ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 10, October 2017

Data Mining System Using Multiple Agents For Stock Market Prediction

[1] Prajyoti P.Sawant Dessai, [2] Prof. Megha Ainapurkar
[1] Master of Engineering (IT), [2] Assistant Professor
[1][2] Padre Conceicao College of Engineering

Abstract: **Intelligent Agent based distributed data mining system to generate stock market database rules and transfer it to the main controller. The combo of data mining technology with multi-Agent technology is expected to provide efficient roadmap for developing highly configurable software approaches that incorporate knowledge and provide decision making capabilities. This paper has offers the performance of every Agent in detail and the basic thoughts of the mining method adopted. The system achieves the goal of mining data accurately and effectively, and offer the personalized service, and improves the efficiency of the system.**

Index Terms— Data Mining ,Agent Technology, Distributed Data Mining, Time Seies Association Analysis, Random Walk, Multi-Agent System, Yahoo! Server, Market BasketData.

I. INTRODUCTION

Data mining is the analysis of (often large) observational datasets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner. Data mining is the extraction of hidden perspective information from large databases, is a powerful new technology with great potential to help companies focus on the most important information. Data mining technology has proven a successful gateway for discovering useful knowledge and for enhancing business intelligence in a range of application fields.

The domain of data mining (DM) and its related techniques and technologies has been greatly expanded in the last few years. Agent technology is more current research. The field of Distributed Data Mining (DDM) deals with these challenges in analyzing distributed data and offers many algorithmic solutions to perform different data analysis and mining operations in a fundamentally distributed manner that pays careful attention to the resource constraints.

To design and develop the automated distributed data mining system using agents. The coupling of data mining and autonomous intelligent agent is expected to provide an efficient roadmap for developing highly reconfigurable software approaches that incorporate domain knowledge and provide decision making capabilities. It introduces Agent's characteristics into the process of data mining, such as autonomy, automatic response, interaction and adaptation to environment, which enables the data mining system to take on high efficient mining ability, and offers better personalized service to user.

II. THE DATA MINING SYSTEM BASED ON

Multi-Agent

The most widely used accepted definitions for this term is "an agent acts on behalf of someone else, after having been authorized". An agent is a physical or virtual entity, which runs approximately as follows: which is capable of acting in an environment, which can communicate directly with other agents, which is driven by a set of tendencies, which possesses resources of its own, which is capable of perceiving its environment.

Agent is the birth and development of AI technology and network technology development in the inevitable result of the artificial intelligence knowledge in the field of engineering Expert systems, decision-making theory and methods, such as the development direction has made the development of intelligent applications possible.

Agents comprise a powerful technology for the analysis, design and implementation of autonomous intelligent systems that can handle distributed problem-solving, cooperation, coordination, communication, and organization in a multiplayer environment. It's also a promising computing paradigm for dealing with system complexities such as openness, distribution, human involvement, societal characteristics, and intelligence emergence. Agent research focuses on theoretical, methodological, technical, experimental, and practical issues and the means to handle system complexities. For several decades, data mining (DM) and machine learning have retained their status among the topmost research and application areas of AI and intelligent information technology.

DDM is a complex system focusing on the distribution of resources over the network as well as data mining processes. The very core of DDM systems is the scalability as the

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 10, October 2017

system configuration may be altered time to time, therefore designing DDM systems deals with great details of software engineer issues such as reusability, extensibility, and robustness. For these reasons, agents' characteristics are desirable for DDM systems. Furthermore, the decentralization property seems to fit best with the DDM requirement. Distributed data mining is a process that uses distributed algorithms and that finds knowledge from logically and physically distributed data sources. The target of distributed data mining is to carry out data mining work on distributed and heterogeneous resources. It can be divided into three steps: identifying data distribution, data pretreatment, data mining. Distributed data mining can go along on both local and global levels.
The frame of the system is designed as figure1 shows.

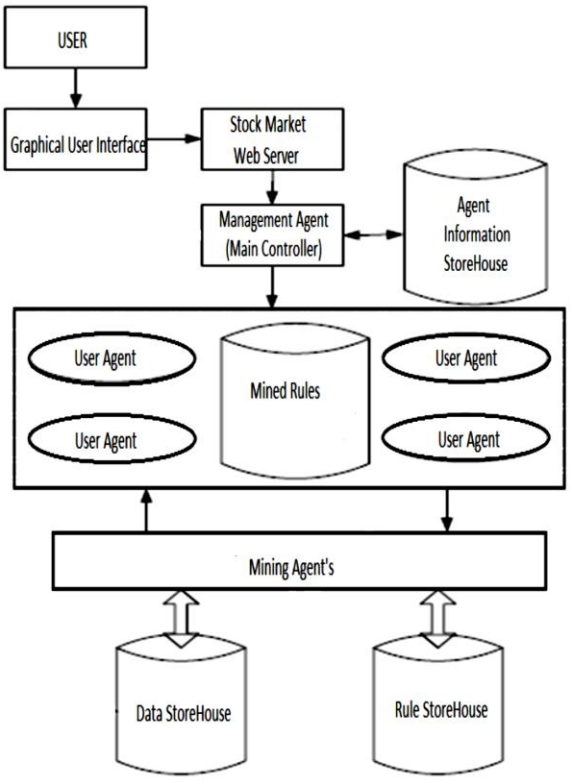

Figure 1. The System Structure

(1) Main Controller System receives login information submitted from servlet and verifies the user name and password in the database. Also releases the user object as well assign the agented to user thread.

(2) Main Controller System is managed by administrator. The administrator is suppose to start agent, kill agent, suspend and resume agents. The administrator normally has fully control of the main controlling system.

(3) Main Controller System starts the mining agents on the database server and delete them after completing their action.

(4) Main Controller System is suppose to start mobile agent on database server which retrieves the mined information from database server and transfer it to the user database on the master machine.

The functions of agent in system

(1) Management Agent
It receives user's request from the user's graphic interface, then look over whether there is user's information in Agent information storehouse. If it does not exist, establish a user Agent for this user, and provide a systematic serial number, and initialize the user's model storehouse; if it does exist, activate the user Agent. And send users' demand to user Agent, then give the information feedback excavated by mining Agent to the user.

(2) User Agent
User Agent is responsible for communicating with management Agent and receiving and passing user's order, to the mining Agent, finally analyses and coordinates the mining result transmitted by mining Agent to upgrade user's model storehouse and preset information storehouse and submit to the management Agent at the same time. According to current user model storehouse, the user Agent provides feedbacks and studies, which make the information storehouse keep information alive all the time, eliminate that outdated, and offer the latest.
User Agent communicates with each other with KQML language.

(3) Mining Agent
The mining Agent is the analytic center of the whole Agent system. It mines the information registered and also, it excavates the data from user's model storehouse according to the rule in the rule store- house.

(4) Information storehouse
There are two kinds of information storehouses in this system: One is user model storehouse, the other is Agent information storehouse. User model storehouse keeps user's basic information and historical web-hit record. While the

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 10, October 2017

Agent information storehouse keeps customer's all materials in website, which benefits analysis and predicts a lot.

III. ALGORITHMS FOR DATA MINING

Time Series Association Analysis
Rule discovery from time series data is a data mining technique that tries to find relationships of sequential data. Finding association rules from time series data is different from finding such rules in traditional data because time series data is ordered data with a sequence that must be preserved.
Stock Data Mining plays an important role to visualize the behavior of financial market. Time series data are difficult to manipulate, but when they are treated as symbols instead of data points interesting patterns are discovered and it becomes an easier task to mine them. The goal of this project is to propose a novel technique to find association rules from time series data. Our technique analyzes the numerical time series and show the resulting rules. Every stock sold is classified into categories, a company is identified based on its primary activities. These categories include Automobile, Banking, Infrastructure, Steel, FMCG, IT, Power and so on.

A. Pre-Processing of stock data

The data preprocessing is an essential part of data mining. We use feature extraction method to preprocess the stock data to get the difference between two consecutive closing values of stocks, for any C_t € C for any stock market data, where C is the set of closing value, we collect pairs of $< C_t\text{-}1, C_t>$ for each stock market data where t is the date of stocks.
The collected feature is then subtracted from each other to see if there is an increase or decrease in the closing values of stock market as

$$\text{Feature} = C_t - C_{t-1}$$

i.e. we subtract the second closing value of stocks with the first closing value and based on its result assign the values. If the feature value of the closing value of the stock market is positive (+) we consider it as 1, if the value is negative (-) or 0 we consider it as 0. This approach is considered as a random walk in the literature.

B. Apriori Algorithm: Finding Frequent itemsets Using Candidate Generation

Apriori is an algorithm for frequent itemset mining and association rule learning over transactional databases. The name of the algorithm is based on the fact that the algorithm uses prior knowledge of frequent itemset properties. It employs an iterative approach known as a level wise search. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those itemsets appear sufficiently often in the database. The frequent item set determined by Apriori can be used to determine association rules which highlight general trends in the database

C. Rules Generation

Association rule learning is a rule-based machine learning method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness. For example, the rule {Onion, Potatoes} => {Burger} found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, they are likely to also buy hamburger meat. Such information can be used as basis for decision about marketing activities.
The goal is to determine which items occur together in shopping baskets frequently. We say a rule X —> Y is true if the presence of item X implies the presence of item Y in a market basket. For example, if many customers who purchase onion also purchase potatoes, we could say the itemset {Onion, Potatoes} is frequent and we could deduce the rule Onion —> Potatoes. To determine how often these items occur together, we use the support statistics: support is an indication of how frequently the itemset appear in the database.

$$\text{Supp}(X \text{ —> } Y) = s(X,Y)/N$$

Where s(X) denotes the frequency of item X and N is the total number of market baskets.
To describe the probability that the presence of item X implies the presence of item Y, we use the confidence statistic: confidence is an indication of how often the rule has been found to be true.

$$\text{conf}(X \text{ —> } Y) = s(X,Y)/s(X)$$

The best rules will have both high support and confidence. The simplest way to generate such rules is to enumerate all possible rules, remove all rules with support less than minimum support threshold T, and sort the remaining rules based on confidence. To save on computation, we do not need to enumerate all rules. The Apriori Principle states that if an itemset is frequent, then all of its subsets must be

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 10, October 2017

frequent as well. Using this principle, once we detect an item set with support less than T, we can prune all supersets of that item set from consideration as well. These basic ideas of grocery store rule generation can be applied to time-series. In our case, we wish to detect which industries are related. That is, we wish to know if a change in stock prices in one industry causes a change in stock prices in another industry.

DATASET

We have taken historical data downloaded from Yahoo! Server of one seventy five companies from a time period of January 2016 to October 2016.

CONCLUSION

Data accessing and mining systems are increasingly becoming critical and need to have enough domain knowledge about the data mining system. In this research we propose a generalized framework of a Data mining system using autonomous intelligent agents. Automation can provide huge advantages in time and cost. Our specific goals in this research include development of a flexible architecture for an automated data mining system using intelligent agents. Also, with the help of an intelligent agent choosing the right algorithm and techniques for the databases.
We showed that the standard "market basket" association analysis could be applied to time series. This can be used to generate rules which point out the industries that move together. The rules generated with the highest confidence and support seemed to be natural, which indicates that our system is working. The unexpected rules will necessarily have lower support and confidence. The stock market data is inherently unpredictable.

REFERENCES

[1] Qiang Wu, "An efficient data mining system using multi-agent of e-commerce", 2009 Seventh ACIS International Conference on Software Engineering Research, Management and Applications, pp. 51-55.

[2] Lizeng Wu, Yongli Zhu, Xueyu Li, Jinsha Yuan, "Application of Multi-agent and Data Mining Techniques in Condition Assessment of Transformers", 2004 lntemallonal Conference on Power System Technology, pp. 823-827.

[3] Andreas L. Symeonidis, Kyriakos C. Chatzidimitriou, Ioannis N. Athanasiadis, Pericles A. Mitkas, "Data mining for agent reasoning: A synergy for training intelligent agents", 2007 Conference on Engineering Applications of Artificial Intelligence.

[4] Sadi Evren SEKER, Cihan MERT, Khaled Al-NAAMI, Nuri OZALP, Ugur AYAN, "TIME SERIES ANALYSIS ON STOCK MARKET FOR TEXT MINING CORRELATION OF ECONOMY NEWS" ,2014 INTERNATIONAL JOURNAL OF SOCIAL SCIENCES AND HUMANITY STUDIES.

[5] Prapanna Mondal , Labani Shit and Saptarsi Goswami, "STUDY OF EFFECTIVENESS OF TIME SERIES MODELING (ARIMA) IN FORECASTING STOCK PRICES ",April 2014 International Journal of Computer Science, Engineering and Applications.

[6] Ehsan Hajizadeh, Hamed Davari Ardakani and Jamal Shahrabi, "Application of data mining techniques in stock markets: A survey", July 2010 Journal of Economics and International Finance Vol. 2(7), pp. 109-118.

[7] H. Yang, L. Chan, and I. King. Support vector machine regression for volatile stock market prediction. Proceedings of the Third International Conference on Intelligent Data Engineering and Automated Learning, 2002.

[8] J. Bollen and H. Mao. Twitter mood as a stock market predictor. IEEE Computer, 44(10):91–94.

[9] Sachin Kamley, Shailesh Jaloree, R.S. Thakur, " An Association Rule Mining Model for Finding the Interesting Patterns in Stock Market Dataset", May 2014, International Journal of Computer Applications.

[10] Zhengxiang Zhu, Dalian and Wuqi Song, " A Multi-Agent and Data Mining Model for TCM Cases Knowledge Discovery", 2008 ISECS International Colloquium on Computing, Communication, Control, and Management.