Prediction of the Next Stock Price using Neural Network for Data Mining

Hiroshi Takaho, Takayuki Arai, Tsuyoshi Otake and Mamoru Tanaka

Sophia University
7-1 kioi-cho, chiyoda-ku, Tokyo, Japan
Phone:+81-3-3238-3878, Fax:+81-3-3238-3321
Email: koguma@mamoru.ee.sophia.ac.jp

Abstract— Adding the dimension of time to databases produces time series databases (TSDB) and introduces new aspects and difficulties for data mining, knowledge discovery and prediction of sample points. In this paper, we introduce the method for the prediction of the next sample point with multi-layer neural network[1] in TSDB. Predicting the next sample point in TSDB includes cleaning and filtering the time series data, identifying the most important predicting attributes, and extracting a set of association rules that can be used to predict the time series behavior in the future. Our method is based on signal processing techniques, and TSDB for the closing price of a stock are used as an example.

1. Introduction

In most knowledge discovery problems and prediction of time series data in futures behavior in the future, it is very difficult without the process to time series data. A regular static (non-series) database contains a set of records, each constructed from a set of attributes. The ordering of records in a static database has no significance, at least from the data mining aspect[2]. In contrast, a time series database (TSDB) contains a set of records in which some of the attributes are associated with a time-stamp.

A TSDB can be a stock market database in which each record includes, besides the static attributes such as stock name and sector, some dynamic attributes such as closing price and transaction volume, which pertain to a specific day. Other types of time series databases include online monitoring systems, sales transactions in a grocery store, and web usage data.

A major difference between static and dynamic (time series) databases lies in the information carried by each attribute. Unlike static databases, where each attribute in a record is independent of attributes in other records, in TSDB some attributes are meaningful only as part of a time segment. In a stock market database, we are more interested in the behavior of the stock over time rather than in its value on a single day. The behavior is due to several components like a periodic, aperiodic components and white noise in TSDB. They make it difficult to predict the next sample point and extract the regulations.

In this paper, we introduce the method for the prediction of the next sample point with multi-layer neural network in TSDB. The process of the prediction of the next sample point in TSDB includes cleaning and filtering of TSDB using the low pass filter and band pass filter. The divided components are applied to inputs of our system.

2. Filtering process using LPF

To clean the data, one can use a low-pass filter (LPF) operator, which eliminates the high frequency waves (which are mostly noise) and leaves only those with low frequency (mostly long-term signal).

There are several LPF operators, in the time and frequency domains. One of the simplest is the finite impulse response (FIR) form[3]. In stock-market data, a simple moving average filter, which is one of the FIR low-pass filters, is often used. The filter is defined as

$$y_L[n] = \frac{1}{L}\sum_{k=0}^{L-1} x[n-k] \tag{1}$$

where

$y_L[n]$	output signal;
$x[n]$	input signal;
L	the days to average;

Fig.1 is TSDB for the stock closing price of the Nikkei 225 futures from 4th Jan '91 to 30th Jan '97. We can find the high frequency components which move quickly in several days and the low frequency components which move slowly and are often called trend line.

In prediction of the next stock price, the former is important. The high frequency components are obtained after the original signal is divided by the processed signal with a simple moving average filter. We think the

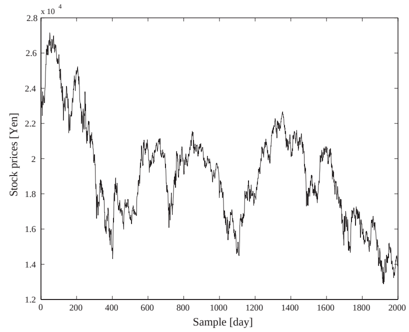

Figure 1: TSDB for the stock closing price of the Nikkei 225 futures from 4th Jan '91 to 30th Jan '97.

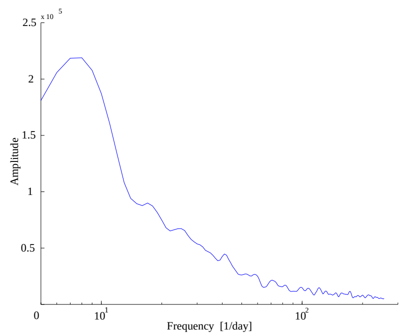

Figure 2: This figure describes frequency characteristic of the Nikkei 225 futures from 4th Jan '91 to 30th Jan '97. We removed the direct current component.

division process method is better than the subtraction process method because the former includes normalization.

3. Filtering process using BPF

We describe our proposal method for extracting the profitable components with band-pass filter (BPF) in this section. The profitable components may be signals which largely affect TSDB. Fig.2 is a frequency characteristic of Nikkei 225 Futures. We can find some peaks in Fig.2, which means that the TSDB include the each frequency components more.

In our proposed method, the more included frequency components in TSDB are extracted using BPF. We propose that the BPF is designed with a moving average filter. The sequences $y_p[n]$ which include the components more is obtained as

$$y_p[n] = \frac{y_{L_1}[n]}{y_{L_2}[n]}, \qquad L_1 < L_2. \tag{2}$$

Table 1: The combination of the optimum L_1 and L_2 using BPF

frequency [1/day]	0.017	0.024	0.040	0.051	0.057
L_1	9	10	14	33	72
L_2	11	13	18	42	80

We also applied the division process for the reasons noted in section 2.

This paragraph describes the method of selecting L_1 and L_2 using the energy ratio to extract the profitable and specific components. In this paper, energy ratio is expressed by amounts of one specific component $Y[\omega_1]$ and amounts of the entire components as

$$E[L_1, L_2] = \frac{Y_{L_1 L_2}[\omega_1]}{\sum_{\omega=0}^{\pi} Y_{L_1 L_2}[\omega]} \tag{3}$$

and combinations of the optimum L_1 and L_2 are obtained by $\arg\max E[L_1, L_2]$. In Table.1, the result of selecting L from TSDB in Nikkei 225 Futures from 4th Jan '91 to 30th Jan '9 is indicated.

4. Simulation

In section 2 and 3, we described the method of filtering the TSDB. In this section, we perform simulations to predict the ceiling rate, which is an index which indicates the buying or selling signal, with multi-layer neural network (in this paper, the former is 1 and the latter is 0, we omit how to find the value of the ceiling rate.).

The ceiling rate is used as a text signal and output. In the input signal, the processed TSDB of the closing price of Nikkei 225 Futures are applied and the simulations are performed in three patterns of the input which is applied to 60 sequences in all patterns. In the first pattern, they are applied to the 60 sequences in TSDB processed with only LPF (Fig.3). Before this work, we could find that the prediction model using $L = 3$ moving average filter as LPF (without BPF) was suitable for prediction. In the second pattern, 55 sequences processed with LPF and 5 sequences processed with BPF are done (Fig.4). In the third pattern, 55 sequences processed with LPF and 5 sequences, which is white noise (WN), are done. Table.2 describes the parameter of neural network and Table.3 describes the result of simulation (20 simulations are performed and the result is the mean of MSE.).

In the cases of using LPF only and using both LPF and BPF, their prediction models are profitable enough

Table 2: The parameters in this simulation are noted.

input	processed TSDB of the closing price of Nikkei 225 futures
number of inputs	60
number of hidden nodes	30
output (text signal)	the ceiling rate
training method	back propagation
the term for training	from 4th Jan '91 to 30th Jan '97
the term for test	from Feb '97 to Feb Jan '98

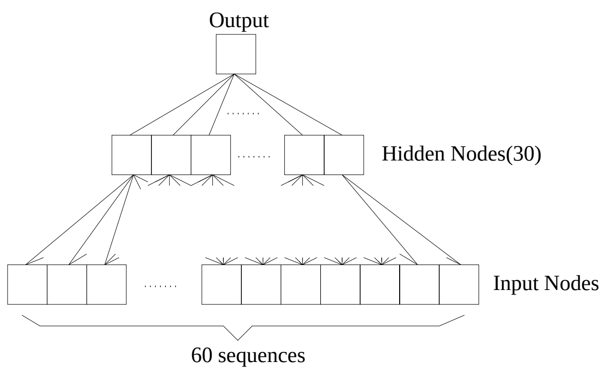

Figure 3: The multi-layer neural network model, 60 sequences are applied to the inputs which are TSDB processed with LPF only.

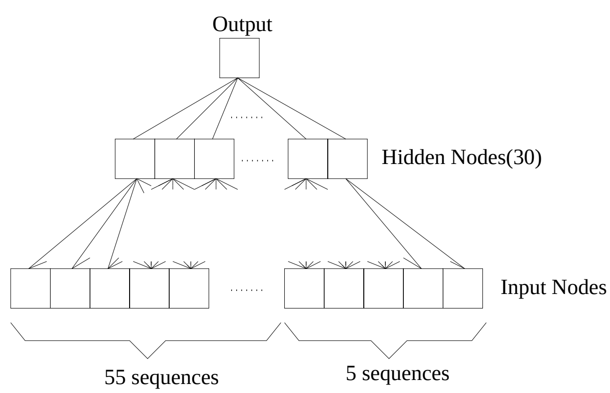

Figure 4: The multi-layer neural network model, 55 sequences are applied to the inputs which are TSDB processed with LPF and 5 sequences are applied to the inputs which are TSDB processed with BPF or white noises.

to trade. The model using both LPF and BPF is more accurate than any other model. The model using both LPF and WN is not good one and the components extracted with BPF are profitable to predict.

Table 3: This table describes MSE (mean square error) in simulation TSDB and test TSDB.

	MSE
inputs using LPF only	0.0532
inputs using both LPF and BPF	0.0523
inputs using both LPF and WN	0.0571

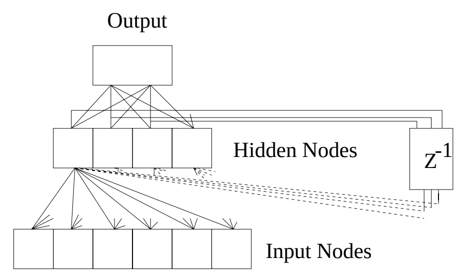

Figure 5: This figure desicribes Elman nueral network which is one of recurrent neural networks. Elman nueral network is a recurrent network with feedback from each hidden node to all hidden nodes.

5. Conclusions

In this paper, we are able to conclude that the method divided into periodic and aperiodic components is more appropriate than the non-divided method. The more accurate prediction model is obtained. However, an another filtering and normalization process should be performed. In this work, we designed some filters, but the results of their simulation are not good. The reason may be that to extract the more specific component only, we increased the orders of filters. The longer delays of information that is most important to predict are brought about by its process so that the model may not be able to predict accurately. We might have to design the filter to take into account the trade-off between amplitude response and delays.

Another neural model, especially a reccurent neural network[4][5], should also be designed. In prediction and extracting the regulations using TSDB, such models are frequently used. We are very interested in applying the TSDB signal processed in our method to inputs of the model.

References

[1] K.Taji, T.Miyake and H.Tamura, “On error back-propagation algorithm using absolute error function,” in Proc Systems, Man, and Cybernetics, 1999. IEEE SMC ’99 Conference Proceedings. IEEE International Conference on, vol.5, 1999

[2] M. Last, Y. Klein and A. Kandel,“Knowledge discovery in time series databases,”Systems, Man and Cybernetics, Part B, IEEE Transactions on, vol.31 Issue.1,

[3] L. Rabiner and C. Rader, Digital Signal Processing, IEEE press, 1972

[4] C. W. Omlin and C. L. Giles, “Extraction of Rules from Discrete-time Recurrent Neural Network,” in Proc. of Neural Network, vol.9, No.1, pp.41-52, 1996

[5] C. Morita and H. Tsukimoto, “Knowledge discovery from numerical data,” in Proc. of Knowledge-based Systems, vol.10, No.7, pp.413-419

[6] S.Lawrence, C.L.Giles and S.Fong, “Natural language grammatical inference with recurrent neural networks,” in Proc. of Knowledge and Data Engineering, IEEE Transactions on, Vol.12 Issue.1, 2000.2.

[7] R. Setiono and H. Liu, “Understanding Neural Networks via Rule Extraction,” in Proc. of The 14th International Joint Conference on Artificial Intelligence, pp.480-485, 1995