

Improving performance of collaborative recommender system using combination of learning techniques

Abstract

As the World Wide Web continues to grow at an exponential rate, the size and complexity of many web sites grow along with it. For the users of these web sites it becomes increasingly difficult and time consuming to find the information they are looking for. User interfaces could help users find the information that is in accordance with their interests by personalizing a web site. Recommender systems provide personalized information by learning the user's interests from traces of interaction with that user. We claim that our method performs better than its existing counterparts. We have performed our experiment on book recommendation dataset available online. Our method makes use of k- means clustering technique of machine learning with some variation.

Keywords: Recommender System, Silhouette index, K-means clustering

Volume 5 Issue 2 - 2019

Nitin Mishra

Data Scientist and Trainer at Excel Solutions, Annamalai University, India

Correspondence: Nitin Mishra, Data Scientist and Trainer at Excel Solutions, Annamalai University, India, Email nitin.mishra@rgmcet.edu.in

Received: January 04, 2019 | **Published:** April 02, 2019

Introduction

It has been found that more than 1 billion peoples are surfing the website these days. From these users, many people's read the books online. Recommender systems are used to recommend the books online. Recommender system collects the information of the books according to user preference. The user preference can be gathered by either explicitly or implicitly.¹ Explicitly collection of information has been done by analyzing the user's behaviour and implicitly collection of information is done by previous ratings. Collaborative information filtering is done using product's previous history.² First it collects the information about the ratings of the textbooks given by peoples and then recommendation is done by people's previous interest and tastes. Clustering techniques are the unsupervised machine learning concept used to partition the data with the help of various similar metrics.³ Searching for some useful data from a pool of huge dataset is sometimes become very trivial. When the amount of data is very vast is of different types then that type of data is known as Big Data to secure this type of data is such a tough task. Biometric systems⁴ are used to secure this data. We also have this kind of data in healthcare,⁵ as in recommender system we are recommending something to someone so that his technique can be useful.

The collaborative recommendation system is explained in next section. The silhouette index is used to improve the accuracy of the recommender system. The k-means clustering technique is found to be better as compare to other state-of-the-arts techniques with respect to time and complexity.⁶ An index when incorporated with k-means clustering gives optimizes results. Silhouette index seems to be more suitable technique to improve the recommendation results accuracy to some extent. It also takes less time as compared to other techniques. The proposed techniques used both silhouette index and k-means clustering technique to get better the results in terms of mean absolute error (MSE), root mean square error (RMSE) and standard deviation (SD). The textbook recommender system is evaluated in terms of these parameters. The experimental results also show the better performance of the recommender systems in terms of reliability and accuracy when compared with other cluster based recommendation techniques. Designing expert and efficient recommendation system is still challenge for new researchers and very difficult to design efficient

clustering technique. To overcome the said problems, we proposed technique which is the combination of silhouette index and k-means clustering techniques. The following contributions are found in our research work.

- We introduced a recommender system using combination of silhouette index and k-means clustering technique.
- Our research is very efficient to recommend textbooks.
- Our framework has 0.63 MAE which is better than existing 0.68 MAE.
- The processing time is also minimizing.

The rest of the paper is organized in such a way that: Next section discussed the various state of arts techniques in the form of related works, Proposed system has been discussed in section 3, Section 4 contains the experiments and results and finally conclusion has been discussed in section 5.

Related works

A recommendation system is used to support information filtering of products and services by analysis of suggestions given by other users.⁷ Collaborative recommendation system is a type of recommendation system based on previous data. Now a day's recommendation system works on individual users not for group activities.⁸ In investment and shopping recommendation system, all the data is taken individually and prediction is very easy but in other recommendation system like movies, products and restaurants, analysis of data is so difficult. The group recommendation systems are mainly based on off-line environments. Various group activities are commonly processed in virtual space these days. The technique of solving group suggestions conflicts are different from off-line environments. Merging of individual suggestion into group suggestion is some time results in the satisfaction of majority of groups but against the minority group.

A group recommendation system proposes the product to group of peoples. Only few researches have been done so far. Mccarthy et al.^{9,10} proposed a framework to select stations to group of peoples working on gym called MusicFx.⁹ Gym members rate all the stations before recommendation. MusicFx plays the stations with higher ratings. The

framework is used to satisfy the groups working on gym. The most important point of that framework is that all the members have to rate the stations in advance. This system might work for small number of stations but it will not properly work in the situation where large number of books is available in library. Oconnor et al.⁸ introduced a movie recommender system called PolyLens which recommends the movies to small group of peoples.¹⁰ They combined the individual recommendation of each member of the group to form group recommendation. PolyLens is able to satisfy all the requirements of the group members as compared to MusicFx. The basis of PolyLens is to satisfy the requirements of least satisfied group member. Hence, a movie which is not good is less accepted to group members but recommended to one member that not likes it but everyone else likes it. Mc Carthy et al.^{9,10} proposed a framework which recommends the restaurant to group of peoples based on location and package called Pocket Restaurant Finder.¹¹ Peoples fill their requirement such as how far they can go, how much they spent, what type of food they wanted, what type of restaurants they wanted, what type of cuisines they wanted etc. Pocket Restaurant Finder combines the group members according to their preferences and then recommends accordingly.

Chao et al.¹¹ introduced a framework that plays a song to a group of listeners having common environments called Adaptive Radio.¹² Music recommendation systems mainly focus on the user preference based on the survey and on their listening habits but Adaptive Radio focus on the dislike songs and ignoring the songs completely. They used negative preference of the songs. Adaptive Radio plays a song which is not familiar and less popular among users.

Yu et al.¹² proposed a framework for in-vehicle multimedia recommendation system which provides multimedia contents for traveler.¹³ It connects all the users using wireless technology. The process of combining user profile consists of two steps: First step is to find the important features of the similar traveller's interests. Second step is to find weights on important features. It is used to satisfy the group member's majority.

Zhou et al.¹³ introduced a framework for recommendation of TV programs to group of users called TV4M.¹⁴ They used the concept of total distance minimization of feature space. The user profile is represented by features with weights according to their importance. They create a subset from the features to represent common interests. Then find the weights for its selected features. They also showed experimentally that TV4M works better for closer member of the group and with similar interests.

Crossen et al.¹⁴ proposed a framework for recommendation of music to group of members based on votes.¹⁵ Ardissono et al.¹⁵ introduced a framework for travel recommendation system and able to solve conflicts between members of the groups.¹⁶

Mishra et al.¹⁷ mentioned various Data Mining and Machine learning techniques¹⁷ through which help one can easily fetch the required details which are very helpful in any recommendation system.

Katarya et al.¹⁸ proposed a novel movie-based collaborative recommender system which utilizes the bio-inspired gray wolf optimizer algorithm and fuzzy c-mean (FCM).¹⁸ clustering technique and predicts rating of a movie for a particular user based on his historical data and similarity of users. Gray wolf optimizer algorithm was applied on the movie lens dataset to obtain the initial clusters, and also the initial positions of clusters are obtained. FCM is used to classify the users in the dataset by similarity of user ratings.

Katarya et al.¹⁹ has proposed another novel recommender system which makes use of k-means clustering by adopting cuckoo search optimization algorithm¹⁹ applied on the movie lens dataset.

Objective of our research

We have discussed many different types of methods in related work section. Our proposed method adds a different approach to solving the problem in recommender systems? Our method can be applicable in domain where other methods may not be applied. The performance of our method is also very comparable to other effective methods. Our method will also give an option to implement the cold start problem solution.²⁰

Our approach

The collaborative recommendation system (CRS) is very popular recommendation system widely used in various applications. The three main steps which are involved in CRS are²¹ represented by following diagram which demonstrate the flow of our proposed model.²² The description of each steps are mentioned below. (Figure 1)

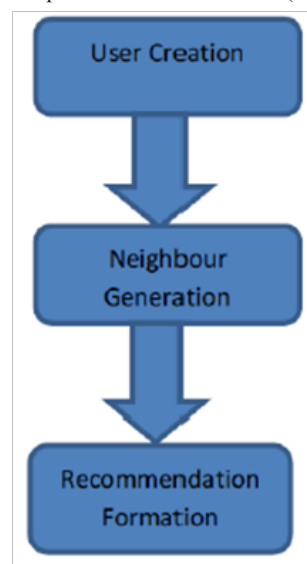


Figure 1 Framework for proposed system.

User creation: User creation, the main step in any recommender system is done using historical data and information regarding ratings which can be easily available on web.

Neighbour generation: As we know Machine Learning is an important area which guarantees more accurate results and these machine learning techniques are applied to get the set of users called neighbour from the past or who had similar behaviours.

Recommendation formation: Once a neighbourhood is formed for the user. CRS has formed a set of products that the user most likely to interested by analyzing the products.

In Figure 1, the flow of our proposed recommender system in which we used BX- Book- Ratings dataset for user creation has been introduced, further silhouette index is used to calculate the initial positions of the clusters, and then k-means clustering is used for the classification of the users in BX-Book-Ratings dataset by using similar user ratings as this clustering technique provides more accurate results as compare to other existing techniques. The final recommendation for the books is provided to the users.

Experiments and results

We are using the BX-Book-Ratings dataset contains the book rating information. Ratings are done in the scale from 1 to 10.⁵ We introduced a combine technique of silhouette index and k-means clustering. The dataset is divided into 70% for training and 30% for testing. Performance of our framework is checked using various parameters such as Root Mean Squared Error (RMSE), Standard Deviation (SD), Precision and Mean Absolute Error (MAE). Participants are using various numbers of recommended books and choose one book from them. We considered the varying of books recommendations.

Mean Absolute Error (MAE)

The mean absolute error for the Bx-Books Ratings is given by

$$MAE = \frac{\sum |P_{ij} - T_{ij}|}{X}$$

Where X is the total number of Textbooks, P_{ij} is the prediction of user I on textbook J and T_{ij} is the true prediction of books. The calculation of MAE has been for different values of clusters ranges from 6 to 46 as depicted in Figure 2. From the Figure 2, it can be observe that increasing the cluster size will results in the decrease of the mean Absoluter Error.

S.No.	No. of Clusters	SD
1	6	.2242582
2	10	.1812353
3	18	.1356231
4	26	.1263289
5	32	.1238030
6	38	.1293452
7	46	.1286983

Figure 2 SD for the clusters.

Standard deviation

The SD values are obtained using the publically available books datasets called BX-Book- Ratings. The no. of clusters increases then the SD values decreases. This indicates that changes between 10-18 is high as compared to the 18-26 is very high which indicates that adding more clusters makes great change in SD values in starting but later on it will decrease slowly.

Precision

For evaluation of performance of our recommendation system precision is the best metrics. It is one of the most popular metrics used in retrieval systems. Precision measuring correctness of recommendation is defined as the ratio of the number of selected items to the number of recommended items:

Precision is the probability of choosing a book from the recommendation systems. In Figure 3, it can be observed that increasing the no. of clusters decrease the precision values. Comparison of Metrics with various techniques: The comparisons of our proposed framework with other state-of-the-art techniques are depicted in Figure 4. Every value is achieved using 46 clusters. The PCA-SOM technique is based on principal component analysis and self-organizing maps. PCA is used to reduce the features and

SOM converts high input space into low input space with the mean of 0.98 and SD of 0.07. SOM-cluster, UPCC, K-means cluster, PCA with k-means, Genetics algorithm based recommendation system has mean of 0.75, 0.81, 0.69, 0.93, 0.76 respectively. PCA with genetics algorithm and k-means cuckoo has 0.98 and 0.68.

S.No.	No. of Clusters	Precision
1	6	.593214
2	10	.475689
3	18	.423563
4	26	.398032
5	32	.379860
6	38	.354798
7	46	.339807

Figure 3 Precision values for the clusters.

In Figure 4, it can also be observed that our proposed framework better than other state- of-the- art technique with mean of 0.63 and SD of 0.08. All experiments are done using 46 clusters in i3 processor with 4 GB RAM with R.

Technique	Mean	Standard Deviation
PCA-SOM	.98	.07
SOM-Cluster	.75	.06
UPCC	.81	.11
k-Means Cluster	.69	.10
PCA-k Means	.93	.12
GAKM-Cluster	.76	.05
PCA-GAKM	.98	.17
K-means Cuckoo	.68	.10
Proposed	.63	.08

Figure 4 Comparison of various techniques.

Conclusion

In this paper, combination of silhouette index and k-means clustering techniques are introduced to enhance the performance of the textbook recommendation system. We also tested our framework using metrics like RMSE, SD, MAE, Precision and found the MAE of 0.63 and improve the accuracy and reliability of our work to great extent. Our proposed framework work better with 46 clusters and in future we will increase the size of the clusters and improve the recommendation results. We will also incorporate new feature extraction technique to improve the accuracy and efficiency.

Acknowledgments

None.

Conflicts of interest

Author declare that there is no conflict of interest.

References

1. Jie Lu, Dianshuang Wu, Mingsong Mao, et al. Recommender system application developments: a survey. *Decision Support Systems*. 2015;74:12–32.

2. Isinkaye FO, Folajimi YO, Ojokoh BA. Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*. 2015;16(3):261–273.
3. John A Hartigan, Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society*. 1979;28(1):100–108.
4. Nalini K Ratha, Jonathan H Connell, Ruud M Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM systems Journal*. 2001;40(3):614–634.
5. Vikash Yadav, Monika Verma, Vandana Dixit Kaushik. Big data analytics for health systems. In *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*. 2015. p. 253–258.
6. Tapas Kanungo, David M Mount, Nathan S Netanyahu, et al. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. 2002;7:881–892.
7. Enrique Frias-Martinez, Sherry Y Chen, Xiaohui Liu. Evaluation of a personalized digital library based on cognitive styles: Adaptivity vs. adaptability. *International Journal of Information Management*. 2009;29(1):48–56.
8. Mark O’connor, Dan Cosley, Joseph A Konstan, et al. Polylens: a recommender system for groups of users. *ECSCW*. 2001. p. 199–218.
9. Joseph F Mccarthy, Theodore D Anagnost. Musicfx: an arbiter of group preferences for computer supported collaborative workouts. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*. 1998. p. 363–372.
10. Joseph F McCarthy. Pocket restaurant finder: A situated recommender system for groups. In *Workshop on Mobile Ad-Hoc Communication at the 2002 ACM Conference on Human Factors in Computer Systems*. 2002.
11. Dennis L Chao, Justin Balthrop, Stephanie Forrest. Adaptive radio: achieving consensus using negative preferences. In *Proceedings of the 2005 international ACM SIGGROUP conference on supporting group work*. 2005. p. 120–123.
12. Yu Zhiwen, Zhou Xingshe, Zhang Daqing. An adaptive in-vehicle multimedia recommender for group users. *2005 IEEE 61st Vehicular Technology Conference*. 2005;5:2800–2804.
13. Zhiwen Yu, Xingshe Zhou, Yanbin Hao, et al. Tv program recommendation for multiple viewers based on user profile merging. *User modeling and user-adapted interaction*. 2006;16(1):63–82.
14. Andrew Crossen, Jay Budzik, Kristian J Hammond. Flytrap: intelligent group music recommendation. In *Proceedings of the 7th international conference on intelligent user interfaces*. 2002. p. 184–185.
15. Liliana Ardissono, Anna Goy, Giovanna Petrone, et al. Tailoring the recommendation of tourist information to heterogeneous user groups. In *Workshop on adaptive hypermedia*. 2001. p. 280–295.
16. Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, et al. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*. 2005. p. 22–32.
17. Nitin Mishra, Vimal Mishra, Saumya Chaturvedi. Tools and techniques for solving cold start recommendation. In *Proceedings of the 1st International Conference on Internet of Things and Machine Learning*. 2017;11:1–11.
18. Rahul Katarya, Om Prakash Verma. Recommender system with grey wolf optimizer and fcm. *Neural Computing and Applications*. 2018;30(5):1679–1687.
19. Rahul Katary, Om Prakash Verma. An effective collaborative movie recommender system with cuckoo search. *Egyptian Informatics Journal*. 2017;18(2):105–112.
20. Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, et al. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*. 2005. p. 22–32.
21. Badrul Munir Sarwar. Sparsity, scalability, and distribution in recommender systems. University of Minnesota. 2001.
22. Rahul Srivastava, Nitin mishra, Saumya Chaturvedi. Research challenges in recommendation system. In *Proceedings of the CONIAPS*. 2014. p. 39–46.