International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-2 Issue-6, August 2013

Collaborative Filtering Recommender System for Financial Market

F.R.Sayyed, R.V.Argiddi, S.S.Apte

***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 analyzing historical user data and then extracting hidden useful information or patterns.**

Collaborative filtering aims at finding the relationships among the new individuals and the existing data items in order to further determine the similarity and provide recommendations. In this paper, a Collaborative Filtering Recommender System is proposed which can be used for financial markets such as stock exchanges for future predictions.

***Index Terms*—Collaborative Filtering, Financial Markets, Recommender System, Stocks Predictions.**

I. INTRODUCTION

Recommender systems store user preferences over items and find the relation between users and items. Recommender systems suggest items to users based upon the users' likes in order to help the users in purchasing items from a large collection of items [1]. Recommender systems provide suggestions and thus increase the likelihood of a customer making a purchase. Personalized recommendations are important in markets where a large number of choices is available, the customer's choice is important and most importantly the price of the item is modest [2]. In short, the main aim of recommender systems is to analyse historical past users' data and then provide recommendations to a new user depending upon the similarity in user behaviour on the basis of 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: object data collection and representation, 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 the definition of similarity may vary in different applications. The degree of similarity between two objects may decide the final predictions to be done.

Similarity decisions may be interpreted in different ways by different collaborative filtering techniques. 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].

Collaborative filtering has two senses, a narrow one and a general one. In general, collaborative filtering is the process of information filtering using techniques involving collaboration among multiple agents, viewpoints, data sources, etc. In the newer, narrower sense, collaborative filtering is a method of making automatic predictions about user's interests by collecting preferences or taste information from many users. The underlying assumption of the collaborative filtering approach is that if a person A and a person B have the same opinion on a certain issue, then person A is more likely to have B's opinion on a different issue x rather than having the opinion on x of any other person chosen randomly. Such predictions are specific to the user, but use information collected from many users. This approach is 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.

The collaborative filtering technique may face challenges in the form 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 changes within short intervals of time. Current users may involve change in their behavior patterns, and at the same time, new users may enter the system at any moment. Data comprising millions of users are to be examined in real time in order to provide recommendations [5]. Searching among millions of users is a lengthy and time-consuming process. For such applications, item-based collaborative filtering algorithms are proposed that enable reductions of computations because properties of items are relatively static [6].

Collaborative filtering is mostly used in those application areas that involve very large data sets. Collaborative filtering techniques have been applied to different kinds of data including sensing and monitoring data such as in mineral exploration, in financial service institutions that integrate many financial sources, in electronic commerce and web 2.0 applications where the user data is of utmost importance and the relationships among different users is to be considered for taking further decisions. In stock markets, the association among different companies can be used for making predictions [7].In this paper, a collaborative filtering recommender system for stock markets is proposed which makes use of the stocks data and gives recommendations to the user for buying or selling the shares of a particular company.

Manuscript published on 30 August 2013.
***** Correspondence Author (s)
F.R.Sayyed, Computer Science and Engineering department, Solapur University/ Walchand Institute of Technology / Solapur-413003, Maharashtra, India.
R.V.Argiddi, Computer Science and Engineering department, Solapur University/ Walchand Institute of Technology / Solapur-413003, Maharashtra, India.
S.S.Apte, Computer Science and Engineering department, Solapur University/ Walchand Institute of Technology / Solapur-413003, Maharashtra, India.

II. LITERATURE REVIEW

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

RecommendationTree (RecTree) is one method using divide-and-conquer approach to improve correlation-based collaborative filtering and performing clustering on movie ratings from users. The ratings are extracted from MovieLens Dataset. Ringo provides music recommendations using a word of mouth recommendation mechanism [4]. Ringo determines the similarity of users based on user rating profiles. Firefly and Gustos are two recommendation systems which employed the word-of-mouth recommendation mechanism to recommend products.

WebWatcher has been designed for assisting information searches on the World Wide Web. WebWatcher suggests those 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 the clustering algorithm and the referral mechanism [10]. The Eigentaste algorithm was proposed to reduce the dimensionality of offline clustering and to perform online computations in constant time. Jester is an online joke recommendation system that is based on this algorithm. The clustering is based on continuous user ratings of jokes.

One of the most famous recommendation systems nowadays is the Amazon.com Recommendation that incorporates a matrix of the 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.

Many methods, algorithms, and models have been proposed to resolve the similarity decisions in collaborative-filtering-based recommendation systems. One of the most common methods to determine the similarity is cosine angle computation. Amazon.com Recommendation system uses this cosine measure to find the similarity between every two items bought by each customer and to generate the item matrix containing item-to-item relationships. Several algorithms that combine the knowledge from Artificial Intelligence (AI), Networking and other fields have also been implemented in the recommendation systems.

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

III. METHODOLOGY

A. *Algorithm*

The following is the algorithm for the proposed recommender system:

1) Start.
2) Collect stock market data from public sources like yahoo or google or use simulated stock transaction data.
3) Preprocess the data as per required input format of Apache Mahout for doing further operations.
4) Process the data using Apache Mahout and Hadoop to output the recommendations in its native format.
5) To obtain the recommendations in user-readable format, process the data using various data mining and statistical techniques.
6) End.

B. *Proposed System*

In this paper, a collaborative filtering recommender system has been proposed 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 architecture of the proposed system is shown in Fig. 1.

In the first step, the stock market data is collected from public sources like yahoo or google. If public stocks data is not available, then simulated data can be used for further processing. 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 data items as well as the past user behavior patterns. It tries to find the similarity between the new user's and the past users' behavior patterns.

After the completion of data collection task, the next step is to prepare or preprocess the data so that the Apache Mahout can work on the data. This process includes tasks like removal of outliers, inserting missing data, etc. This preprocessing is done so that the raw data is converted to a form that is suitable for further processing.

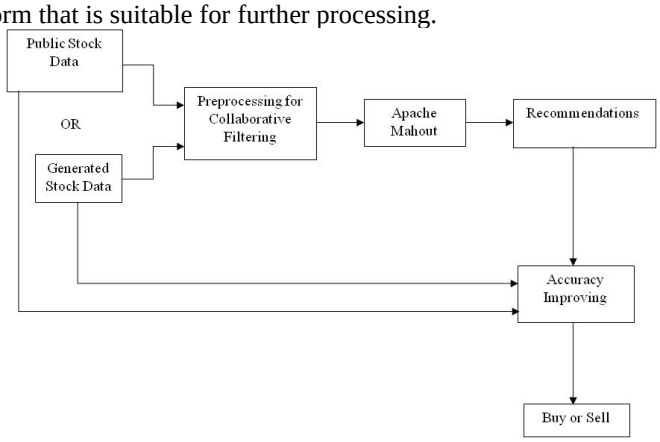

Figure 1. System architecture

Then, the Apache Mahout either as a standalone or in a distributed setup using Hadoop cluster will process the data using the packages provided for collaborative filtering and would output the recommendations. Due to this, huge amounts of stock market data could be processed easily and would generate recommendations efficiently.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The next step is to make the recommendations generated by Apache Mahout user-readable. For this purpose, the output generated by Apache Mahout undergoes processing via various techniques which may involve Data Mining techniques in combination with Statistical and Mathematical analysis techniques. The aim of this proposed recommender system would involve improvement of the recommendations' accuracy. Also, it would establish a good correlation between the original data and the recommendations generated. The advantage of using such a system for stock market predictions is that because of the use of Hadoop, it could handle processing on large amounts of data efficiently. Also, the use of Collaborative Filtering technique improves the predictability features of the system.

C. Advantages

- Collaborative filtering technique used in this system improves the stock market predictability features.
- The use of Hadoop in this system enables to perform processing on very large amount of data called Big Data.
- 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.

IV. CONCLUSION

Stock Predictions can be done in various ways using different Data Mining techniques. However, the Collaborative filtering technique using Apache Mahout and Hadoop proposed in this paper would work effectively. Thus, the recommender system would generate better recommendations for new users in the field of Stock Market.

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.

Farooque R. Sayyed is a Post Graduation student pursuing M.E in Computer Science and Engineering from Walchand Institute of Technology, Solapur. He received his Bachelor of Engineering degree from Walchand Institute of Technology, Solapur affiliated to Solapur University in 2010. His research interests lie in the area of Data Mining and its application in the field of stock predictions, specifically focusing on developing a recommender system using collaborative filtering technique.

Mr. Rajesh V. Argiddi is an Associate Professor in Computer Science and Engineering Department at Walchand Institute of Technology, Solapur. He received his B.E degree from Shivaji University, Kolhapur and M.E degree from Shivaji University, Kolhapur. He is currently doing his Ph. D from Solapur University, Solapur. His research area lies in Data Mining. Currently he is working for Indian stock market behavior analysis using Data Mining techniques. For this, he has published papers in various renowned journals.

Dr. Sulbha S. Apte is the Professor and Head of Department in Computer Science and Engineering at Walchand Institute of Technology, Solapur. She received her B.E degree from Pune University, Pune and M.E degree from Shivaji University, Kolhapur. She received her Ph. D degree from S.R.T.M Nanded, Maharashtra, India. Book published: Digital Systems and Microprocessor Fundamentals published by Jaico Publishing House, Bombay in 1993. Her name included in World 'Who's who' by Marquis, US, also her name included in 'The Directory of Computer Professionals' and she acquired a Session Chair at ICINT: International Conference on Information & Network Technology, Shanghai, China (23 June 2010). She is a member of following Reputed Bodies:
Life Member - ISTE (No. LM2349), Fellow of Institute of Engineers (No. F-114920-9).

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)