2011 Conference on Technologies and Applications of Artificial Intelligence

Application of Bayesian Network for Nikkei Stock Return Prediction

Yi Zuo
Graduate School of Information Science
Nagoya University
Nagoya, Japan

Masaaki Harada
Graduate School of Information Science
Nagoya University
Nagoya, Japan

Takao Mizuno
Graduate School of Information Science
Nagoya University
Nagoya, Japan

Eisuke Kita
Graduate School of Information Science
Nagoya University
Nagoya, Japan
Email: kita@is.nagoya-u.ac.jp

***Abstract*—This paper describes the stock price return prediction algorithms by using Bayesian network. In the first algorithm, the clustering algorithm transforms the stock price return distribution to the discrete values set. The Bayesian network gives the probabilistic graphical model that represents previous stock price returns and their conditional dependencies via a directed acyclic graph. The network is applied for the stock price return prediction. The second algorithm uses, in addition to the previous stock price return, the prediction error data of the first algorithm for determining the Bayesian network. Finally, two algorithms are compared with the time-series prediction algorithm in NIKKEI stock return prediction.**

***Keywords*-Bayesian network; NIKKEI stock average; Stock return; Time-series prediction:**

I. Introduction

Several time-series prediction algorithms have been applied for the stock price prediction [1]–[3]; Auto Regressive (AR) model, Moving Average (MA) model, Auto Regressive Moving Average (ARMA) model and Autoregressive Conditional Heteroskedasticity (ARCH) model. After that, several prediction models are developed from the ARCH model; e.g., Generalized AutoregRessive Conditional Heteroskedasticity (GARCH) model [4], Exponential General AutoRegressive Conditional Heteroskedastic (EGARCH) model [5] and so on.

The time-series forecast algorithms represent the error distribution according to the normal distribution. Recent studies in econophysics point out that the distribution of the stock price fluctuation does not follow the normal distribution [6]. The algorithm based on the normal distribution may not forecast the stock price accurately. Therefore, the stock price forecast by using Bayesian Network [7], [8] is presented in this study. In the present algorithms, the use of the Bayesian network defines the conditional dependencies between the previous stock prices in order to prediction of the future stock price return.

Two algorithms are presented in this study. In the first algorithm, the stock price distribution is discretized by using the clustering algorithms and then, the Bayesian network defines the stochastic dependencies between the discrete value set of the previous stock prices. The network is directly used for the future stock price prediction. In the second algorithm, the prediction error of the first algorithm is also used for the stock return prediction. Finally, the algorithms are applied for the NIKKEI stock price return prediction in order confirm their accuracy.

The remaining part of the manuscript is as follows. In section II, time-series prediction algorithms are explained briefly. Bayesian network algorithm is described in section III and the present algorithms are explained in sections IV and V. Numerical results are shown in section VI. The results are summarized again in section VII.

II. Time-series Prediction Algorithms

A. AR Model [1], [2]

The notation r_t denotes the stock price return at time t. In AR model AR(p), the return r_t is approximated with the previous return r_{t-i} $(i = 1, \cdots, p)$ and the error term u_t as follows

$$r_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i} + u_t \qquad (1)$$

where α_i $(i = 0, \cdots, p)$ is the model parameter. The error term u_t is a random variable from the normal distribution centered at 0 with standard deviation equal to σ^2.

B. MA Model

In MA model MA(q), the stock price return r_t is approximated with the previous error term u_{t-j} $(j = 1, \cdots, q)$ as follows

$$r_t = \beta_0 + \sum_{j=1}^{q} \beta_j u_{t-j} + u_t \qquad (2)$$

978-0-7695-4601-8/11 $26.00 © 2011 IEEE
DOI 10.1109/TAAI.2011.41

IEEE computer society

where β_j $(j = 0, \cdots, q)$ is the model parameter.

C. ARMA Model

ARMA model is the combinational model of AR and MA models. In ARMA model ARMA(p, q), the stock price return r_t is approximated as follows.

$$r_t = \sum_{i=1}^{p} \alpha_i r_{t-i} + \sum_{j=1}^{q} \beta_j u_{t-j} + u_t \tag{3}$$

D. ARCH Model

In ARCH model ARCH(p, q), the stock price return r_t at time t is approximated as follows

$$r_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i} + u_t. \tag{4}$$

The error term u_t is given as follows

$$u_t = \sigma_t z_t \tag{5}$$

where $\sigma_t > 0$ and the function z_t is a random variable from the normal distribution centered at 0 with standard deviation equal to 1.

The volatility σ_t^2 is approximated with

$$\sigma_t^2 = \beta_0 + \sum_{j=1}^{q} \beta_j u_{t-j}^2. \tag{6}$$

E. Determination of Model Parameters

In each model, the model parameters p and q are taken from $p = 0, 1, \cdots 10$ and $q = 0, 1, \cdots, 10$. Akaike's Information Criterion (AIC) is estimated in all cases. The parameters p and q for maximum AIC are adopted.

III. Bayesian Network

A. Conditional Probability Table

The Bayesian network represents the conditional dependencies among a set of random variables with a directed acyclic graph.

If the random variable x_i depends on the random variable x_j, the variable x_j and x_i are called as a parent and a child, respectively. Their dependency is represented with $x_j \rightarrow x_i$. If more than one parents exist for the child x_i, the notation $Pa(x_i)$ denotes the parents set for x_i. Conditional dependency probability between x_j and x_i is represented with $P(x_i|x_j)$, which means the conditional probability of x_i given x_j.

The strength of relationships between random variables is quantified with the conditional probability table (CPT). The notation Y^m and X^l denote the m-th state of $Pa(x_i)$ and the l-th state of x_i, respectively. The conditional probability table is given as follows.

$$\begin{matrix} P(X^1|Y^1), P(X^2|Y^1), & \cdots, & P(X^L|Y^1) \\ \vdots & & \vdots \\ P(X^1|Y^M), P(X^2|Y^M), & \cdots, & P(X^L|Y^M) \end{matrix}$$

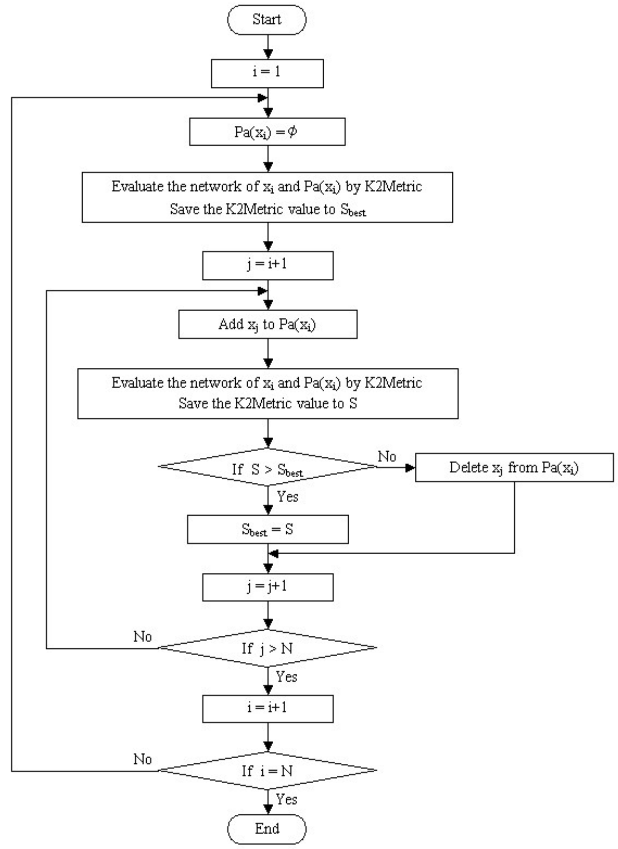

Figure 1. K2 algorithm

where M and L are total numbers of the states of $Pa(x_i)$ and x_i, respectively.

B. Determination of Graph Structure

In this study, the Bayesian networks are determined by using K2 algorithm [7], [8] with K2Metric [7]–[9] as the estimator of the network.

K2Metric is given as follows [9], [10].

$$K2 = \prod_{i=1}^{N} \prod_{j=1}^{M} \frac{(L-1)!}{(N_{ij} + L - 1)!} \prod_{k=1}^{L} N_{ijk}! \tag{7}$$

where

$$N_{ij} = \sum_{k=1}^{L} N_{ijk}. \tag{8}$$

The notation N, L, and M denote total number of nodes, total numbers of states for x_i and $Pa(x_i)$, respectively. Besides, the notation N_{ijk} denotes the number of samples of $x_i = X^k$ when $Pa(x_i) = Y^j$.

K2 algorithm determines the network from the totally ordered set of the random variables.

The K2 algorithm is illustrated in Fig.1 and summarized as follows.

1) $i = 1$
2) Set the parents set $Pa(x_i)$ for the note x_i to an empty set.
3) Estimate K2 metric S_{best} of the network composed of x_i and $Pa(x_i)$.
4) For $j = i+1, \cdots, N$,
   a) Add x_j to $Pa(x_i)$.
   b) Estimate K2 metric S of the network composed of x_i and $Pa(x_i)$.
   c) Delete x_j from $Pa(x_i)$ if $S < S_{best}$.
5) $i = i + 1$
6) Go to step 2 if $i \leq N$.

C. Probabilistic Reasoning

When the evidence e of the random variable is given, the probability $P(x_i|e)$ is estimated by the marginalization with the conditional probability table [11].

The probability $P(x_i = X^l|e)$ is given by the marginalization algorithm as follows.

$$P(x_i = X^l|e) = \frac{\sum_{j=1, j\neq i}^{N} \sum_{x_j=X^1}^{X^L} P(x_1, \cdots, x_i = X^l, \cdots, x_N, e)}{\sum_{j=1}^{N} \sum_{x_j=X^1}^{X^L} P(x_1, \cdots, x_N, e)} \tag{9}$$

where the notation $\sum_{x_j=X^1}^{X^L}$ denotes the summation over all states $X^1, X^2, \cdots X^L$ of the random variable x_j.

IV. Prediction Algorithm 1

A. Process

The process of the prediction algorithm 1 is summarized as follows.

1) The stock price return is discretized according to the Ward method.
2) The Bayesian network B is determined by the set of the discretized stock prices.
3) The stock price return is predicted by using the network B.

B. Discrete Value Set of Stock Price Return

The stock price return r_t is defined as follows

$$r_t = (\ln P_t - \ln P_{t-1}) \times 100 \tag{10}$$

where the notation P_t denotes the closing stock price at time t.

When the stock price return is transformed into the set of some clusters with the Ward method, the notation C_l and c_l denote the cluster and its center.

The discrete value set of the stock price return

$$\{r^1, r^2, \cdots, r^L\}, \tag{11}$$

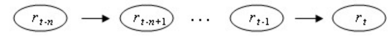

Figure 2. Total order of stock price returns

is given as follows

$$\{r^1, r^2, \cdots, r^L\} = \{c_1, c_2, \cdots, c_L\} \tag{12}$$

where the notation L denotes the total number of the discretized values.

C. Ward Method

Ward method defines clusters so that the Euclid distances from samples to the cluster centers are minimized. The notation z, C_i and c_i denote the sample, the cluster and its center, respectively. The estimator is given as

$$\begin{aligned} D(C_i, C_j) &= E(C_i \cup C_j) \\ &\quad - E(C_i) - E(C_j) \end{aligned} \tag{13}$$

$$E(C_i) = \sum_{z \in C_i} d(z, c_i)^2 \tag{14}$$

where the notation $d(z, c_i)$ denotes the Euclid distance between z and c_i.

D. Stock Price Return Prediction

For determining the network B by K2 algorithm, the total order of the random variable sets is necessary. The stock price return are totally ordered according to the order of their time-series (Fig.2).

Once the network B is determined, the stock price return r_t is predicted so as to maximize the probability $P(r^l|B)$:

$$r_t = \arg\max_{r^l}(P(r^l|B)) \tag{15}$$

V. Prediction Algorithm 2

A. Process

The process of the prediction algorithm 2 is summarized as follows.

1) The prediction algorithm 1 described in section IV is applied for predicting the previous stock price returns which have been already observed.
2) The prediction error of the prediction algorithm 1 is estimated from the results in the step 1).
3) The prediction error distribution is transformed into the discrete value set by the Ward method.
4) The Bayesian network B' is determined from the stock price return and the prediction error.
5) The future stock price return is predicted by using the network B'.

The discretization of the prediction error and the determination process of the Bayesian network are summarized as follows.

Figure 3. Total order of stock return and prediction error

Table I
DISCRETE NUMBER VERSUS AIC ON STOCK RETURN

Discrete number L	AIC
2	2.0442
3	1.7830
4	1.6066
5	1.5193
6	1.4597
7	1.8168
8	1.7478
9	1.8126
10	1.7729

B. Prediction Error Discretization

When the stock price return at the time t and its prediction value estimated from equation (15) are given as $\bar{r}_t$ and r_t, respectively, the prediction error u_t is defined as follows

$$u_t = r_t - \bar{r}_t. \tag{16}$$

The prediction error is transformed into the discretized values set by using Ward method in section IV-C.

When the discretized prediction error and its total number are given as u^l and L, respectively, the prediction error set is defined as follows.

$$\{u^1, u^2, \cdots, u^L\} \tag{17}$$

The Akaike's information criterions (AICs) are estimated for different value of the total number L. The total number L for the smallest AIC is adopted.

C. Stock Price Return Prediction

For determining the network B by K2 algorithm, the total order of the random variable sets is necessary. The stock price return are totally ordered according to the order shown in Fig.3. Then, the Bayesian network B' is determined by the algorithm shown in section III.

The stock price return is predicted according to the maximum likelihood estimation in the network B' as follows.

$$r_t = \arg\max_{r^l}(P(r^l|B')) \tag{18}$$

where the notation $P(r^l|B')$ denotes the probability when $r_t = r^l$ at the network B'.

VI. NUMERICAL EXAMPLE

A. Network Determination

The stock price data from Feb. 22th 1985 to Dec. 30th 2008 are used for determining the Bayesian network.

Table II
DISCRETE NUMBER SET OF STOCK RETURN

Cluster	$(C_l)_{\min}$, $(C_l)_{\max}$	$c_l(r^l)$
C_1	[-16.138%, -3.000%)	-4.30%
C_2	[-3.000%, -0.730%)	-1.48%
C_3	[-0.730%, -0.065%]	-0.37%
C_4	(-0.065%, 0.947%]	0.39%
C_5	(0.947%, 3.800%]	1.73%
C_6	(3.800%, 13.235%]	5.45%

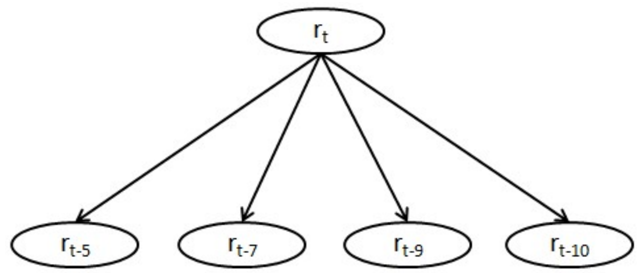

Figure 4. Bayesian network determined with NIKKEI stock average return only

1) Prediction Algorithm 1: The Bayesian networks are determined for $L = 2, 3, \cdots, 10$ and then, their AICs are compared in Table I. We notice that AIC is smallest at $L = 6$. In this case, the clusters are listed in Table II. The parameter $c_l(r^l)$ denotes the cluster center (discrete value). The parameters $(C_l)_{\min}$ and $(C_l)_{\max}$ are the minimum and the maximum values included in the cluster C_l. The Bayesian network determined from the above data set is shown in Fig.4. The figure denotes that the stock return r_t at the time t depends on the 5-day prior return r_{t-5}, 7-day prior return r_{t-7}, 9-day prior return r_{t-9} and 10-day prior return r_{t-10}.

2) Prediction Algorithm 2: Next, the network is determined from the stock return and the prediction error. The discrete number of the stock price return is fixed to be $L = 6$. Then, the discrete number of the prediction error distribution is specified to be $L = 2, 3, \cdots, 10$. AICs are compared in Table III. We notice that the AIC is smallest at $L = 6$. In this case, the clusters are listed in Table IV. The Bayesian network determined from the above data set is shown in

Table III
DISCRETE NUMBER VERSUS AIC ON STOCK RETURN PREDICTION ERROR

Discrete number L	AIC
2	1.6659
3	1.4587
4	1.5764
5	1.5132
6	1.3203
7	1.4596
8	1.4599
9	1.4602
10	1.4606

Table IV
DISCRETE NUMBER SET OF STOCK RETURN PREDICTION ERROR

Cluster	$(C_l)_{\min}$, $(C_l)_{\max}$	$c_l(u^l)$
C_1	[-16.67%, -2.87%)	-4.13%
C_2	[-2.87%, -0.92%)	-1.59%
C_3	[-0.92%, -0.00%]	-0.47%
C_4	(0.00%, 1.40%]	0.55%
C_5	(1.40%, 3.71%]	2.26%
C_6	(3.71%, 16.74%]	5.33%

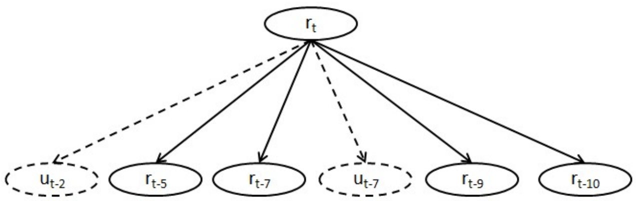

Figure 5. Bayesian network determined with stock return and error

Fig.5. The figure denotes that the stock return r_t at the time t depends on the 5-day prior return r_{t-5}, 7-day prior return r_{t-7}, 9-day prior return r_{t-9}, 10-day prior return r_{t-10}, 2-day prior error u_{t-2} and 7-day prior error u_{t-7}.

B. Comparison of Prediction Algorithms

The networks are employed for predicting the stock price return from January 1st to March 31st, 2009. The prediction accuracy and the correlative coefficient for the actual stock price are compared with the time-series prediction algorithms in Tables V and VI, respectively. The label AR(2), MA(2), ARMA(2,2) and ARCH(2,9) denote the AR model with $p = 2$, MA model with $q = 2$, ARMA model with $p = 2$ and $q = 2$ and ARCH model with $p = 2$ and $q = 9$, respectively. Table V illustrates that the prediction algorithm 2 shows the smallest maximum error and the smallest average error. Table VI illustrates that the prediction algorithm 2 shows the largest correlative coefficient.

C. CPU Time

For Bayesian network determination and the stock return prediction, CPU times of BN-1 and BN-2 are compared in Table VII. The results show that the network determination takes 529 seconds in BN-1 and 908 seconds in BN-2, respectively, and that the prediction takes 0.691 seconds in BN-1 and 0.887 seconds in BN-2, respectively. We notice that Bayesian network determination is very time-consuming process and CPU time of BN-2 is almost 1.7 times that of BN-1.

Table V
COMPARISON OF PREDICTION ACCURACY

	Maximum	Minimum	Average
AR(2)	5.7648	0.0604	2.4923
MA(2)	5.7746	0.0758	2.4953
ARMA(2,2)	5.9389	0.0076	2.5150
ARCH(2,9)	5.8342	0.0178	2.5116
Prediction Algorithm 1	6.2948	0.0057	2.9905
Prediction Algorithm 2	4.9016	0.0105	2.3467

Table VI
COMPARISON OF CORRELATIVE COEFFICIENT

AR(2)	0.9278
MA(2)	0.9276
ARMA(2,2)	0.9250
ARCH(2,9)	0.9268
Prediction Algorithm 1	0.9099
Prediction Algorithm 2	0.9375

Table VII
COMPARISON OF CALCULATION COST

	Network Search	Prediction
BN-1	529sec	0.691sec
BN-2	908sec	0.887sec

VII. CONCLUSIONS

The stock price return prediction algorithms by using Bayesian network were presented in this study. The stock price return distribution is discretized by clustering with the Ward method. Bayesian network models the stochastic dependency among previous return distribution. The prediction accuracy and the correlation coefficient with respect to the actual stock price are compared with the traditional time-series prediction algorithms such as AR, MA, ARMA and ARCH models in the NIKKEI stock average prediction. The prediction algorithm 1 and the prediction algorithm 2 show 26% and 30% better than the time-series prediction algorithms.

REFERENCES

[1] G. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control.* Prentice-Hall, 1994.

[2] P. Brockwell and R. Davis, *Time Series: Theory and Methods.* Springer, 2009.

[3] R. F. Engle and V. K. Ng, "Measuring and testing the impact of news on volatility," *Journal of Finance*, vol. 48, no. 5, pp. 1749–78, 1993.

[4] T. Bollerslevb, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, pp. 307–327, 1986.

[5] D. B. Nelson, "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica*, vol. 59, pp. 347–370, 1991.

[6] H. Takayasu, *Practical Fruits of Econophysics.* Springer, 2006.

[7] J. Pearl and S. Russell, "Bayesian networks," in *Handbook of Brain Theory and Neural Networks*, M. A. Arbib, Ed. MIT Press, 2002, pp. 157–160.

[8] I. Ben-Gal, "Bayesian networks," in *Encyclopedia of Statistics in Quality and Reliability*, F. Ruggeri, R. S. Kennett, and F. W. Faltin, Eds. John Wiley & Sons, 2007, vol. 1.

[9] D. Heckerman, D. Geiger, and D. Chickering, "Learning Bayesian networks: The combination of knowledge and statistical data," *Machine Learning*, vol. 20, no. 3, pp. 197–243, 1995.

[10] G. F. Cooper and H. E., "A Bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 9, no. 4, pp. 309–347, 1992.

[11] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Representation and Reasoning Series*. Morgan Kaufmann, 1988.