A Novel Data Mining Neural Network

Dr. Tony K.Y. Chan, Eng Chong Tan, Neeraj Haralalka

German Abstract

Unter der Annahme, dass sich die Preisentwicklung auf dem Wertpapiermarkt nicht zufällig sondern chaotisch verhält, sind gewinnbringende Prognosen denkbar. Dies führte zur Entwicklung einer Anwendung, die eine gezielte Datensuche und Analyse von ausgewählten Wertpapieren innerhalb eines Zeitraums ermöglicht. Die durch Graphen veranschaulichte Auswertung liefert dem Benutzer neue Informationen über den Markt, die anderen nicht zur Verfügung stehen. Den Kern der Anwendung bildet ein modifiziertes neuronales Netzwerk, das als Bollinger Band Crossover Supervised Network (BBCSN) bekannt ist.

Introduction

There is justifiable scepticism about the idea that it is possible to make money by predicting price changes in a stock market based only on its past behaviour and a number of publicly available indicators. This scepticism is based on a variety of reasons; many of which are explained by the efficient market hypothesis. The hypothesis states that the market follows a random walk that cannot be predicted by past prices. Any chance of potential profits is snapped up immediately, removing the opportunity almost as soon as it is created, and certainly before the technical analyst can view it. The hypothesis relies on perfect knowledge, implying that perfect prediction technology would only serve to enforce conditions in which that technology is useless.

However, there is a possible way around the efficient market hypothesis because it relies on the public availability of market information. If prices do not follow a random walk, but a chaotic one, then anybody who is able to model the price structure and make valid predictions using the model will have access to information that is not available to others. The efficient market hypothesis will not apply to that person until everyone gains access to the same technology and things even out once more. This presents that person with a well-designed and capable model with tremendous opportunities for profit. This forms a major attraction to develop good models, which can analyse the chaotic stock series, and construct a predictive model.

We aim to research and develop an application that will provide the user with a suitable technique to perform data mining and to predict the stock series of their choice, and present them with the results of the project in the form of a visualization graph for their own analysis. The data-mining algorithm poses significant research opportunities since there are various approaches available and each of them has its own advantages and disadvantages.

Neural networks are probably the most common data mining technique, sometimes synonymous with data mining. They are simple models of neural interconnections in brains, adapted for use on digi-

Figure 1: Prediction of JadeTech by the BBCSN network

tal computers. They learn from a training set and generalize patterns inside it for classification and prediction. They can be applied in both undirected data mining and in time series prediction. Back propagation neural networks are most successfully used for time series prediction and provide a good quality and consistent performance.

The network proposed here utilises the Back-propagation algorithm, with modifications to include the temporal factor and the concept of Bollinger Band Crossover in it. This network is known as the Bollinger Band Crossover Supervised Network (BBCSN).

Architecture

Data is central to any data-mining application. It is needed for the proper training and initialisation of the Data Mining Algorithm. Our data, sourced from the Metastock database, is transformed to text file representation for every stock ticker. The text file has the Price (High, Low, and Close) and the Volume of the stock ticker with the corresponding date listed in a tabular fashion. The length of time period for which the data is available is different for different tickers, some ranging for as long as the past seven years while others as less than three years. We have chosen the Intra-Day High price as the representative ticker.

We assume that the nature of the stock data is generally a chaotic one. There are no limits on the ticker price, which can shoot up to great heights or dip down to the zero level, all within a very short period of time. To handle this, the data in the input series is normalized to a value between zero and one, allowing the data mining algorithm to comprehend the data more intelligently and make valid deductions from the input series. The algorithm involves a well-known and tested temporal back-propagation algorithm using Input Delayed Neural Networks (IDNN). To improve performance of this network, the concept of Bollinger Band Crossover is used in combination with the supervised network and this network is called the Bollinger Band Crossover Supervised Network (BBCSN). BBCSN is an adaptation of the popular Multi-layer Feed Forward Networks which are commonly used as benchmarks for determining the performance of any other network. In our implementation, the output layer comprises of a single neuron, the output of which is the prediction of the network for the next time instant. The hidden layers have twenty and fifteen neurons respectively. All the layers use unipolar sigmoidal function as the transfer function.

The input is in the form of an ordered series, which has the intra-day high price of the stock ticker to be predicted. The data mining algorithms discussed in the previous section do not process the full series at one go. They process the series in a sequence of patterns, while its length is determined by the Window Size of the time series.

Results

Figure 1 shows the graph of the actual data and the predicted data against time (approximately 5 years) of the stock ticker JadeTech as predicted by the supervised BBCSN Network. As the graph shows, the predicted values are very close to the actual values. It seems, this network is able to predict the values quite accurately.

Figure 2: Portfolio of JadeTech by the BBCSN network

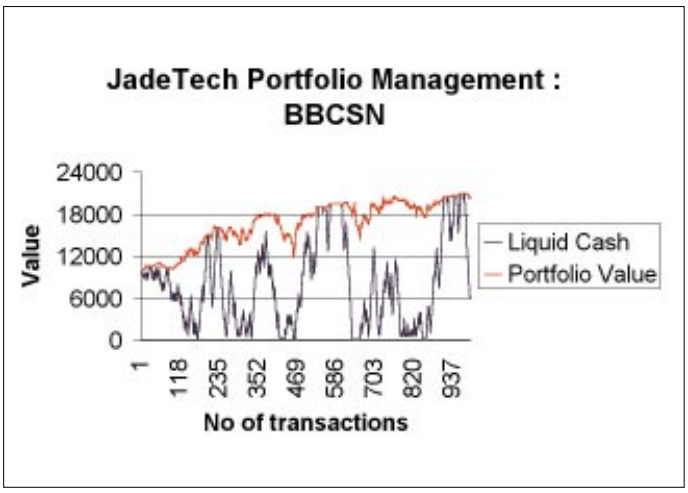

Figure 2 shows the portfolio of JadeTech, built using the supervised BBCSN Network. The portfolio starts off with an initial value of 10,000 in 'Cash'. Buying and selling is done in over approximately 1000 transactions. The resulting final Portfolio is valued at approximately 21,000 thus resulting in a profit of 11,000. Cash at the end is approximately 12,000 thereby signifying that about 9,000 is still invested in stocks. It has to be noted that during the 1000 transactions, the active buying and selling of stock was going on almost constantly as can be seen in the figure.

The Supervised BBCSN network performs with a reasonable level of accuracy in terms of prediction. However, it can be thrown off balance if the input series is very chaotic in nature. This algorithm shows very good performance when it comes to predicting the temporal series and using intelligent investment strategies to benefit from the knowledge gained ahead of time. More work is needed to establish a level of confidence in the technique.

Point of contact

Dr. Tony K. Y. Chan
Centre for Advanced Media Technology, Singapore
Email:
tonychan@camtech.ntu.edu.sg