GENERATING RECOMMENDATIONS FOR STOCK MARKET USING COLLABORATIVE FILTERING

F. R. Sayyed, R. V. Argiddi, S. S. Apte
[1]Department of CSE, WIT, Solapur University, Solapur, India.

ABSTRACT

Recommendation systems apply different data mining techniques on data and try to find the similarity among a huge collection of data items. These systems work on the technique of collaborative filtering, which makes use of the past users' data and gives recommendations to new users. Collaborative filtering technique can be applied to various application areas where there is similarity between the past users' behaviour pattern and the current new users. The recommendation system for stock markets described in this paper works using the technique of collaborative filtering and makes use of Apache Mahout to generate recommendations for new users.

KEYWORDS: *Recommendation System, Stock Market, Collaborative Filtering, Stock Predictions.*

1. INTRODUCTION

Recommendation systems are developed with an aim of providing recommendations to the new users in any field. The main task of such systems is to provide meaningful recommendations to new users for items or products of their interest. Recommender systems apply different techniques of data mining on data and find similarities among various data items [1]. Some real-world examples of recommender systems are suggestions for books on Amazon, suggestion for movies on Netflix, etc. The recommendation systems used in these examples have a design which depends on the particular domain and the characteristic features of the data available on which the data mining techniques are to be applied. For providing recommendations to users, the system may require to perform operations on both the data related to products and the data related to users. The recommender systems analyze the product data and user data and generate a user-item similarity matrix which indicates up to what extent the users are similar. This similarity would then be used by the system to give recommendations to new users. Recommender systems use the technique of collaborative filtering or content-based filtering. Collaborative filtering (CF) systems build a model based on a user's past behaviour as well as similar decisions taken by other users, and then try to use this model to predict items that the user may be interested in [2]. Content-based filtering systems try to use the characteristics of an item in order to recommend new items with similar properties [3].

Collaborative filtering technique makes use of past user activities. The algorithms used for collaborative filtering identify the relationships between users and items, and make predictions depending upon the similarities between current user and past users' behaviour patterns. Recommender systems help new users by searching through a collection of various products to find the most important product for them depending upon their past taste information [4]. The fundamental assumption of collaborative filtering is that if users A and B rate m items similarly, or have similar behaviours, then they will rate or act on other items similarly [5]. The CF techniques use a database of preferences for items by users to predict additional topics or products in which a new user might be interested. In a typical CF scenario, there is a list of m users $\{u_1, u_2, \ldots, u_m\}$ and a list of n items $\{i_1,$

$i_2, \ldots, i_n\}$ and each user has a list of items which the user likes. This is represented by a user-item ratings matrix. The collaborative filtering algorithms can be applied on this matrix to predict items for the new user [6].

2. LITERATURE REVIEW

Many existing recommender systems are based on collaborative filtering which have been designed and implemented to provide recommendations to users [7]. GroupLens project is a recommendation system for netnews and has investigated the issues on automated collaborative filtering [8]. In the system design, the Better Bit Bureaus were developed to predict user preferences that are based on computing the correlation coefficients between users and on averaging ratings for one news article from all. MovieLens is a recommendation system for movies which is based on the GroupLens technology [9].

RecommendationTree (RecTree) is a method that makes use of divide-and-conquer approach to improve correlation-based collaborative filtering and performs clustering on movie ratings from users. Ringo gives music recommendations using a word of mouth recommendation mechanism [10]. Ringo finds the similarity among users based on user rating profiles. Firefly and Gustos are the two recommendation systems employing the word-of-mouth recommendation mechanism and accordingly recommending products to users.

WebWatcher was designed with an aim of assisting information searches on the World Wide Web. WebWatcher suggests those hyperlinks to the user which would give the intended information to the user. The general function serving as similarity model is generated by learning from a sample of trained data logged from users. A multi-agent matchmaking system, Yenta, is implemented with the clustering algorithm and the referral mechanism [11]. The Eigentaste algorithm was proposed to reduce the offline clustering dimensionality and to perform online computations in constant time. An online joke recommendation system that is based on this algorithm is Jester. There is a continuous user rating of jokes, on which the clustering is based.

One of the most famous recommendation systems being used today is the Amazon.com Recommendation that makes use of a matrix of the item similarity [12]. The formulation of the matrix is performed offline. A few successful examples of collaborative-filtering-based recommender systems are music on Yahoo!, Launch, Cinemax.com, TV Recommender, MovieCritic, etc. Many algorithms and models have been proposed to resolve the similarity decisions in collaborative-filtering-based recommender systems. Among the methods used for determining similarity, the most common one is cosine angle computation. Amazon.com Recommendation system makes use of this cosine measure to determine the similarity between every two items bought by each customer and generates the item matrix containing item-to-item relationships. Several algorithms that combine the knowledge from different fields like Artificial Intelligence (AI), Networking, etc. have also been implemented in the recommendation systems.

3. METHODOLOGY

The recommender system for stock markets described in this paper is based on the collaborative filtering technique. The share prices of various stock companies may change rapidly during the course of time. The prices may sometimes go up, sometimes remain low while at other times, they may remain at a constant level. The recommender system described in this paper uses the technique of collaborative filtering to generate recommendations to users which may help them to decide in which company they should invest their money. The detailed procedure of the system and the results obtained after following it are described in the further part of this section.

In the initial step, the stock market data was collected from public sources. This data formed the database for the system on which the further processing was performed. The next step performed was pre-processing of data. This pre-processing was done so that the Apache Mahout can work on the data. It means that the data was converted into a form on which the Apache Mahout can work and generate recommendations. The stocks data for companies was first analysed to find the occurrence of frequent patterns in the stock prices. The prices of different companies may go high, low or remain unchanged over the course of time. Such behaviour of different companies was checked to see if they

resemble or not. Actually, the companies do not resemble each other completely. But the aim was to find out groups of companies which resemble up to a large extent. Accordingly, those companies for which the behaviour is same for many instances were found out and were concluded as having similar behaviour patterns. This similarity helped to find out what the future state of the company's shares will be. Then collaborative filtering technique was applied on the stocks data and user transactions data using the packages provided for the same by Apache Mahout. When companies behave similarly and are in profit, users buy the shares of these companies. This similarity in user behaviour was taken into consideration so as to suggest the new user to buy the shares of that particular company. The following is the similarity matrix resembling the similarity between different companies.

Table 1. Similarity Matrix

Companies ⇨ ⇩	**A**	**B**	**C**	**D**
A	--	188	112	132
B	188	--	146	119
C	112	146	--	166
D	132	119	166	--

As shown in Table 1 above, the similarity matrix depicts the number of times the behaviour of the companies was found to be similar over a period of time. This table clearly shows that company A is more similar to company B than the remaining companies, C and D. Similarly, companies C and D are more similar to each other than any other company. This similarity matrix was formed after performing processing on stocks data of a few months. After this, the further work was to make use of the user transaction data which helped to give further recommendations to users. This means that if a company behaves similar to another company and both are in profit, and at the same time users have bought shares of that company, then this system would recommend a new user to buy the shares of that company with a view that the user would be in profit in future. This system could be used for any other financial market where collaborative filtering can be applied so as to provide recommendations to new users.

4. CONCLUSION

There are different Data Mining techniques used for stock predictions. But the collaborative filtering technique used by the recommender system described in this paper improves the predictability features. Thus the system generates better recommendations for new users in the field of stock market trading. Also, the system is scalable.

REFERENCES

[1] Anne Yun-An Chen & Dennis McLeod, (2006) "Collaborative Filtering for Information Recommendation Systems", Encyclopedia of E-Commerce, E-Government, and Mobile Commerce, Chapter 20, pp. 118-123.

[2] Prem Melville & Vikas Sindhwani (2010) "Recommender Systems", Encyclopedia of Machine Learning, Chapter No. 00338, pp. 1-9.

[3] R. J. Mooney and L. Roy, (1999) "Content-based book recommendation using learning for text categorization", Proceedings of the Workshop on Recommender Systems: Algorithms and Evaluation (SIGIR '99), Berkeley, Calif , USA.

[4] Xiaoyuan Su & Taghi M. Khoshgoftaar, (2009) "A Survey of Collaborative Filtering Techniques", Advances in Artificial Intelligence, Vol. 2009, Article ID 421425, 19 pages.

[5] K. Goldberg, T. Roeder, D. Gupta & C. Perkins, (2001) "Eigentaste: a constant time collaborative filtering algorithm", Information Retrieval, Vol. 4, No. 2, pp. 133-151.

[6] B. N. Miller, J. A. Konstan & J. Riedl, (2004) "PocketLens: toward a personal recommender system", ACM Transactions on Information Systems, Vol. 22, No. 3, pp. 437-476.

[7] W. Hill, L. Stead, M. Rosentein & G. Furnas, (1995) "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.

[8] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom & J. Riedl, (1994) "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 & J. Riedl, "MovieLens Unplugged: Experiences with an Occasionally Connected Recommender System", Proceedings of the 2003 Conference on Intelligent User Interfaces, pp. 263-266.

[10] U. Shardanand & 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.

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

[12] G. Linden, B. Smith & J. York, (2003) "Amazon.com Recommendations: item-to-item collaborative filtering", IEEE Internet Computing, Vol. 7, No. 1, Jan-Feb 2003, pp. 76-80.

AUTHORS

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, especially focusing on developing a recommender system using collaborative filtering technique.

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 behaviour analysis using Data Mining techniques. For this, he has published papers in various renowned journals.

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 Techology, 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).