

# Comparative Study of Hybrid and Context Based Approach for Recommender System

Kimi Bhima Ramteke

Department of Computer Engineering  
Pune Institute of Computer Technology (PICT)  
Pune, Maharashtra, India

Arati R. Deshpande

Department of Computer Engineering  
Pune Institute of Computer Technology (PICT)  
Pune, Maharashtra, India

**Abstract**— Day by day user & item information is increasing in the recommender system. Hence, traditional recommender systems (RS) becoming inefficient to give the personalized recommendation. Hybrid technique is trying to take the advantages and to minimize the disadvantages of traditional ones. Recommender system has taken the place of agents, guides, friends due to its vast world wide information and less time for recommendation. In this paper, we proposed the system having hybrid technique with contextual approach. In hybrid technique, we used the combination of clustering and association rules whereas, user's demographic information have been taken as the contextual approach. Finally the performances of 'hybrid RS with context' with 'hybrid RS without context' is compared.

**Keywords**— Recommender system, data mining, clustering, classification and association algorithm, hybrid technique.

## I. INTRODUCTION

Today recommender systems have taken the place of different recommending agents like tourist agents recommending tourist points, friends that share their experience or opinions. Recommender systems are chosen due to their vast world wide information, minimum time of recommendations, personalized recommendation which is easily available at your e-device. Though the people are using it, there is need of continues improvement in prediction of users likings.

There are well known recommender systems techniques exist like content based filtering, collaborative filtering and combination of both techniques called hybrid approach. This paper uses hybrid approach to remove the drawbacks of first two techniques in some extent like sparsity, new user or new item problem i.e. cold start, grey sheep problems. It uses model based collaborative filtering through building user model using bisecting k-means clustering algorithm and classification based association for building association rules offline. Uses content based filtering for prediction of items for personalized recommendation. In addition to this uses contextual post filtering for improving the accuracy of recommendation to the user.

The paper is organized as follows. Section II gives the review of recommendation system techniques, section III presents the working of proposed system through its architecture, mathematical model. Section IV discusses experimental results and finally paper is concluded in section V.

## II. RELATED WORK

Recommender system techniques have been classified into collaborative filtering (CF) and content based filtering. Collaborative filtering uses the other user choices, whereas content based filtering uses other purchased items records for recommendation purpose. Collaborative filtering has some drawbacks like grey sheep, early rater, scalability, sparsity. Content based method is not used solely. In collaborative filtering memory based methods takes large data history, large memory and takes much time to recommend. Model based methods build offline model, hence takes little time to recommend.

According to [2] item based approach gives better prediction quality and the performance. Incorporating data mining into recommendation systems (RS) like rule induction, neural networks and Bayesian networks which builds model based method and produces the results as accurate as CF given by [8] and [15]. For the clustering purpose [12] has concluded that bisecting k-means is the best for clustering the documents. [5] and [9] has used k-means for clustering the user for tourism application. According to [1], [3], and [14] association rules may produce prediction module so as to minimize the prediction time at the time of recommendation in RS. Classification based on association yields better result than other classification techniques commented in [1], [13], [14]. As per [5], [9], [10] hybrid technique with different combination may affect recommendation quality where they proposed combination of clustering technique and association rules mining. Fuzzy logic has been used by [5] & [9] through fuzzy clustering (k-means) and fuzzyfied association rules respectively. According to [4] and [7] incorporating contextual data may yield better personalized RS than any other pure traditional RS methods. They have proposed three types of context filtering i.e. pre-filtering, post filtering and combination of the two in cascaded manner. In [11] different evaluation measures have been studied like precision, recall, MAE, RMSE etc. For our system precision and recall are used for recommendations.

## III. PROPOSED SYSTEM

### A. Architecture

The architecture of the proposed system shows the flow of working. Architecture consists of mainly two modules on-line module and off-line module. When user logs in to the system,

user's cluster-ID is identified. Then for that cluster CBA rules are identified according to his last transaction. Then items are extracted from those rules and items are filtered out using contextual information. Finally filtered items are recommended to the user. The architecture is shown as below.

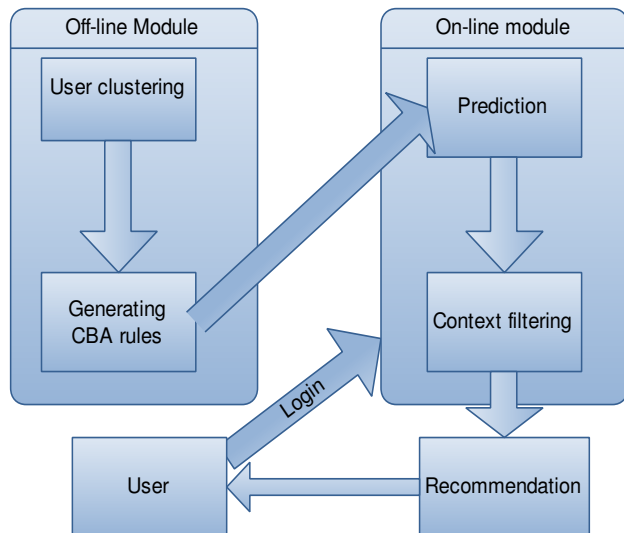


Fig. 1 Architecture showing the working of proposed system

### B. User Clustering

In offline module all users profile consisting of their age, gender, marital status, etc. are taken as input and their clustering is done using bisecting k-means method. Bisecting k-means proved best over k-means by [10], [12]. Better clustering of users would result into better recommendation to the user. Steps for bisecting k-means are given as below.

#### Basic Bisecting K-means Algorithm for finding $K$ clusters.

1. Pick a cluster to split.
2. Find 2 sub-clusters using the basic K-means algorithm. (Bisecting step)
3. Repeat step 2, the bisecting step, for ITER times and take the split that produces the clustering with the highest overall similarity.
4. Repeat steps 1, 2 and 3 until the desired number of clusters is reached.

Similarity between the two users is measured by cosine measure. Let similarity between user profiles  $u_1$  and  $u_2$  is given by,

$$\text{cosine}(u_1, u_2) = \frac{(u_1 * u_2)}{(|u_1| |u_2|)} \quad (1)$$

The frequency of each term in profile is calculated through Term Frequency–Inverse Document Frequency (TF-IDF). TF gives frequency of each term in that profile where as IDF gives frequency in all profiles. Let  $t$  be the term in user profile  $p$  and  $P$  be the total number of user profiles. Then TF-IDF for the term  $t$  is given by  $w_{t,p}$ ,

$$w_{t,p} = (1 + \log f_{t,p}) * \log_{10} (P/pf_t) \quad (2)$$

Where,  $f_{t,p}$  be the frequency of  $t$  in profile  $p$ .

### C. Generating Association Rules

Let  $D$  be the dataset consists of user profile, cluster ID and age-band over which CBA method is applied. Let  $I$  be the set of items and  $T$  be the set of transactions. Each transaction  $t$  in  $T$  is subset of  $I$ . Then association rule mined from  $D$  would be in the form  $X \rightarrow Y$ , where  $X, Y \subseteq I$  and  $X \cap Y = \phi$ . Let  $s$  be the support of the rule which is percentage of transactions in  $D$  that contains both  $X$  and  $Y$ . Confidence  $c$  is out of all transactions that contains  $X$  and percent that contains  $Y$  as well [16]. Support and Confidence is calculated by following formulas:

$$S_{x,y} = \frac{|T_{x,y}|}{|T|} \quad (3)$$

$$C_{x,y} = \frac{S_{x,y}}{S_x} \quad (4)$$

where  $S_{x,y}$  be the support of  $x$  &  $y$ .  $T_{x,y}$  be the transaction containing items  $x$  &  $y$  both.  $C_{x,y}$  be the confidence of  $x$  &  $y$ . Class Based Association (CBA) has two parts rule generator (CBA-RG) which incorporates apriori method to generate the association rules and classifier builder (CBA-CB) generates classifier from the rules generated by CBA-RG. Here we select small set of rules to form classifier. In our case class classifier is the cluster ID through which the rule with highest confidence is selected for prediction purpose. Algorithms for CBA-RG and CBA-CB are given in the figure 2 & 3 respectively.

Algorithm	Generating Association Rules
Inputs:	UpperBoundMinSupport, LowerBoundMinSupport, delta, minNumberOfRules, minConfidence
Output:	rules
1.	rules = $\emptyset$ ;
2.	freqItemsets = $\emptyset$ ;
3.	support = UpperBoundSupport;
4.	while (support $\geq$ LowerBoundSupport AND rules.size < minNumberOfRules) do
5.	$L_1 = \{1\text{-item itemsets}\}$ ;
6.	for ( $k = 2; L_{k-1} \neq \emptyset$ ) do
7.	$C_k = \text{generateCandidates}(L_{k-1})$ ;
8.	$L_k = \text{evaluateCandidates}(C_k)$ ;
9.	freqItemsets $\cup L(k)$ ;
10.	end for
11.	maxFreqItemsets = genMaxFreqItemset(freqItemsets);
12.	rules = GenerateAllRules(maxFreqItemsets, minConfidence);
13.	support = support - delta;
14.	freqItemsets = $\emptyset$ ;
15.	end while
16.	return rules;

Fig. 2 shows steps for CBA rule generator algorithm

### D. Prediction and Context filtering

For the prediction purpose, there are two cases one for existing and another for new user. Our aim is to recommend the items to both types of users. For the existing user, we match his last transaction  $T$  with the rules  $R$  and obtained subset  $R'$  of  $R$ . In the proposed system post filtering of context is done. The context is nothing but the demographic information of user like profile as mentioned before. Then context filtering is done by filtering out the rules with user's age-band  $A$  and last seen item-type  $IT$ . We consider 4 age-bands and 5 item-types in our dataset. Then the prediction is

done by extracting the items of the rules  $R'$ . For the new user, items are recommended based on his profile by finding best cluster for him.

```

Algorithm CBA-CB
Inputs: rules R, training set instances D
Output: classifier C
1. R = sort(R);
2. for each rule r in R in sequence do
3.   temp = 0
4.   for each instance d in D do
5.     if d satisfies the conditions of r then
6.       store d.id in temp and mark r if it correctly classifies d;
7.     end if
8.   end for
9.   if r is marked then
10.    insert r at the end of C;
11.    delete all the cases with the ids in temp from D;
12.    select the default class for the current C;
13.    compute the total number of errors of C;
14.   end if
15. end for
16. Find the first rule p in C such that Cp, the list of rules in C up to p, has the lowest total number of errors. and drop all the rules.
17. Add the default class associated with p to the end of C, and return C

```

Fig. 3 shows steps for CBA classifier builder algorithm

### E. Recommendation

When old or existing user logs in to the system, system follows the steps mentioned over. Topmost rules are selected from his cluster and then top 10 non-repeated items are recommended to that user. Though in the case of new user with no last transaction but system is able to recommend items to him. So here we have succeeded to overcome cold start problem for both user and items. For the new user using his profile most nearest cluster is chosen and according to the topmost rules satisfying pre-defined support and confidence, top 10 items are extracted and recommended to the user.

### F. Evaluation

For evaluation purpose, we are taking users last transaction as actual value and matching it across predicted value. Prediction is done on the basis of user's transaction without last one. Precision and recall are two measures used to give quality and quantity of recommendation respectively [11]. We consider quality of recommendation for our experiment to be fruitful. Precision and recall are given by following two formulas.

$$\text{Precision} = \frac{\text{no. of correct recommendations}}{\text{total no. of recommendations}}$$

$$\text{Recall} = \frac{\text{no. of correct recommendations}}{\text{total no. of correct recommendations}}$$

### G. Experimental Results

For experimental purpose, the system has been evaluated with 20 users, 500 items and 2000 transactions. Accuracy, precision and recall are used for performance evaluation of the proposed recommender system. Precision is used to measure how many are correct out of total forecasts for RS. Recall is the measure used to ensure quantity of total correct forecasts.

Experiment has been performed over 19:1 cross-fold for different numbers of users. We are not using rating because rated items found to be very less; hence it may limit the recommended items to the user. In this experiment we are comparing the performance of hybrid approach with our system i.e. Hybrid + context approach on the same over-mentioned dataset. Experimental results for precision and recall are shown through the graphs no. 4 & 5 respectively.

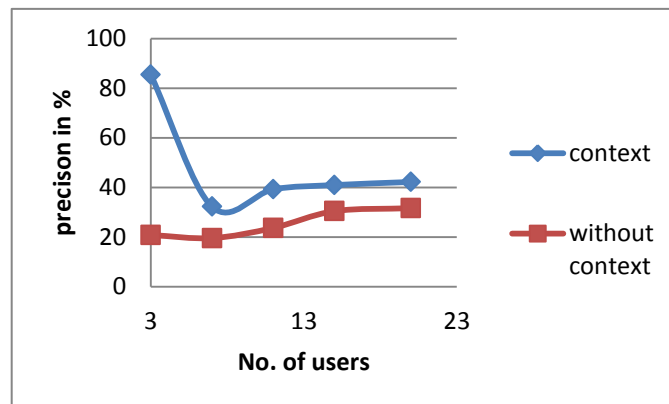


Fig. 4 Graph showing comparison between context & without context RS for Precision values for different no. of users

Experimental results show that precision and recall values of RS with context are always greater than that of RS without context. In the experimental results it has been observed that, for users with consistent choices Context based RS gives better recommendation whereas, for users with random choices RS without Context gives better recommendation.

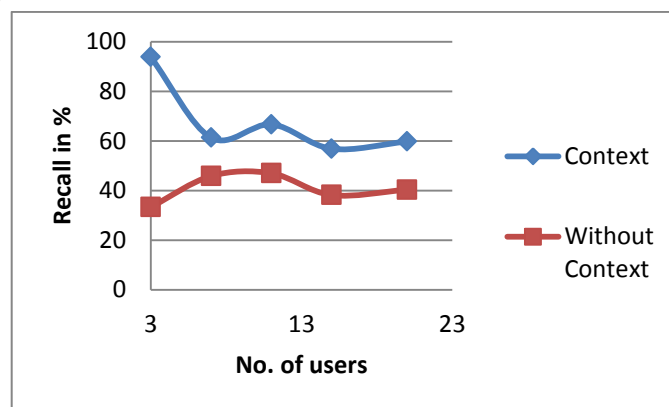


Fig. 5 Graph showing comparison between context & without context RS for Recall values for different no. of users

## IV. CONCLUSIONS

Clustering users results into good recommendations. Bisecting proves better over K-means. Different combinations of clustering and classification must be tried for improvement in RS. In our experiments recall is greater than precision means proposed system is returning most of the relevant items to the user. Hence proposed system with context gives better personalized recommendations than Recommender system without context. Explicit

feedback/opinion does not guarantees good prediction, hence by taking Implicit feedback in future from the user may guarantee prediction of users right choices, which may results into perfect recommendation.

#### ACKNOWLEDGMENT

I take this opportunity to express my deep sense of gratitude towards my esteemed guide Arati R. Deshpande for giving me this splendid opportunity to select and present this topic. I wish to express my thanks to our HOD Prof. G.P. Potdar for encouragement & providing me with the best facilities for my preliminary dissertation work. I thank all the staff members, for their indispensable support, priceless suggestions and for most valuable time lent as and when required. I also thank my parents for their continuous support. I thank my friends for their help in collecting information without which this research paper not have seen the light of the day.

#### REFERENCES

- [1] A. Alsalama. "A Hybrid recommendation system based on association rules", digitalcommons 2013.
- [2] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms", May 1-5, 2001, Hong Kong. ACM 1-58113-348-0/01/0005.
- [3] F. Muhlenbach and R. Rakotomalala, "Discretization of continuous attributes", hal-00383757, version 2 - 13 May 2009.
- [4] F. Ullah, G. Sarwar, "Hybrid recommender system with temporal information", ICOIN ©2012 IEEE 978-1-46730250-0/12.
- [5] G. Fenza, E. Fischetti, D. Furno, V. Loia "A hybrid context aware system for tourist guidance based on collaborative filtering ", IEEE 2011.
- [6] G. Adomavicius and A. Tuzhilin, "Towards the next generation of recommender systems: A Survey of the State-of-the-Art and Possible Extensions", IEEE transaction on KADE 2005.
- [7] G.s Adomavicius, A. Tuzhilin, "Context-aware recommender systems", IEEE 2011.
- [8] J. Ben Schafer "The Application of data-mining to recommender systems", University of Northern Iowa, 2004.
- [9] J. P. Lucas, N. Luz, María N. Moreno, R. Anacleto, A. Almeida Figueiredo, C. Martins, "A Hybrid recommendation approach for a tourism system" Expert Systems with Applications, Elsevier 2013.
- [10] K. Ramteke, A. Deshpande, "Context based recommender system using hybrid approach", Third post graduate conference for computer engineering (CPGCON), March-2014.
- [11] M. Mortensen, "Design and evaluation of a recommender system", Faculty of Science Department of Computer Science University of Troms, February 5, 2007.
- [12] M. Steinbach, G. Karypis, V. Kumar, "A Comparison of document clustering techniques", KDD workshop on clustering, 2000.
- [13] S. Gambhir, N. Gondliya, "A Survey of associative classification algorithms", International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 9, November – 2012.
- [14] P. Cremonesi, R. Turrin, E. Lentini and M. Matteucci, "An Evaluation methodology for collaborative recommender systems", Contentwise, Whitepaper March 2010.
- [15] X. Amatriain, A. Jaimes, N. Oliver, and J. M. Pujol, "Data mining methods for recommender systems", Recommender Systems Handbook, © Springer Science+Business Media, LLC 2011.