An Innovative Use of Historical Data for Neural Network Based Stock Prediction

Tak-chung Fu*, Tsz-leung Cheung, Fu-lai Chung and Chak-man Ng

Abstract

—Using artificial neural network is a common approach for the stock time series prediction problem. Unlike variety of researches that focus on selecting different indicators, network training, network architecture, etc., we are focusing on the selection of appropriate time points from the time sequence to serve as the input of the neural network prediction system for dimensionality reduction. We propose to select the time points based on data point importance using perceptually important point identification process. The empirical result shows that the proposed method generally outperformed the traditional method using uniform time delay.

1. INTRODUCTION

In stock time series, future moving trend is suggested to be predictable according to the previous price movement. The study of such prediction of time series stock data is called technical analysis. Generally, technical analysis is built on the fundamental premise that prices move in trends that persist and are predictable to the investor [1]. The financial analyzer must read through the information about a stock price as it comes in and determine if the stock price is continuing on its trend or suddenly changing direction. Therefore, using Artificial Neural Network (ANN) to process the available indictors (e.g. opening, closing, high, low, volume) for prediction is commonly adopted.

Researches on neural network based stock prediction are focused on different prospective like feature selection [2], network training [3], network architecture [4] or even network performance [5].

However, most approaches use the indicators in all the time points within the manipulating period to form the input vector. For example, using 30 daily time points for monthly prediction or 24 hourly time points for daily prediction [6]. It can be easily being seen that the dimension increases highly for a longer period. It is infeasible to retrieve all indicator of each time unit (e.g. hourly, daily) for a long time sequence since the increasing complexity of the system.

The simplest way is to sample the time points from the time sequence in fixed interval [7]. It can be considered as a uniform time delay method. However, this over simplified method cannot guarantee to obtain the characteristic points of the time sequence for the prediction process.

In this paper, instead of studying on the selection of different indictors as the features or the architecture of the ANN, we are focusing on the determination of the suitable time points to serve as the features of the prediction process. Based on the concept of data point importance, a guideline of selecting which time points (e.g. days) as the input features for the neural network-based prediction is proposed. Also, this concept supports prediction on different investigation periods (e.g. long-term and short-term) without changing the architecture of the ANN (i.e. same number of input features). The paper is organized into five sections. The feature selection process is presented in section 2. Section 3 describes the ANN model used in this paper. The preliminary results are reported in section 4 and the conclusion is made in the final section.

* Corresponding Author: Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (e-mail: cstcfu@comp.polyu.edu.hk).

2. TIME POINT SELECTION FOR PREDICTION

In this section, the data point importance evaluation scheme is described first for selecting the appropriate time points to serve as the input for prediction. The preparation step for the input of neural network is then introduced.

A. Data Point Importance

In view of the importance of salient points in stock time series, the identification of Perceptually Important Points (PIP) in time series is first introduced in [8] and used for pattern matching of technical (analysis) patterns in financial applications. The frequently used stock patterns are typically characterized by a few salient points. For example, the head and shoulders pattern consists of a head point, two shoulder points and a pair of neck points. These points are perceptually important in the human identification process and should be considered as having higher importance and identified as PIPs. The proposed scheme follows this idea by identify n PIPs from the sequence P of m length. The PIPs are located according to the steps in Fig.1.

```
Procedure PIPIdentification (P,n)
   Input: sequence P[1..m], integer n
   Output: sequence SP[1..n]
Begin
   Set SP[1]=P[1], SP[n]=P[m]
   Repeat until SP[1..n] are all filled
   Begin
      Select point P[j] with maximum distance to the
      adjacent points in SP (SP[1] and SP[n]
      initially)
      Add P[j] to SP
   End
   Return SP
End
```

Fig. 1. The PIP identification process

The distance measurement depicted in Fig.2 is the vertical distance between the test point p_3 and the line connecting the two adjacent PIPs, i.e.,

$$VD(p_3, p_c) = |y_c - y_3| = \left| \left(y_1 + (y_2 - y_1) \cdot \frac{x_c - x_1}{x_2 - x_1} \right) - y_3 \right| \quad (1)$$

where $x_c = x_3$. This measurement intends to capture the fluctuation of the sequence and the highly fluctuated points would be considered as PIPs while the PIPs may bias to a small segment compare to the whole sequence.

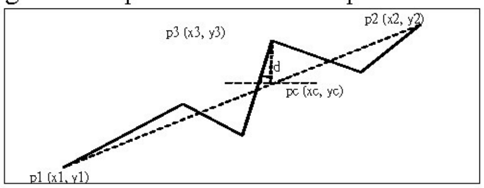

Fig. 2. Vertical distance measure: PIP-VD

B. Dataset Preparation

In order to serve as the input of the ANN for the training and prediction process, preprocessed dataset of the time series data has to be obtained. Given a stock time series of m length, sliding window is used to convert the time series to a set of subsequences to serve as the input records. The size of the window w controls the period for prediction, a larger w for long-term investigation and visa versa. Therefore, totally m-w+1 subsequences will be formed (see the bottom of Fig.3).

For each subsequence, n time points are identified to reduce the dimension of each record. The fundamental method is using uniform time delay approach to sample the time points. However, it may not be a reasonable method as the time points selected by this method may not show the characteristic of the subsequence. Therefore, the PIPs identified by the proposed method are the better choice for the time point selection. It is because the PIPs identified are the data points have high importance according to the shape of that subsequence. Each PIP should more or less contribute to the overall shape of that subsequence and capture the trends in the subsequence. Therefore, using PIPs as the features in the subsequence is expected to be more meaningful (i.e. it can also be viewed as capturing the shape of each subsequence to serve as the input record). The normalized value of each time point would be used as an input node in the input layer of the neural network. As an example, during the short-term investigation in Fig.3, the vertical dot lines are the seven time points selected by using the uniformed time delay method. Obviously, the shape of the latest subsequence cannot be captured well. On the other hand, the seven time points which label "s" are the PIPs identified. A head and shoulders pattern is captured.

Furthermore, the PIP identification process can facilitate to identify the same number of time points across different periods (i.e. long-term vs. short term). Again, as shown in Fig.3, both short-term and long-term investigations in this example captured a head and shoulders pattern represented by 7 time points (i.e. short-term labeled by "s" and long-term labeled by "l").

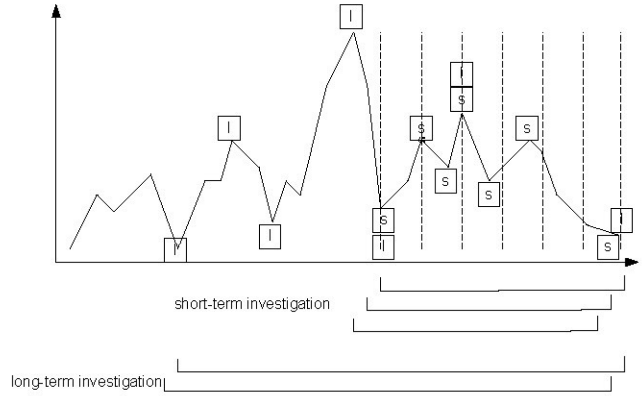

Fig. 3. Example of the input formation for the prediction process

In contrast, expected output of the network illustrates the moving trend of the stock pricing value while the trend measures the moving direction from the last point of each subsequence to the coming trade day after the subsequence and it can be represented by binary method. The moving trend of the subsequence is described in binary number form. There are two binary number nodes controlling the moving trend: uptrend node and downtrend node.

Furthermore, previous day trend can be also considered as extra input nodes of the ANN. Previous day trend describes the moving trend of last day in the subsequence regarding the immediate previous (second last) day record. This idea can illustrate how the previous day moving trend would influence the network training performance for the prediction of moving trend of the coming trade day.

To sum up, each subsequence's pattern, previous day trend and expected output trend, are considered as a record to serve as an input of the ANN.

3. NEURAL NETWORK BASED PREDICTION

The prepared records then will be served as the input of a traditional Artificial Neural Network (ANN) for the training and prediction processes. In the training module, it involves MultiLayer Perceptron (MLP), random sampling of dataset and the main core of this module: backpropagation neural network training. After completing the training process, the network configuration setting and weighting will be saved for stock prediction afterward.

A. Architecture

A structure of MultiLayer Perceptron (MLP) has to be created before the training process. MLP is a feedforward neural network trained with the error backpropagation, which is a two-pass weight-learning algorithm to adjust the weights in between nodes continually in order to reduce error of the output [9]. In general, the architecture of MLP is consisted of input layer, hidden Layer and output Layer. The number of nodes in the input layer and output layer should have the same data dimension as in the preprocessed dataset and it can have one or more hidden layers of various numbers of computational nodes in network setting.

As shown in Fig. 4, an example of MLP is created. 5 time points is used in each record and moving trend in previous day is considered. From the dimension of the dataset, there should be 7 nodes in the input layer and 2 nodes in the output layer. Moreover, there are 2 hidden layers in the network setting and each layer has 10 and 20 computational nodes correspondingly. The MLP architecture is (7-10-20-2).

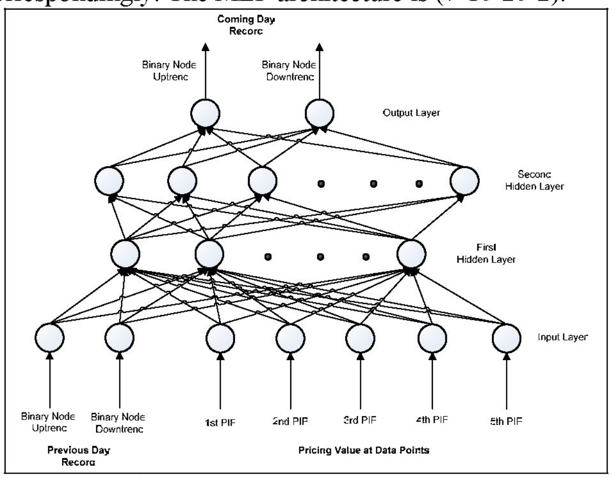

Fig. 4. A MLP architecture of (7-10-20-2)

After the structure of MLP is created, each node weight is first initialized randomly. Furthermore, there are several parameters that will affect the performance of the network training and testing: sigmoid curve flatness, learning rate and momentum rate.

B. Training

During the training process (Fig.5), training records are sequentially presented to the network. These inputs are propagated forward layer by layer until a final output is computed. By comparing the final output to the target output of the corresponding sample, error rate is computed from the sum of Mean Squared Error (MSE) of each output node.

```
Initialize the weight Wij for each node in MLP
   Loop Start and Break when average MSE < expected
   error
      Process all the case patterns in Training Set
      Compute the gradient of the average MSE by
      Cross Validation Set
      Calculate the weight change of each □Wij
      Update the Wij by □Wij
   End of Loop
End
```

Fig. 5. Pseudo code of the neural network training

Di are desired outputs and *Yi* are computed outputs. This error acts as feedback for the training of the weight of nodes in a backpropagation direction. During the training of the network, the minimization of the error function is carried out using a gradient descent learning technique [10]. It starts at a random point in the weight space and moves "downhill" until an error (or "loss") function $L(W)$ is minimized.

During the training, the updated weights of the network for each iteration n are obtained by calculating the partial derivative of the error function in relation to each weight w_{ij}, which gives a direction of steepest descent. The delta rule (Eqt.2) shows how learning rate (η) and momentum rate (α) influence the amount of weight change $\triangle w_{ij}$ based on the gradient direction:

$$\Delta w_{ij}(n) = -\eta \frac{\partial E(n)}{\partial w_{ij}(n)} + \alpha \Delta w_{ij}(n-1) \qquad (2)$$

C. Feedforward Neural Network Prediction

With similar function of feedforward neural network training, each new record is to be fed forward in the MLP for prediction purpose. After the network feedforward mechanism, an error rate of the prediction is calculated. This error is a performance index of the creditability of the stock prediction process.

The moving trend of each new record can then be predicted, in comparing with real output values of the record, a prediction accuracy can then be measured as a performance index of the stock prediction.

4. PERLMINARY RESULTS

Based on the proposed time point selection approach, a stock time series moving trend prediction system is developed using Java. The neural network engine in the system is adopted from [11]. The performance of the proposed PIP-based time point selection method will be compared to the traditional uniform time delay method in term of accuracy. Six Hong Kong Hang Seng Index (HSI) constituent stocks are selected from three stock categories: Banking, Public Affair and Telecommunication. Due to the space limit, only part of the results is reported in this paper. However, similar results are obtained for those stock datasets. For the trade day value, daily closing price is selected since it is usually viewed as a reference stock price by the public.

In the evaluation section, if without special declaration, the default structure of MLP has 2 hidden layers. For each parameter setting of the MLP: sigmoid curve flatness = 1.0, learning rate = 0.1 and momentum rate = 0.5. The dataset sampling ratio of (training set: cross validation set: testing set) is set as (0.6, 0.2, 0.2). Each stock time series subsequence, with parameters of time trend = 30 and number of identified data point = 10, is prepared from a 1-year-long time series sequence from 1st Jan to 31st Dec. Moreover, the expected error is set default as 0.49.

A. 4.2.1 Comparison of Time Point Selection Method

First, the accuracy of the proposed PIP-based time point selection method is compared to the traditional uniform time delay method. As shown in table 1 and 2, the prediction results (i.e. accuracy) using the proposed method on two stocks from different categories are outperformed the traditional method. In the tables, the accuracy of 60% or higher is bolded while the better result between the two methods is highlighted.

TABLE 1. THE PREDICTION ACCURACY ON A BANKING STOCK FROM 1994 TO 1998

	PIP-based	Uniform Time Delay
1994	**0.641**	0.581
1995	**0.605**	**0.605**
1996	**0.614**	0.545
1997	0.579	0.558
1998	0.591	0.568

TABLE 2. THE PREDICTION ACCURACY ON A TELECOMMUNCATION STOCK FROM 1994 TO 1998

	PIP-based	Uniform Time Delay
1994	0.562	0.581
1995	**0.604**	0.452
1996	**0.642**	0.545
1997	**0.638**	**0.612**
1998	0.542	0.515

B. Comparison of Different Subsequence Lengths and Numbers of Time Point Selected

In this subsection, the performance in different subsequence lengths and using different numbers of time point is studied. The result from a telecommunication stock in 1997 is shown in table 3. As shown in the table, a longer subsequence length can generally obtain better prediction result. In other words, a prediction based on a longer period (long-term investigation) is better.

On the other hand, the number of time point used for the prediction process seems not affected greatly by using 10 or more time points. That means in most of the cases, 10 time points can already capture the shape of the time sequence for prediction. Moreover, as the number of input feature is reduced from the original length of the subsequence to the number of time point, the speed for both training and prediction process can be greatly reduced.

As a benchmarking, table 4 shows the same matrix but the traditional uniform time delay method is adopted. By comparing the result, our proposed method outperformed the traditional method except in one case. It is believed that the good performance of the proposed method over traditional one is due to the PIPs identified from the sequence can capture the shape of the subsequences' patterns. Prediction based on these patterns is similar to the traditional technical analysis method and is confirmed to be effective.

TABLE 3. THE PREDICTION ACCURACY ON A TELECOMMUNCATION STOCK, NUMBER OF TIME POINT SELECTED VS. SUBSEQUENCE LENGTH USING PIP-BASED METHOD

Length \ No. of pt.	5	10	15
200	**0.64**	**0.670**	**0.652**
100	0.587	**0.667**	**0.639**
50	**0.6**	**0.642**	**0.600**
30	0.567	**0.659**	0.523
20	0.502	0.553	0.509

TABLE 4. THE PREDICTION ACCURACY ON A TELECOMMUNCATION STOCK, NUMBER OF TIME POINT SELECTED VS. SUBSEQUENCE LENGTH USING UNIFORM TIME DELAY METHOD

Length \ No. of pt.	5	10	15
200	**0.624**	**0.623**	**0.600**
100	0.524	0.583	**0.612**
50	0.546	0.531	0.575
30	0.467	0.531	0.475
20	0.523	0.483	0.527

V. CONCLUSION

Instead of selecting the time points with fixed interval to serve as the input of a stock neural network prediction system, we proposed to select the characteristic points from the stock time series based on data point importance. The data point importance is evaluated by the perceptually important point identification process which is a method simulates how human identify points from a time series data. Experiments show that the proposed approach generally outperformed the traditional uniform time delay method. We are now working on evaluating the performance of using different features to serve as the input of the prediction network. That is, instead of using closing price only, opening, high, low, volume or even the combination of them retrieved from the selected time points can be served as the input for prediction.

REFERENCES

[1] R. Romeu and U. Serajuddin: Technical Analysis for Direct Access Trading. Direct Access Trader Series (2001)

[2] R. J. van Eyden: The Application of Neural Networks in the Forecasting of Share Prices. Finance and Technology Publishing (1996)

[3] C. Klimasaushas Applying neural network. In Neural networks in Finance and Investing. Probus Publishing Company (1993) 47-72

[4] T. Kimoto, K. Asahawa, M. Yoda and M. Takeoka: Stock market prediction system with modular neural networks. Proc. of the Int'l Joint Conf. on Neural Networks, Vol.1 (1990) 1-6

[5] Y. Yoon and G.Wsales: Predicting stock price performance: A neural network approach. In Neural Networks in Finance and Investing. Probus Publishing Company (1993) 329-342

[6] S.C. Kuo, S.T. Li, Y.C. Cheng and M.H. Ho: Knowledge Discovery with SOM Networks in Financial Investment Strategy. Proc. of the 4th HIS (2004) 98-103

[7] G. Tsibouris and M. Zeidenberg: Neural networks as an alternative stock market model. In Neural Networks in the Capital Markets. John Wiley and Sons (1995) 127-136

[8] F.L. Chung, T.C. Fu, R. Luk, V. Ng: Flexible Time Series Pattern Matching Based on Perceptually Important Points. In: The 17th IJCAI Workshop on Learning from Temporal and Spatial Data (2001) 1-7

[9] G. Thimm, P. Moerland and E. Fiesler: The Interchangeability of Learning Rate and Gain in Backprogpagation Neural Networks. Neural Computation, Vol.8, No.2 (1996)

[10] M. Moreira and E. Fiesler: Neural Networks with adaptive Learning Rate and Momentum Terms. IDIAP Technical Report (1995)

[11] A. Gurel: Feed Forward Neural Networks – A JAVA Implementation v2.0. http://aydingurel.brinkster.net/neural (2006)