Development of Context Based Collaborative Filtering System for Recommendation

[1]Mr.Rohit M.Pawar, [2]Mr.V.V.Bag
PG Scholar, Associate Professor
Department of Computer Science and Engineering,
N.K. Orchid College of Engineering and Technology, Solapur, India.

Abstract **- Recommender systems suggest items to users by utilizing the techniques of Collaborative filtering based on historical records of items that users have purchased. Recommender systems make use of data mining techniques to determine the similarity among a huge collection of data items, by analysing historical user data and then extracting hidden useful information or patterns. Goal of Collaborative filtering is finding the relationships among the individuals and the existing data items in order to further determine the similarity and provide recommendations. This paper, proposes the Context based Collaborative Filtering Recommender System, which can be used for any commercial online-marketing. Experimental evaluation of results and comparing them with traditional collaborative filtering approach, concludes that context based collaborative approach provide dramatically better performance than traditional-based algorithms, while at the same time providing better recommendation as per customer point of view.**

Keywords **- Collaborative filtering, Context based collaborative filtering, Clustering, Recommender, User-Context**

I. INTRODUCTION

In everyday life, people rely on recommendations from other people by spoken words, reference letters, and news reports from news papers, general surveys, travel guides, and so on. For online users the recommendation systems are available. With the recent explosive growth of e-marketing, recommendation system has been accepted by large no of users. The goal of a Recommender System is to generate meaningful recommendations to a collection of users for items or products that might interest to them. With the explosive growth of the web,72% of consumers trust Online reviews as much as Personal Recommendations[2].Recommendation systems are a subclass of information filtering system that aims to predict the 'rating' or 'preference' that user would give to an item. Suggestions for books on website like Amazon, or movies on Netflix, are real world examples of the operation of recommendation systems. One approach to the design of recommender systems that has been seen widely used is collaborative filtering.

Collaborative filtering methods are based on collecting and analyzing a large amount of information on users' behaviors, activities or preferences and predicting what users will like based on their similarity to other users. Advantage of the collaborative filtering approach is that it does not rely on machine analyzable content. Many algorithms have been used in measuring user similarity or item similarity in recommender systems. For example, the k-nearest neighbors (k-NN) approach.

Collaborative Filtering (CF) is the most popular approach used in recommendation systems [1,8,9].Now a day recommendation systems have been widely accepted by users [1,3]. Personalized recommendation approaches have gained great momentum both in the commercial and research areas [4,5].Collaborative filtering is a method of making automatic predictions about the interests of a user by collecting preferences or taste information from many users (collaborating).

The motivation for collaborative filtering comes from the idea that people often get the best recommendations from someone with similar choices to themselves. Collaborative filtering explores techniques for matching people with similar interests and making recommendations on this basis.

Collaborative filtering approach often requires:

1. User's active participation
2. An easy way to represent users interests to the system
3. Techniques those are able to match people with similar interests

Typically, the workflow of a collaborative filtering system is:

1. A user expresses his or her preferences by rating or selecting items (e.g. books, stocks, movies or CDs) of the system.
2. Approximate representation of the user's interest in the corresponding domain.
3. The system matches this user's weight against other users and finds the people with most "similar" tastes.
4. With similar users, the system recommends items that the similar users have rated highly but not yet being rated by this user.

Even though the Collaborative filtering approach has the advantages, it has some drawbacks such as low prediction accuracy, scalability etc. To overcome these disadvantages, it needs improvements. For improvement the Context based filtering approach is suggested. In this paper a new approach i.e. Context based collaborative filtering is introduced. Context is a broad concept that

can be studied across different research disciplines, including computer science. Since context has been studied in multiple disciplines, each discipline tends to take its own view that is somewhat different from other disciplines.

The underlying assumption of this approach is that if a entity A has the same opinion with set of entity E on an issue, A is more likely to have set of E's opinion on a different issue *x* than to have the opinion on x of a person chosen randomly. For example, a Context collaborative filtering recommendation system for television could make predictions about which television show a user should like from given a partial list of that user's tastes. For implementations of context based approach selection of context or user attributes is important aspects as it directly affects the accuracy of recommendation.

II. LITERATURE REVIEW

Recommender systems assist and augment the natural social process to help people look through available books, articles, web pages, movies, music, restaurants, jokes, grocery products, and so forth to find the most interesting and valuable information for them. Recommendation algorithms often operate in a challenging environment, especially for large online shopping companies like eBay and Amazon. Some popular systems are available in literature.

Two popular music recommender systems: Last.fm and Pandora Radio.

- Pandora uses the properties of a song or artist a subset of the 400 attributes provided by the Music Genome Project in order to seed a "station" that plays music with similar properties. User feedback is used to refine the station's results, deemphasizing certain attributes when a user "dislikes" a particular song and emphasizing other attributes when a user "likes" a song. This is an example of a content-based approach.
- Last.fm creates a "station" of recommended songs by observing what bands and individual tracks that the user has listened to on a regular basis and comparing those against the listening behavior of other users. Last.fm will play tracks that do not appear in the user's library, but are often played by other users with similar interests. As this approach leverages the behavior of users, it is an example of a collaborative filtering technique.

Each type of system has its own strengths and weaknesses. In the above example, Last.fm requires a large amount of information on a user in order to make accurate recommendations. While Pandora needs very little information to get started, it is far more limited in scope (for example, it can only make recommendations that are similar to the original seed). The developers of one of the recommendation system, Tapestry[1],generates the phrase "collaborative filtering (CF)," which has been widely adopted regardless of the facts that recommenders may not explicitly collaborate with recipients and recommendations may suggest particularly interesting items, in addition to indicating those that should be filtered out [3]. The fundamental assumption of CF is that if users rates items similarly or have similar behaviors (e.g., buying, watching, listening), and hence will rate or act on other items similarly [11].Item-based Collaborative Filtering (CF) is one of the popular approach for determining the recommendations A common problem of current item-based CF approaches is that all users have the same weight. To improve the quality of recommendations, there is incorporation of the weight of a user, into the computation of item similarities and differentials [5].

A key problem of collaborative filtering is how to combine and weight the preferences of user neighbours. Sometimes, users can immediately rate the recommended items. As a result, the system gains an increasingly accurate representation of user preferences over time.

Currently, item-based collaborative filtering approaches had some problem related to produce accurate recommendation within particular time span in a System. In this, analysis of the ranking of users and prediction of items by recommendation system, experimental results show that user context information helps to improve the prediction results and the stability of the recommendation system.

The majority of existing approaches to recommender systems focus on recommending the most relevant items to individual users and do not take into consideration any contextual information, such as time, place and the company of other people (e.g. for watching movies or dining out). In other words, traditionally recommender system deal with applications having only two types of entities, users and items, and do not put them into a user context when providing recommendations.

However, in many applications, such as recommending a vacation package, personalized content on a Web site, or a movie, it may not be sufficient to consider only users and items it is also important to include a contextual information into the recommendation process in order to recommend items to users in certain situations. For example, using the temporal context, a travel recommender system would provide a vacation recommendation in the winter that can be very different from the one in the summer. Similarly, in the case of personalized content delivery on a Web site, it is important to determine what content needs to be delivered (recommended) to a customer and when. More specifically, on weekdays a user might prefer to read world news when he/she logs on in the morning and the stock market report in the evening, and on weekends to read movie reviews and do shopping.

From the literature survey, it seems that existing recommendation system mainly focused on Customer (user) and product (items) only. Existing system needs improvement in the areas mainly of predictions accuracy. Hence context based collaborative system is developed. The methodology used to develop the system is presented as given below.

III. METHODOLOGY

The general concept of context is very broad, hence, in this work, the attempt is made to focus on those fields that are directly related to recommender systems. To generate higher accuracy in recommendations, obtaining user contextual information is a measure task. The contextual information can be obtained in the following ways:

- Explicitly-by directly approaching relevant people and other sources of contextual information and explicitly gathering this information either by asking direct questions. For example, a website may obtain contextual information by asking a person to fill out a web form or to answer some specific questions before providing access to certain web pages.

- Implicitly from the data or the environment, such as a change in position of the user detected by a mobile telephone company. Alternatively, temporal contextual information can be implicitly obtained from the timestamp of a transaction. Nothing needs to be done in these situations in terms of interacting with the user or other sources of contextual information the source of the implicit contextual information is accessed directly and the data is extracted from it.

- Inferring the context using statistical or data mining methods. For example, the household identity of a person watching the TV channels may not be explicitly known to a cable TV company; but it can be inferred with accuracy by observing the TV programs watched and the channels visited using various data mining methods. In order to infer this contextual information, it is necessary to build a predictive model.

Methodology is divided in the following phases:

1. Data attributes with in dataset are used for recommendation system. The dataset is randomly fetched in database.
2. Preprocessing of data attributes.
3. Calculation of attributes percentage using serial counting approach.
4. According to attribute percentage store the data within particular cluster.
5. Context based Collaborative filtering approach is applied on data values and finally at fifth step expected recommendations are to be produced.

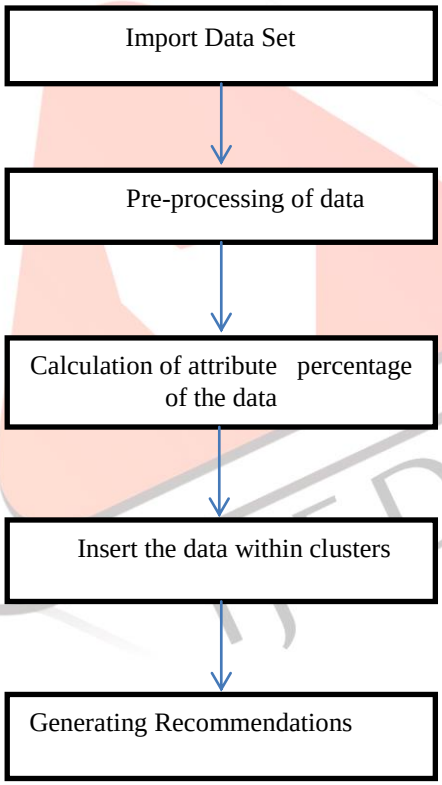

Fig1:Flow Diagram Context based Collaborative Filtering System

Algorithm for Context based System:

Input:
Data set D - A set of n records within dataset D

Output:
Set of recommendation R={r1,r2,r3..$r_{n.}$}

Algorithm:
1: Import dataset D
2: Store all user items (I) within dataset D
3: Fetch all User context Uc from user U

4:Compute context percentage(P) for each Ui.
5: Initialize n clusters $C_{(1...n)}$ each having cluster boundary
For each user (Ui∈Ci) i=1,2..n.
6. Compute all weight (W) of all User (Ui).
7. for each recommendation (Ur) for all Ui
Compute user with highest K_W among all users Ui.
8. if (Item I ∈ (K_W))
Recommend item (I)
9. else repeat step 8
10 return.

Recommendation approaches have gained great response in commercial and research areas. Our proposed system is applicable to current recommendation system, as Context based filtering approach has flexibility in the implementation with in commercial and personal web which will provide effective techniques for collaborative filtering based on context based method. After determining the actual design detail Overall, the field of context based recommendation systems is a relatively new and underexplored area of research, and much more work is needed to explore it. We provided suggestions of several possible future research directions that were presented throughout the paper.

IV. EXPERIMENTAL RESULTS

Context based collaborative algorithm is applied on various data set like on stock data, shopping item data set, IBM stock values etc. To calculate attribute percentage n attributes are considers from m for evaluation. To apply an algorithm an environment is created with user friendly GUI using PHP and Mysql. Equation to calculate mean similarity among the items within dataset.

$$W_{a,u} = \frac{\sum i \varepsilon u + \sum i \varepsilon a}{2} \quad - \quad (1)$$

Where i is the user ,purchased item u after the item a.

In step 2,

$$k=[w_{a,u} \approx w_{a,u}] \quad - \quad (2)$$

where k is the highest$w_{a,u}$ similarity with the active user.
For evaluating the performance of a prediction accuracy of a proposed approach, comparison is done with traditional context based collaborative approach on a Stock data [14].

Table 1 Sample stock data

Record No	Date	Open	High	Low	Close	Volume	Adj Close
100	2/21/2014	184.25	185.71	182.62	182.79	5699300	181.73
200	9/27/2013	188.87	188.94	186.45	186.92	3905500	183.84
300	5/7/2013	202.81	203.73	201.65	203.63	3451000	198.35
400	12/11/2012	193.18	194.8	193.15	194.2	4144000	188.37
500	7/18/2012	184.15	188.59	183.55	188.25	8019500	181.03
600	2/24/2012	198.38	198.56	196.61	197.76	3353200	189.38
700	9/30/2011	176.01	178.27	174.75	174.87	7807400	166.14
800	5/10/2011	169.25	170.9	169.22	170.38	4934800	161.17
900	12/15/2010	145.13	145.72	144.31	144.72	4447400	135.75
1000	7/26/2010	128.18	128.43	127.14	128.41	5174900	119.33

Prediction accuracy =100 – RMSE(Root mean square error) - (3)
and

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_{obs,i} - X_{model,i})^2}{n}} \quad - \quad (4)$$

Where X obs is observe record for i, and X_{model} is calculated value by the model.

For calculation of accuracy of each record value of n=1.
Considering the above data and formulae the Predicted values for 10 records are calculated and presented in the following table No.2.

Table 2 Prediction Accuracy

Record No	100	200	300	400	500	600	700	800	900	1000
Accuracy of TCFM*	25	30	45	62	60	54	69	51	45	42
Accuracy of CCFM**	32	41	55	65	65	87	87	90	81	89

*TCFM –Traditional Collaborative Filtering Model, **CCFM-Context based Collaborative Model

Using Traditional collaborative and context based collaborative filtering algorithm prediction accuracy is measured and results are displayed in graphical form.

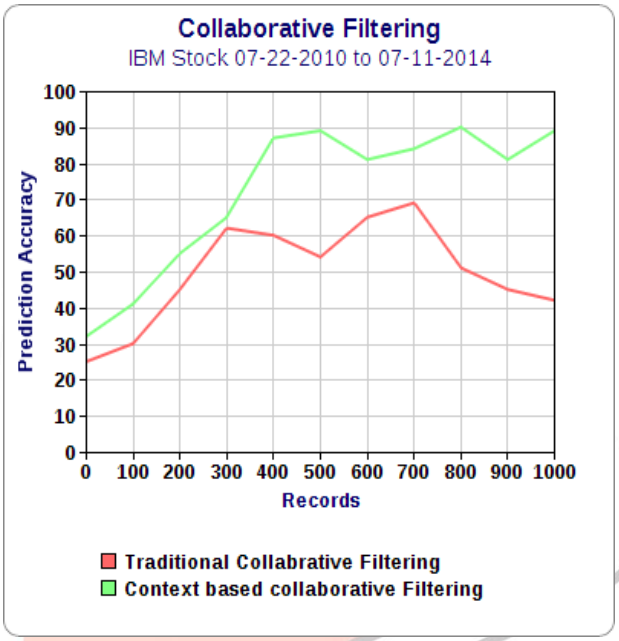

Fig. 2:Graphical representation shows comparison of collaborative filtering.

Similar work is carried out on the stock data [15]from other stock history source and results are shown in the following graph.

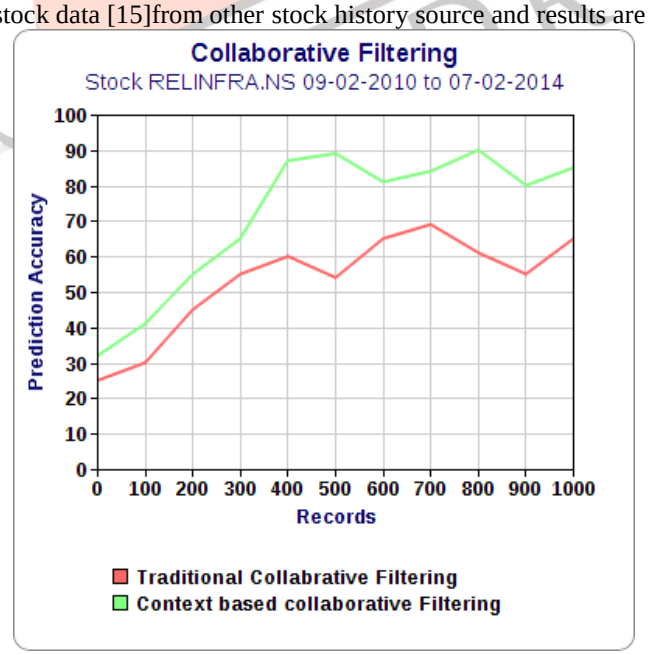

Fig.3:Graphical representation shows comparison of collaborative filtering.

Figure 2,3 illustrates the graphical representaion of performance of improved Context based collaborative filtering approach compare to Traditional collaborative aproach. From the graph it is clear that Context based collaborative filtering approcach generates higher prediction accuracy than the traditional collaborative filtering approcah.

After this, experimentation is done on dataset 10000 records Context based collaborative approach also works well on more than 10000 records within dataset. Context based Collaborative filtering technique used in this system improves the item predictability features. The system is able to perform processing on very large amount of data records .Data processing can be done for a single system or can also be done for distributed applications.

V. CONCLUSION

- Considering the percentage of accuracy of recommendations, current recommendation systems need to be improved.
- To improve the present recommendation systems, a context based collaborative filtering approach can be utilized and implemented in huge commercial business web approach.
- Context collaborative approach proves that it has better performance than the existing collaborative approaches.

So it constitutes a newly developing and promising research area with many interesting and practically important research problems.

REFERENCES

[1] M. Gao, Z. Wu, Personalized context-aware collaborative filtering based on neural network and slope one, in: Lecture Notes in Computer Science, vol. 5738, 2009, pp. 109–116,2011

[2] http://searchengineland.com/study-72-of-consumers-trust-online-reviews-as-much-as-personal-recommendations-114152

[3] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," Communications of ACM, vol. 35, no. 12, pp. 61–70, 2011

[4] P. Resnick and H. R. Varian, "Recommender systems," Communications of the ACM, vol. 40, no. 3, pp. 56–58, 2010

[5] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, "Eigen taste: a constant time collaborative filtering algorithm," Information Retrieval, vol. 4, no. 2, pp. 133–151, 2010.

[6] Min Gaoa,Zhongfu Wu , Feng Jiang User rank for item-based collaborative filtering recommendation, 2010

[7] S. M. McNee, J. Riedl, and J. A. Konstan, "Accurate is not always good: how accuracy metrics have hurt recommender systems," in Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '06), 2009.

[8] J.A. Konstan, B.N. Miller, D. Maltz, J.L. Herlocker, L.R. Gordon, J. Riedl, Group-Lens: applying collaborative filtering to Usenet news, Communications of the ACM 40 (2008) 77–87,2010.

[9] Y. Li, L. Lu, L. Xuefeng, A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce, Expert Systems with Applications 28 (2005) 67–77,2010.

[10] X. Su, T. M. Khoshgoftaar, and R. Greiner, "A mixture imputation-boosted collaborative filter," in Proceedings of the 21th International Florida Artificial Intelligence Research Society Conference (FLAIRS '08), pp. 312–317, Coconut Grove, Fla, USA, May 2008.

[11] M. Deshpande and G. Karypis, "Item-based top-N recommendation algorithms," ACM Transactions on Information Systems, vol. 22, no. 1, pp. 143–177, 2008.

[12] Rong Pan and Martin Scholz. Mind the gaps: Weighting the unknown in large-scale one-class collaborative filtering. In 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2009.

[13] C. Palmisano, A. Tuzhilin, and M. Gorgoglione.Using context to improve predictive modeling of customers in personalization applications. IEEE Transactions on Knowledge and Data Engineering, 20(11):1535–1549, 2008.

[14] http://finance.yahoo.com/q/hp?s=IBM&a=06&b=22&c=2010&d=06&e=14&f=2014&g=d.

[15] http://finance.yahoo.com/q/hp?s=RELINFRA.NS&a=01&b=9&c=2010&d=01&e=7&f=2014&g=d.