ORIENTAL JOURNAL OF
COMPUTER SCIENCE & TECHNOLOGY
An International Open Free Access, Peer Reviewed Research Journal
Published By: **Oriental Scientific Publishing Co., India.**
www.computerscijournal.org

ISSN: 0974-6471
December 2012,
Vol. 5, No. (2):
Pgs. 263-267

Review on Financial Forecasting using Neural Network and Data Mining Technique

MEHZABIN SHAIKH and GYANKAMAL J. CHHAJED[2]

Department of Computer Engineering, VPCOE, Baramati, India.

(Received: April 12, 2011; Accepted: June 04, 2011)

ABSTRACT

The rise of economic globalization and evolution of information technology, financial data are being generated and accumulated at an extraordinary speed. As a result, there has been a critical need for automated approaches to effective and efficient utilization of massive amount of financial data to support companies and individuals in strategic planning and investment decision-making. The competitive advantages achieved by data mining include increased revenue, reduced cost, and much improved marketplace responsiveness and awareness. There has been a large body of research and practice focusing on exploring data mining techniques to solve financial problems. This paper describes data mining in the context of financial application from both technical and application perspective by comparing different data mining techniques.

Key words: Data mining in Financial Forecasting, Data Mining Algorithm, Technical Indicators, BP neural network.

INTRODUCTION

Data mining techniques have been used to uncover hidden patterns and predict future trends and behaviours in financial markets. The competitive advantages achieved by data mining include increased revenue, reduced cost, and much improved marketplace responsiveness and awareness. Data mining has been applied to a number of financial applications, including development of trading models, investment selection, loan assessment, portfolio optimization, fraud detection, bankruptcy prediction, real-estate assessment, and so on

Literature survey

Classification and issues of data mining in financial application

Data mining aims to discover hidden knowledge, unknown patterns, and new rules from large databases that are potentially useful and ultimately understandable for making crucial decisions.

Based on the type of knowledge that is mined, data mining can be mainly classified into the following categories:

Association rule mining

Uncovers interesting correlation patterns

among a large set of data items by showing attribute-value conditions that occur together frequently. A typical example is market basket analysis, which analyzes purchasing habits of customers by finding associations between different items in customers' shopping baskets.

Classification and prediction

Is the process of identifying a set of common features and models that describe and distinguish data classes or concepts. The models are used to predict the class of objects whose class label is unknown. A bank, for example, may classify a loan application as either a fraud or a potential business using models based on characteristics of the applicant. A large number of classification models have been developed for predicting future trends of stock market indices and foreign exchange rates.

Clustering analysis

Segments a large set of data into subsets or clusters. Each cluster is a collection of data objects that are similar to one another within the same cluster but dissimilar to objects in other clusters. In other words, objects are clustered based on the principle of maximizing the intra-class similarity while minimizing the inter-class similarity.

Sequential pattern and time-series mining example, clustering techniques can be used to identify stable dependencies for risk management and investment management.

looks for patterns where one event (or value) leads to another later event (or value). One example is that after the inflation rate increases, the stock market is likely to go down.

First, data mining needs to take ultimate applications into account. For example, credit card fraud detection and stock market prediction may require different data mining techniques.

Second, data mining is dependent upon the features of data. For example, if the data are of time series, data mining techniques should reflect the features of time sequence.

Third, data mining should take advantage of domain models. In finance, there are many well-developed models that provide insight into attributes that are important for specific applications. Many applications combine data mining techniques with various finance and accounting models (e.g., capital asset pricing model and the Kareken–Wallace model)[1].

Existing data mining techniques

Overview of Data Mining Techniques

Neural Networks

Artificial neural networks are computer models built to emulate the human pattern recognition function through a similar parallel processing structure of multiple inputs. A neural network consists of a set of fundamental processing elements (also called neurons) that are distributed in a few hierarchical layers. Most neural networks contain three types of layers: input, hidden, and output.

Genetic Algorithms

The basic idea of genetic algorithms is that given a problem, the genetic pool of a specific population potentially contains the solution, or a better solution. Based on genetic and evolutionary principles, the genetic algorithm repeatedly modifies a population of artificial structures through the application of initialization, selection, crossover, and mutation operators in order to obtain an evolved solution.

Statistical Inference

Statistics provides a solid theoretical foundation for the problem of data analysis. Through hypothesis validation and/or exploratory data analysis, statistical techniques give asymptotic results that can be used to describe the likelihood in large samples. The basic statistical exploratory methods include such techniques as examining distribution of variables, reviewing large correlation matrices for coefficients that meet certain thresholds, and examining multidimensional frequency tables.

Rule Induction

Rule induction models belong to the logical, pattern distillation based approaches of data mining. Based on data sets, these techniques produce a set of if-then rules to represent significant patterns and create prediction models. Such models are fully transparent and provide complete

explanations of their predictions. One commonly used and well-known type of rule induction is the family of algorithms that produce decision trees.

Data Visualization

"Seeing" the Data: Data are difficult to interpret due to its overwhelming size and complexity. In order to achieve effective data mining, it is important to include people in the data exploration process and combine the flexibility, creativity, and general knowledge of people with the enormous storage capacity and computational power of today's computers. Data visualization is the process of analyzing and converting data into graphics, thus taking advantage of human visual systems. large number of variables while still presenting useful information.

Applications of data mining in finance

Prediction of the Stock Market

Investors in the market want to maximize their returns by buying or selling their investments at an appropriate time. Since stock market data are highly time-variant and are normally in a nonlinear pattern, predicting the future trend (i.e., rise, decrease, or remain steady) of a stock is a challenging problem. The dominant data mining technique used in stock market prediction so far is neural network modeling, including back-propagation (BP) networks, probabilistic neural networks, and recurrent neural networks The basic assumption is that similar input time series should produce similar output time series while ignoring intra-day fluctuations compared to regression models with a back-propagation network using the same data for stock prediction. Results showed that back-propagation network was a better predictor.

Portfolio Management

Portfolio management is a major issue in investment. It concerns how individuals decide which securities to hold in investment portfolios and how funds should be allocated among broader asset classes, such as stocks versus bonds, and domestic securities versus foreign securities. The primary goal is to choose a set of risk assets to create a portfolio in order to maximize the return under certain risk or to minimize the risk for obtaining a specific return.

Bankruptcy Prediction

Predicting bankruptcy is of great benefit to those who have some relations to a firm concerned, for bankruptcy is a final state of corporate failure. In the 21st Century, corporate bankruptcy in the world has reached an unprecedented level. It results in huge economic losses to companies, stockholders, employees, and customers, together with tremendous social and economical cost to the nation.

Foreign Exchange Market

Foreign Exchange is the simultaneous buying of one currency and selling of another. The foreign exchange market is the largest financial market in the world, with a daily average turnover of over US$1 trillion. Data mining has been applied to identifying such technical trading rules. The neural networks were trained using more than 21 years of data to predict 1-day future spot rates for several nominal exchange rates, and achieved 58% accuracy for trading the British Pound and 57% accuracy for trading the German Mark.

Fraud Detection

Credit card transactions continue to grow, taking an ever larger share of the U.S. payment system and leading to a higher rate of stolen account numbers and subsequent losses by banks. According to Meridian Research, financial institutions lost more than US$1 billion in credit and debit card fraud in 2001. Therefore, fraud detection is becoming a central application area of data mining, which aims at searching for patterns indicative of fraud. Improving fraud detection is essential to reducing the loss and maintaining the viability of the payment system.

Data Mining in Other Financial Applications

In addition to the above applications that are discussed, above data mining techniques have also been applied to other financial applications such as loan risk analysis and payment prediction, mortgage scoring and real estate services. Data mining systems can determine whether or not a customer will be able to pay off their loans based on his/her income, age, and historical credit information, etc. Neural networks have been used for providing recommendations to grant or deny a loan based on financial ratios, past credit ratings, and loan records.

BP Neural Network

Artificial neural network is a large broad network with a number of processing units (neurons) connected. It is an abstract, simplified and simulation to human brain, and reflects the basic characteristics of the human brain. Generally, the neural network is the multi-layered network topology, including the input layer, hidden layer and output layer. In dozens of neural network models that were put forward, researchers often use the Hopfield network, BP network[2].

Hopfield network is the most typical feedback network model, it is one of the models which are most commonly studied now. The Hopfield network is the monolayer constituted by the same neuron, and is also a symmetrically connected associative network without learning function. It can implement the restriction optimization and associative memory BP network is the back-propagation network. It is a multi-layer forward network, learning by minimum mean square error. It is one of the most widely used networks. It can be used in the field of language integration, identification and adaptive control, etc.

BP network is semi supervised learning. First of all, artificial neural network needs to learn a certain learning criteria, and then it can work. Guidelines for e-learning (Electronic Learning) can be listed as below. If the result yielded by network is wrong, then the network should reduce the possibility of making the same mistake next time through learning.

Back propagation network with the following indicator can be used to develop the system which can be used for financial forecasting.

Economic indicator used as input to the BP neural network

Moving average (MA)

Moving average line is a statistical mean, which sums up stock price of certain days and gives out an average, and then connects them into a line to observe the price trend. The function of moving average is obtaining the average cost during a certain period, and we can use the average cost curve and the movement of daily closing price changes in the line analysis of the change of bull and bear to study and to determine possible changes in stock.

Random indicator (KDJ)

There are a total of three lines standing for random indicators in the stock, namely K line, D line and the J line. Random indicator not only considers the highest price, the lowest price in the calculation period, but also takes into account of the random amplitude in the course of the fluctuation of stock price. Therefore, researchers always think that random indicator can more truly reflect the volatility of stock price, and it plays an important role in prompting.

$K=2Kt^{-1} / 3+RSV/3$

$D=2\ Dt^{-1} /3+ K/3$

$RSV=100(Cn-Ln)(Hn-Ln)$.

In the formula: C represents the daily closing price means the lowest price that day; H stands for the highest price that day.

J = 3K - 2D

Moving Average Convergence/Divergence (MACD)

The principle of MACD is to use the functions of the signs for aggregation and separation of fast moving average and slow moving average, in addition to double smoothing operation in order to study and determine the timing of buy and sell.

Relative Strength Index (RSI)

Relative Strength Index is an indicator comparing the average of closing high and the average of closing low .We can use it to analyze the market and strength in order to forecast the future of the market trend.

RSI = [average of increase / (average of increase + average of decline)] *100

OnBalanceVolume (OBV)

OBV is the degree of the active investors in the stock market. If there are a lot of buyers and sellers, the stock prices and the volume of trade will rise, the atmosphere of the stock market is warm, commonly known as the bull market; if there are a few of buyers and sellers, the stock prices and the volume of trade will decline. It can be seen that the impact of the rise and fall of stock prices and the volume of trade is the OBV of stock. The volume of

shares and stock price can also reflect the degree of the rise and fall of popularity of stock.

BIAS

BIAS is the ratio between the application index and the moving average. Base on BIAS, it is possible to observe the degree which the stock price deviate from the moving average price to decide to buy or sell.

RSI = [average of increase / (average of increase +average of decline)] *100

Increase scope

Increase scope= (the stock market closing price of today the stock market opening price of today) / the stock market opening price of today.

In order to avoid the situation that the dimensions of larger data have larger influence in the results than that of smaller data normalized the original data.

The formula of Normalization is as follows:

$$A_{ij}\text{-Min } (A_j) \,/\, \text{Max } (A_j)\text{-Min } (A_j)$$

The data were normalized to between 0 and 1.

Analysis

Even though different exiting techniques can be used for financial forecasting, but BP neural network will be the best to develop a software tool that can be used for financial forecasting.

REFERENCES

1. Stock data analysis based on BP neural network Zhou Yixin (software engineering department of international college of Qingdao University) QingDao, ChinaJie Zhang.Information Engineering College of Qingdao (2010)
2. Discovering Golden Nuggets: Data Mining in Financial Application by Dongsong Zhang and Lina Zhou-IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, VOL. 34, NO. 4
3. Fengjing Shao, Zhong qing yu. Principle and algorithm of data mining. Sinohydro Press (2008)
4. Huadong Qin. Based on neural network forecasting stock market trends. Chinese dissertation database (2009).
5. Zhijun Peng. Some new data mining method and its application in Chinese securities market. Chinese dissertation database (2009).