Table 3. Results

Evaluation Metrics	**Recommendation Approach**	
	Cosine similarity	**Tanimoto coefficient**
MAE	0.09	0.19
RMSE	0.11	0.24
Coverage	85%	77%

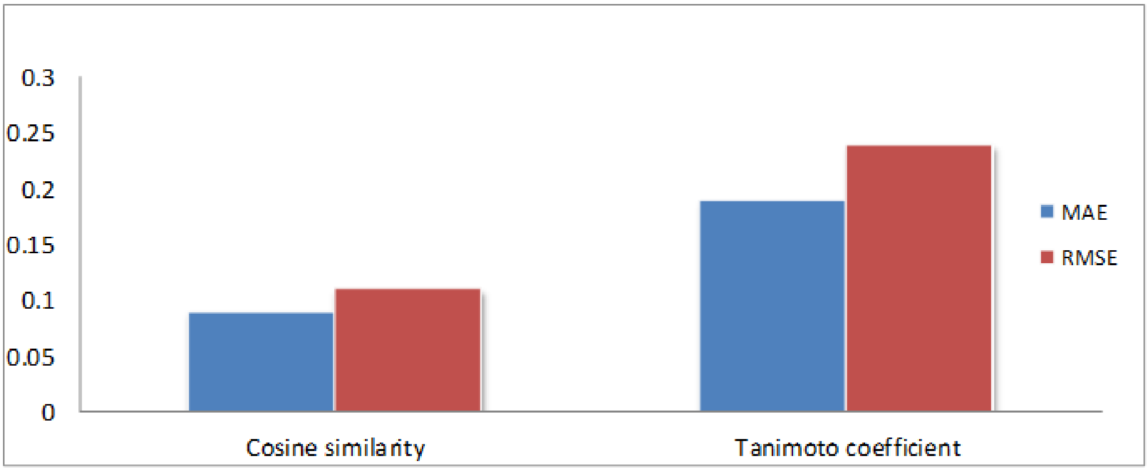

Fig.8. Graphical Representation of MAE and RMSE

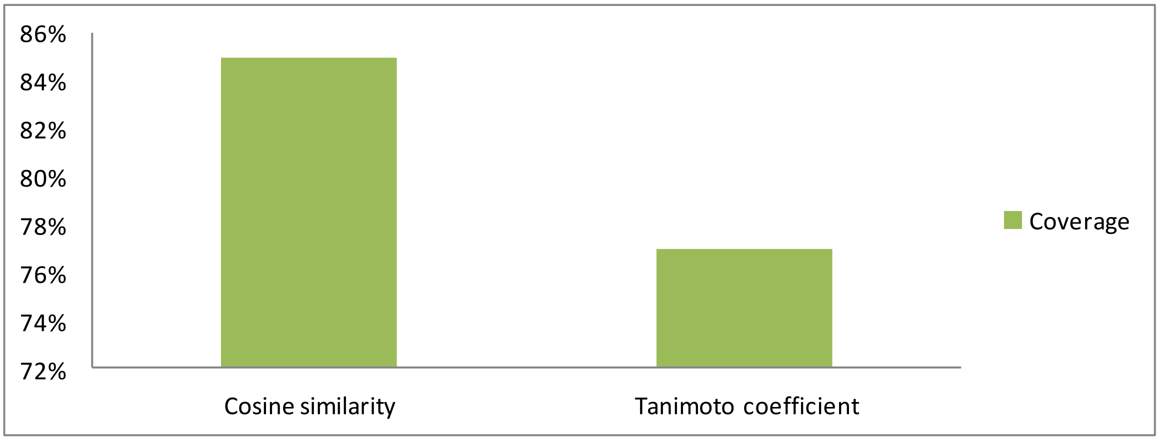

Fig.9. Graphical Representation of Coverage

Based on Figure 8, it is clear that the results obtained from Cosine similarity have a lower MAE and RMSE value as compared to Tanimoto coefficient. Therefore, it is concluded that Cosine similarity has a higher accuracy (lower the value of MAE and RMSE, higher the accuracy) as compared to Tanimoto coefficient. From Figure 9, it can be concluded that Cosine similarity has a higher coverage value as compared to Tanimoto coefficient. Higher coverage implies that the number of different research areas that were recommended was higher in case of Cosine similarity.

Therefore, based on the results it is concluded that Cosine similarity performed the best as compared to Tanimoto coefficient with more accurate results and a higher coverage.

5. Conclusion

In this paper, we introduced a novel recommender system RWARS that recommends a research work area to the user based on the similarity of user characteristics between the new user and the ones who have already worked in some research area. A user has to answer a simple questionnaire asking him about his skills, subjects of interests, future objectives etc. and based upon these answers he is provided with the recommendations.

The main objective of this work is to extend the utility of Recommender systems in the field of education. Collaborative filtering approaches have been deployed for building RWARS: cosine similarity and tanimoto coefficient. To compare the results obtained using both approaches, the values of Mean absolute error (MAE), Root mean square error (RMSE) and coverage are calculated. All this has been done manually. From the results, it is concluded that Cosine similarity performed the best for the proposed work with higher accuracy and more coverage as compared to Tanimoto coefficient.

The proposed system is a novel idea and it can be extended further for better results and better reliability. So far only Collaborative filtering techniques have been deployed for the system, but in future other recommendation approaches too can be applied and then the results obtained from each approach can be compared. Since this system is offline at present, it can also be converted into an online version where the data can be gathered online and the results can also be evaluated online only. Also, just the name of the research area might not be sufficient for the user to reach to a certain decision, so in future; the aim is to provide the links of the most recent literature review of the research areas too along with the names thereby making it convenient for the user to take the decision more effectively.

References

[1] Ricci, Francesco, Lior Rokach, and Bracha Shapira. *Introduction to recommender systems handbook*. Springer US, 2011.

[2] Jannach, Dietmar, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. *Recommender systems: an introduction*. Cambridge University Press, 2010.

[3] Ekstrand, Michael D., John T. Riedl, and Joseph A. Konstan. "Collaborative filtering recommender systems." *Foundations and Trends in Human-Computer Interaction* 4, no. 2 (2011): 81-173.

[4] Jannach, Dietmar, and Gerhard Friedrich. "Tutorial: recommender systems." In *Proceedings of the International Joint Conference on Artificial Intelligence, Barcelona*. 2011.

[5] Huttner, Joseph. "From Tapestry to SVD: A Survey of the Algorithms That Power Recommender Systems." (2009).

[6] Yazdanfar, Nazpar, and Alex Thomo. "Link recommender: Collaborative-Filtering for recommending URLS to Twitter users." *Procedia Computer Science* 19 (2013): 412-419.

[7] Zahra, Sobia, Mustansar Ali Ghazanfar, Asra Khalid, Muhammad Awais Azam, Usman Naeem, and Adam Prugel-Bennett. "Novel centroid selection approaches for KMeans-clustering based recommender systems." *Information Sciences* 320 (2015): 156-189.

[8] Kothari, Aansi A., and Warish D. Patel. "A Novel Approach Towards Context Based Recommendations Using Support Vector Machine Methodology." *Procedia Computer Science* 57 (2015): 1171-1178.

[9] De Nart, Dario, and Carlo Tasso. "A personalized concept-driven recommender system for scientific libraries." *Procedia Computer Science*38 (2014): 84-91.

[10] Dutta, P. and Kumaravel, A., 2016. A Novel Approach to Trust based Identification of Leaders in Social Networks. *Indian Journal of Science and Technology*, *9*(10).

[11] Reddy, C.A. and Subramaniyaswamy, V., 2015. An Enhanced Travel Package Recommendation System based on Location Dependent Social Data. *Indian Journal of Science and Technology*, *8*(16), p.1.

[12] Singh, G., Boparai, R.S. and Kathpal, M., 2016. A Novel Hybrid K-Means PLSA Technique for Music Recommender. *Indian Journal of Science and Technology*, *9*(16).

[13] Verbert, Katrien, Hendrik Drachsler, Nikos Manouselis, Martin Wolpers, Riina Vuorikari, and Erik Duval. "Dataset-driven research for improving recommender systems for learning." In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, pp. 44-53. ACM, 2011.

[14] Salehi, Mojtaba, Mohammad Pourzaferani, and Seyed Amir Razavi. "Hybrid attribute-based recommender system for learning material using genetic algorithm and a multidimensional information model." *Egyptian Informatics Journal* 14, no. 1 (2013): 67-78.

[15] Hong, Kwanghee, Hocheol Jeon, and Changho Jeon. "Personalized Research Paper Recommendation System using Keyword Extraction Based on UserProfile." *Journal of Convergence Information Technology* 8, no. 16 (2013): 106.

[16] Wang, Pu. "A Personalized Collaborative Recommendation Approach Based on Clustering of Customers." *Physics Procedia* 24 (2012): 812-816.

[17] Nagpal, Diksha, Sumit Kaur, Shruti Gujral, and Amritpal Singh. "FR: A Recommender for Finding Faculty Based on CF Technique." *Procedia Computer Science* 70 (2015): 499-507.

[18] Perone, C. S. "Machine Learning: Cosine Similarity for Vector Space Models (Part III). Pyevolve." *Availa ble: http://pyevolve. sourcefor ge. net/wordpress*.

[19] Wikipedia contributors, "Jaccard index," *Wikipedia, The Free Encyclopedia*, https://en.wikipedia.org/w/index.php?title=Jaccard_index&oldid=704793261(accessed May 25, 2016).

Authors' Profiles

Richa Sharma: Belongs to Bilaspur, Himachal Pradesh, India and born on March 6, 1992. She is a Research student for Masters Degree for Computer Science Engineering in Chandigarh University. She received her Bachelor's Degree from Punjab Technical University in 2014. The field of her thesis is Recommender Systems. Her areas of interests for research are R language for building Recommender Systems, Data Privacy and Security.

Sharu Vinayak: Belongs to Ludhiana, Punjab, India and born on October 8, 1992. She is currently pursuing Masters in Computer Science and Engineering from Chandigarh University and has done Bachelor's degree in Information Technology from Ludhiana College of Engineering and Technology. Programming and learning about new technologies interests the author.

Rahul Singh: Belongs to Kanpur, Uttar Pradesh, India and born on April 15, 1989. He is an Assistant Professor in CSE Department of Chandigarh University. He has received his Masters degree from Thapar University. His areas of interests are Theory of Computation and Software Engineering.

How to cite this paper: Richa Sharma, Sharu Vinayak, Rahul Singh,"Research Work Area Recommendation based on Collaborative Filtering", International Journal of Education and Management Engineering(IJEME), Vol.7, No.2, pp.50-62, 2017.DOI: 10.5815/ijeme.2017.02.06