International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 3 | ISSN : 2456-3307

A Comparison between Intra-Transaction Association, Inter-Transaction Association and Collaborative Filtering

F. R. Sayyed

Department of Computer Science and Engineering, WIT, Solapur, Maharashtra, India

ABSTRACT

Using the techniques of Collaborative filtering based on historical records of items purchased by users, Recommender systems suggest items to users. Recommender systems use techniques of data mining for determining the similarity among a collection of data items, by analyzing user data and finding hidden useful information or patterns. The Collaborative filtering technique tries to find relationships between the existing data and new data and determines further the similarity and provide recommendations. In this paper, the intra-transaction association, inter-transaction association and collaborative filtering approaches are compared. The similarity between companies is compared in different cases using collaborative filtering technique and accordingly recommendations are generated for users interested in investing in stock market.

Keywords: Recommender Systems, Financial Markets, Stock Predictions, Financial Investments, Recommendation Generation

I. INTRODUCTION

Recommender systems generate recommendations for items to users depending on the users' likes in order to help the users in purchasing items from a large collection of items [1]. These systems store user preferences for items and try to find the relationships between users and items. Recommender systems thus provide suggestions and increase the likelihood of a customer making a purchase. Personalized recommendations are important in markets where there are huge number of choices available, the customer's choice is important and most importantly the price of the item is affordable or modest [2]. The main aim of recommender systems is to analyse historical data of past users and then generate recommendations to new users based upon the similarity in user behaviour based on their buying or selling patterns.

Collaborative filtering (CF) techniques have been successful in enabling the prediction of user preferences in the recommender systems [3]. There are three steps involved in the working of recommender systems: collection of data and its representation in desired format, similarity decision and recommendation computations. Collaborative filtering tries to find the hidden relationships among the new individuals and the existing data in order to further determine the similarity and provide recommendations. Defining the similarity is an important issue and its definition of similarity may be interpreted differently in different applications. The maximum degree of similarity between two objects may decide the final predictions to be done. Similarity decisions may be interpreted in different ways by different techniques of collaborative filtering. For example, people that like certain movies and dislike some other movies in the same categories would be considered as the ones with similar behavior [4]. Similarly, people that like certain novels or books and dislike some other specific books in the same categories can be considered to have similar likes and dislikes. This similarity in their likes and their dislikes can turn out to be an important factor in generating recommendations for new users in the same category in future.

Collaborative filtering is a procedure that comprises filtering of information and involves techniques like collaboration among multiple agents or data sources or viewpoints, etc. It may be considered as a method of deriving automatic predictions about interests of users by gathering various users' information about their preferences or taste. The basic assumption for the collaborative filtering approach is that if a person P and a person Q have the similar opinion on a certain issue, then person P is more likely to have person Q's opinion on any different issue x rather than having the opinion on x of any other person chosen randomly. This means that the technique of collaborative filtering gives priority to similarities between users of the same category and accordingly generates recommendations to the new user. Such predictions are user-specific, but may use the information collected from many users. This approach is somewhat different than the simpler approach that gives an average score for each item of interest, where the score is calculated depending upon the number of votes, the item has gained. This means that the item is rated depending upon similarity with other items rather than being rated on the basis of maximum votes.

The collaborative filtering technique may face challenges in terms of issues like scalability. Conventional algorithms explore the relationships among users in large datasets. User data are dynamic, which means that the data may undergo various changes within a short span of time. Current users may involve change in their behavior patterns, and at the same time, new users may enter the system at any point of time. Data comprising millions of users are to be scanned in real time in order to generate recommendations [5]. Searching among millions of users is a time-consuming and tedious process. Collaborative filtering algorithms that are item-based are proposed for such applications which help in the computations reduction as items properties are static relatively [6].

Collaborative filtering is mostly used in those application areas that involve very large data sets. Collaborative filtering techniques have been applied to different varieties of data including financial service institutions, monitoring and sensing data as in mineral exploration, in web 2.0 applications and electronic commerce where the user data is important and the users relationships is taken into account for taking further decisions. In stock markets, the association among different companies can be used for making predictions [7]. The collaborative filtering recommender system for stock markets makes use of the stocks data in different situations and accordingly gives recommendations to the user for buying or selling the shares of a particular company depending upon the behavior of different companies in different scenarios. This means that the companies whose stock prices either go high or low simultaneously are considered to be similar. The maximum degree of similarity of any company with any other company can be used to check with which company a particular company is similar.

II. LITERATURE REVIEW

Several Collaborative filtering-based recommender systems have been designed and implemented to provide satisfying recommendations to new users [3]. GroupLens is a project, a recommender system that has investigated the issues on automated collaborative filtering [8]. In the system design, the Better Bit Bureaus were developed so as to predict users preferences by computing the coefficients of correlation between users and on averaging ratings for one news article from all. MovieLens is a recommendation system for movies that is based on the GroupLens technology [9].

RecommendationTree (RecTree) is a method that uses the approach of divide-and-conquer for improving correlation-based collaborative filtering and then performs clustering on users' ratings regarding movies. The ratings are extracted from MovieLens Dataset. Ringo provides music recommendations using a word of mouth recommendation mechanism [4]. Ringo tries to find the similarity of users based on user rating profiles. Gustos and Firefly are systems based on recommendation which employed the recommendation mechanism so as to recommend new products for users. WebWatcher has been designed for assisting the information searches on the World Wide Web. WebWatcher suggests such hyperlinks to the user that would give the information which the user was searching for. The general function serving as the similarity model is generated by learning from a sample of trained data logged from users. Yenta is a multi-agent matchmaking system implemented with clustering algorithm and referral mechanism [10]. The Eigentaste algorithm was proposed to reduce the dimensionality of offline clustering and performs online computations in constant time. Jester is an online joke recommendation system based on this algorithm. The clustering is based on continuous user ratings for jokes.

One of the most famous recommendation systems nowadays is the Amazon.com Recommendation that incorporates a matrix of item similarity [11]. The formulation of the matrix is performed offline. Other successful examples of collaborative-filtering-based recommendation systems are music on Yahoo!, Cinemax.com, Launch, TV Recommender, Moviecritic, etc.

A variety of algorithms, methods and models have been suggested in order to resolve the similarity based decisions in recommendation systems based on collaborative filtering. One of the most common methods to determine similarity is cosine angle computation. The Recommendation system of Amazon.com uses this cosine measure for finding the similarity between every two items bought by a particular customer and to generate the item-to-item relationships matrix. Several algorithms that combine the knowledge from Networking, Artificial Intelligence and other fields have also been implemented in the recommender systems.

Expectation Maximization (EM) algorithm provides a standard procedure to estimate the maximum likelihood of latent variable models, and this algorithm is applied to estimate different variants of the aspect model for collaborative filtering [12]. Heuristic of EM algorithm may be applied on latent class models so as to perform aspect extracting or clustering.

III. METHODOLOGY

In this paper, a collaborative filtering recommender system has been described that could be used for making predictions in the stocks market. The aim of this system would be to study and evaluate Collaborative Filtering technique in stock market predictions. The description of the working of the system has been explained below in detail.

In the first step, the stock market data for daily prices of companies was collected from public sources like yahoo. The collaborative filtering technique can be applied on this data and recommendations can be generated for the user. The advantage of using collaborative filtering technique is that it generates recommendations by processing the stock prices data of various companies. It tries to find the similarity between the various companies in different conditions.

After the completion of data collection task, the next step performed was that of pre-processing the data so that the Apache Mahout can work on the data. This process included tasks like removal of erroneous data, inserting missing data, etc. This pre-processing was done so that the raw data is converted to a form suitable for further processing.

Then, the Apache Mahout was used to process the data using the packages provided for collaborative filtering and recommendations were generated. Due to this, huge amounts of stock market data were processed in an efficient manner and the similarity between companies was calculated in the form of general similarity matrix to check which company is in general more similar in behavior to a given company. The following table-I displays the general similarity matrix. The values indicate the number of occurrences when the stock prices for companies were similar for three consecutive days.

TABLE-I
GENERAL SIMILARITY MATRIX

Companies	A	B	C	D
A	-	164	162	62
B	164	-	133	50
C	162	133	-	72
D	62	50	72	-

The above table-I shows that company A is more similar to company B in general than any other company while company D is more similar to company C.

After calculating the general similarity between various companies, the next step was to find the similarity between different companies especially when the prices of the different companies go high simultaneously. For this purpose, another similarity matrix was calculated and accordingly a company with which the given company is similar when stock prices go high was found out. The following table-II displays the Up-similarity matrix. The values indicate the number of occurrences when the stock prices for companies were high at the same time.

TABLE-III
Up Similarity Matrix

Companies	A	B	C	D
A	-	11	7	2
B	11	-	3	1
C	7	3	-	4
D	2	1	4	-

The above table-II shows that company A is more similar to company B when the prices of both companies go high. Similarly, company D is more similar to company C when their prices go high. The Collaborative Filtering technique used in this system improves the predictability features of the system and that the similarity between companies is calculated based upon the fact that the prices of both companies should go high simultaneously.

The next final task was to find the similarity between different companies when their prices go low simultaneously. For this purpose, the third type of matrix was calculated and accordingly a company with which the given company is similar when their stock prices go low was found out. The following table-III displays the Down-similarity matrix. The values indicate the number of occurrences when the stock prices for companies were low at the same time.

TABLE-IIIII
Down Similarity Matrix

Companies	A	B	C	D
A	-	31	37	22
B	31	-	33	41
C	37	33	-	14
D	22	41	14	-

The above table-III shows that company A is more similar to company C when the prices of both companies go low. Similarly, company B is more similar to company D when their prices go low. The Collaborative Filtering technique used in this system improves the predictability features of the system and that the similarity between companies is calculated based upon the fact that the prices of both companies should go low simultaneously.

IV.COMPARISON BETWEEN ASSOCIATION RULE MINING AND COLLABORATIVE FILTERING

A. Intra-transaction Association

Most of the previous studies on mining association rules are on mining intra-transaction associations i.e. the associations among items within the same transactions, where the notion of transaction could be the items bought by the same customer, the events happened on the same day, etc. The Intra Transaction Association Rule Mining tries to find the associations between the variations occurring within the same company. Depending on these variations occurring for a company, the similarity in the current situation of the stock prices of the company with the already occurred previous situations of that same company can be found out. This will help to determine the percentage of occurrence of all the patterns of the same company and to find out which pattern has occurred for a maximum number of times.

In the case of intra-stock mining, the search is for the repetitive pattern on the selected stock and then generates association rules based on this symbolic pattern. Consider the following symbol sequence obtained by the numeric-to-symbolic process:

X X X X X A B C X X X A B C X X A B C X X X X X X A B C

where A, B and C denote the symbols of interest, the sequence ABC occurred 4 times. Depending upon such sequences, the occurrence of that pattern is checked to find out the support for such sequence i.e. the percentage of occurrence of such pattern in the whole set of transactions. Thus, this type of rule generation tries to find how much a given sequence has occurred for a company in the given stocks data.

B. Inter-transaction Association

While intra-stock pattern mining is to find the association rules within a time series, inter-stock association rule mining is concerned with more than one stock company among which the patterns (associations) are mined. Since more than one stock company can be picked, we can find the inter-relationships of companies from same industrial domain, e.g. we can find the association between different IT stock companies. For e.g. when the prices of IBM and SUN go up, 80% of time the price of Microsoft goes up (on the same day).

Association rules that express the association among items from different transaction records are called as inter-transaction association rules. Similarly, in inter-transaction association rule mining, the association between two companies is found out and accordingly association rules are generated. After this, association rules having a support and confidence value more than or equal to a predefined value for minimum support and confidence can be taken into consideration for further activities. Now, a disadvantage of such type of association rule generation is that such types of association rules generated do not always give favourable results. For e.g. the similarity of a company when its prices go high may be with a certain company whereas when its prices go low, it may be similar to some another company. This fact may be taken into consideration while building the Recommendation System for Stock Market using the Collaborative Filtering Technique.

C. Collaborative Filtering

Collaborative Filtering Technique has been widely used by e-commerce sites for finding the relationships between buying patterns of different users. This concept can be used to find the similarity between different stock companies. The similarity may change depending upon the scenario being taken into consideration. The company may be similar to one company in general, similar to a second company when its prices go high considerably and similar to a third company when its prices go low.

In this technique, three similarity matrices were generated for different scenarios. The similarity of a company was checked in all the three cases especially when its prices go high, go low or remain stable. For every company, the similarity is checked with each and every other company to find out which company out of the remaining companies is more similar to this company under consideration. The collaborative filtering technique compares each and every company with the given company and finds the company with maximum similarity for three continuous stock trading days. The General-Similarity matrix has values which indicate how many no. of times a company was similar to every other company for three continuous stock trading days and out of that with which company it had the maximum similar behavior.

Collaborative Filtering Technique can then be used for finding the relationships between the stock price behaviors of different stock companies especially when their prices remain high simultaneously for some time. This concept is the up-similarity between companies which is different than the general-similarity between companies. The similarity may change depending upon the scenario being taken into consideration. The company may be similar to one company in general, but may be similar to another company when its prices go high considerably. The up-similarity matrix indicates this behaviour. For every company, the similarity is checked with each and every other company when their prices simultaneously go high for three consecutive days. The collaborative filtering technique compares each and every company with the given company for the prices of three continuous stock trading days and finds the company with maximum similarity with the current company under consideration.

The values in the up-similarity matrix indicate the no. of instances when the two companies were simultaneously going high. This matrix indicates the values of similarity between any two companies especially when their prices go high simultaneously for three consecutive stock trading days. But, the matrix may have some entries as ‘0’. This indicates that the two concerned companies may have gone high simultaneously for one day or two consecutive days but they never went high simultaneously for three consecutive days. Thus, this up-similarity matrix is useful for analysis of stock prices of companies for providing recommendations in those cases where the companies currently resulting mostly in profit are considered for stocks purchasing.

Similarly, the collaborative filtering technique can also be used for finding the relationships between the stock price behaviors of different stock companies especially when their stock prices remain low for some time. This concept is the down-similarity between companies which is different than the general-similarity and the up-similarity between companies. This similarity may change depending upon for how much time the prices are being taken into consideration. The company may be similar to one company in general, but may be similar to another company when its prices go low considerably.

The down-similarity matrix is generated similar to the previous two matrices. For every company, the similarity is checked with each and every other company when their prices simultaneously remain low for three consecutive stock trading days. The collaborative filtering technique compares each and every company with the given company for the prices of three consecutive stock trading days and finds the company with maximum down-similarity with the current company under consideration. The values in the down-similarity matrix indicate the no. of instances when the two companies were simultaneously low for three consecutive stock trading days. Thus, such types of matrices could be used in stock markets for generating buy/sell recommendations for users.

D. Advantages

- Collaborative filtering technique used in this system improves the stock market predictability features.
- Data processing can be done for a single system or can also be done for distributed applications.
- Because of the use of Apache Mahout, the system is scalable.
- Because of the generation of different types of matrices in different cases, the similarity generation for companies and the further recommendations generated are improved.

V. CONCLUSION

Stock Predictions have been done in various ways using different Data Mining techniques. However, the Collaborative filtering technique using Apache Mahout described in this paper works effectively and generates recommendations depending upon the similarities between companies. Also, the recommender system described in this paper goes one step ahead and checks to see if companies' stock prices are simultaneously going high, going low or remaining stable and accordingly calculates the similarity between companies in these different cases using the different similarity matrices generated.

VI. REFERENCES

[1] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating Collaborative Filtering Recommender Systems", ACM Transactions on Information Systems, Vol. 22, No. 1, January 2004, pp. 5-53.

[2] G. Adomavicius, and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 6, June 2005, pp. 734-749.

[3] W. Hill, L. Stead, M. Rosentein, and G. Furnas, "Recommending and Evaluating Choices in a Virtual Community of Use", Proceedings of ACM CHI '95 Conference on Human Factors in Computing Systems ACM, New York, pp. 194-201.

[4] U. Shardanand, and P. Maes, "Social Information Filtering: Algorithms for Automating 'word of mouth' ", Proceedings of ACM CHI '95 Conference on Human Factors in Computing Systems ACM, New York, pp. 210-217.

[5] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. T. Riedl, "An Algorithmic Framework for Performing Collaborative Filtering", Proceedings of the 22nd International Conference on Research and Development in Information Retrieval (SIGIR '99) ACM, New York, pp. 230-237.

[6] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based Collaborative Filtering Recommendation Algorithms", Proceedings of

the 10th International Conference on World Wide Web, 2001, pp. 285-295.

[7] R. V. Argiddi, and S. S. Apte, “Future Trend Prediction of Indian IT Stock Market using Association Rule Mining of Transaction Data”, International Journal of Computer Applications, Vol. 39, No. 10, Feb 2012, pp. 30-34.

[8] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J Riedl, “GroupLens: an Open Architecture for Collaborative Filtering of Netnews”, Proceedings of the 1994 ACM conference on Computer Supported Collaborative Work, pp. 175-186.

[9] B. N. Miller, I. Albert, S. K. Lam, J. A. Konstan, and J. Riedl, “MovieLens Unplugged: Experiences with an Occasionally Connected Recommender Systems”, Proceedings of the 2003 Conference on Intelligent User Interfaces, pp. 263-266.

[10] L. N. Foner, “Yenta: A Multi-Agent, Referral-Based Matchmaking System”, Proceedings of the First International Conference on Autonomous Agents, ACM, 1997, pp. 301-307.

[11] G. Linden, B. Smith, and J. York,“Amazon.com Recommendations: item-to-item collaborative filtering”, IEEE Internet Computing, Vol. 7, No. 1, Jan-Feb 2003, pp. 76-80.

[12] C. D. Charalambous, and A. Logothetis, “Maximum Likelihood Parameter Estimation from Incomplete Data via the Sensitivity Equations: The Continuous-Time Case”, IEEE Transactions on Automatic Control, Vol. 45, No. 5, May 2000, pp. 928-934.