Bayesian Network Based Prediction Algorithm of Stock Price Return

Yi Zuo, Masaaki Harada, Takao Mizuno, and Eisuke Kita

Abstract. This paper describes the stock price return prediction using Bayesian network. The Bayesian network gives the probabilistic graphical model that represents previous stock price returns and their conditional dependencies via a directed acyclic graph. When the stock price is taken as the stochastic variable, the Bayesian network gives the conditional dependency between the past and future stock prices. In the present algorithm, the stock price return distribution is transformed to the discrete values set by using Ward method, which is one of the clustering algorithms. The Bayesian network gives the conditional dependency between the past and future stock prices. The stock price is determined from the discrete value set of the stock prices so that its occurrence probability is maximized. Finally, the present algorithm is compared with the traditional time-series prediction algorithms in the TOYOTA motor corporation stock price prediction. The present algorithm show 20% better than the time-series prediction algorithms.

1 Introduction

For predicting the stock price, several time-series prediction algorithms have been studied by many researchers[1, 2, 3, 4, 5]. Autoregressive (AR) model[1, 2], Moving Average (MA) model[3], Autoregressive Moving Average (ARMA) model[4] and Autoregressive Conditional Heteroskedasticity (ARCH) model[5] are very popular time-series prediction algorithms. Recently, the extensions of ARCH model have been presented; Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model[6] and Exponential General Autoregressive Conditional Heteroskedastic (EGARCH) model[7].

Yi Zuo · Masaaki Harada · Takao Mizuno
Graduate School of Information Science, Nagoya University

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

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 1, SIST 15, pp. 397–406.
springerlink.com

The time-series forecast algorithms represent the future stock price by the linear combination of the past stock prices and the error term following to the normal distribution. Recent studies in econophysics point out that the stock price distribution does not follow the normal distribution[8]. Therefore, the error term modeled according to the normal distribution may not predict the stock price accurately. Therefore, the stock price prediction by using Bayesian Network is presented in this study. In the present algorithm, the stock price is transformed to the discrete values set by using Ward method, which is one of the clustering algorithms. The Bayesian network[9, 10] gives the probabilistic graphical model that represents previous stock price returns and their conditional dependencies via a directed acyclic graph. The future stock price can be predicted by using the network. The stock price is determined from the discrete value set of the stock prices so that its occurrence probability is maximized. Finally, the present algorithm is applied for prediction of the TOYOTA Motor Corporation stock price. The results are compared with the time-series prediction algorithm.

The remaining part of the manuscript is as follows. In section 2, time-series prediction algorithms are explained briefly. Bayesian network algorithm and the present algorithm are explained in sections 3 and 4. Numerical results are shown in section 5. The results are summarized again in section 6.

2 Prediction Algorithms

2.1 Time-Series Prediction

2.1.1 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

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

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.

2.1.2 MA Model[3]

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

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

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

2.1.3 ARMA Model[4]

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.

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

2.1.4 ARCH Model[5]

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

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

2.1.5 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)[11] is estimated in all cases. The parameters p and q for maximum AIC are adopted.

2.2 Bayesian Network Model

In the time-series algorithms, the stock price distribution is assumed to be according to the normal distribution. Recent studies, however, point out that the distribution of the stock price fluctuation does not follow the normal distribution[8]. Figure 1 shows the stock price return of TOYOTA Motor Corporation. This figure is plotted with the P/E ratio as the horizontal axis and the data distribution as the vertical axis. The bar chart and the solid line denote the actual stock price distribution and the normal distribution. We notice from this figure that the actual distribution is very far from the normal distribution and especially, the standard deviation around $\pm\sigma$ and $\pm 3\sigma$ is conspicuous. As a result, the normal distribution may not forecast the stock price accurately. For overcoming this difficulty, Bayesian network is applied for the stock price forecast in this study.

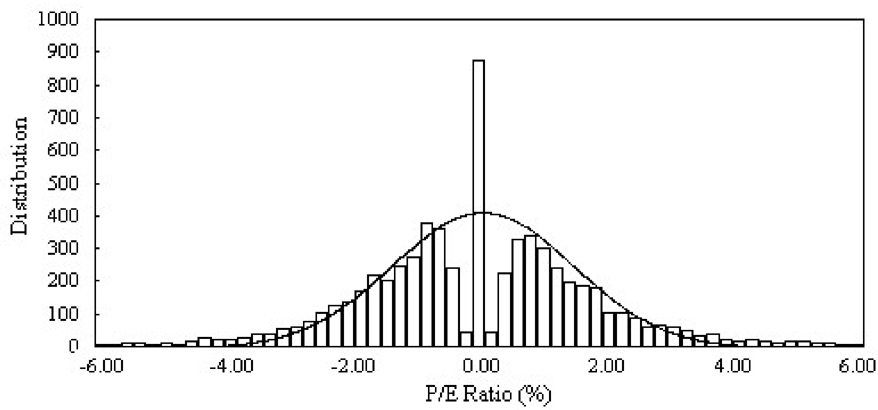

Fig. 1 P/E ratio frequency distribution of TOYOTA motor corporation stock price

The Bayesian network represents the conditional dependencies of random variables via a directed acyclic graph. The Bayesian network model can predict the stock price without the normal distribution error model. Figure 2 shows the relationship between the AR model, one of the ordinary models, and the Bayesian networkD While, in the AR model, the stock price is approximated with the linear model of the stock prices, the present method gives the nonlinear and stochastic model of the stock prices by using Bayesian network.

3 Bayesian Network

3.1 Conditional Probability Table

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 nodes, respectively. Their dependency is

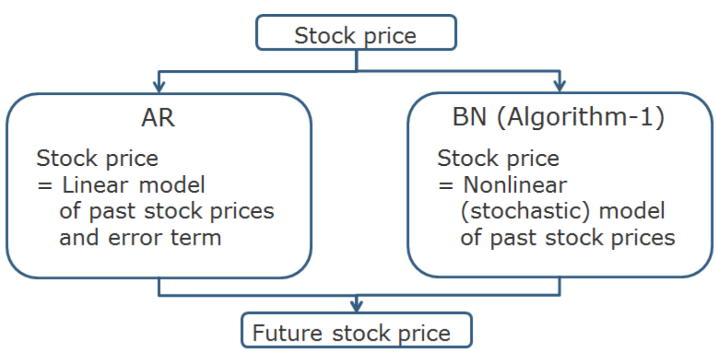

Fig. 2 Comparison of time-series and present algorithms

represented with $x_j \rightarrow x_i$. If the child node x_i has more than one parent nodes, the notation $Pa(x_i)$ denotes the parent node 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.

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

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

3.2 Determination of Graph Structure

In this study, the Bayesian networks are determined by using K2 algorithm[9, 10] with K2Metric[9, 10, 12] as the estimator of the network.

K2Metric is given as follows[12, 13]

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

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. 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 which is 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 $i \leq N$.

3.3 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[14].

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.

4 Prediction Algorithm

4.1 Process

The process of the prediction algorithm 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.

4.2 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 is given as follows

$$\{r^1, r^2, \cdots, r^L\} = \{c_1, c_2, \cdots, c_L\} \tag{11}$$

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

4.3 Ward Method

Ward method[15] 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

$$D(C_i, C_j) = E(C_i \cup C_j) - E(C_i) - E(C_j) \tag{12}$$

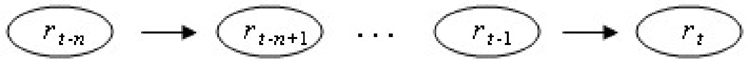

Fig. 3 Total order of stock price returns

$$E(C_i) = \sum_{z \in C_i} d(z, c_i)^2 \tag{13}$$

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

4.4 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.3).

Once the network B is determined, the stock price r_t is determined from the discrete value set of the stock prices so that its occurrence probability $P(r^l|B)$ is maximized.

$$r_t = \arg\max_{r^l} (P(r^l|B)) \tag{14}$$

Table 1 Discrete number versus AIC on forecast error of TOYOTA Motor Corp. stock return

Discrete number L	AIC
2	2.8452
3	2.5892
4	2.3913
5	2.3243
6	2.5309
7	2.7240
8	2.4935
9	2.5941
10	2.5994

5 Numerical Example

TOYOTA motor corporation stock price is considered as an example. Bayesian network B is determined from the TOYOTA motor corporation stock price from February 22nd 1985 to December 30th 2008. Then, the network is applied for predicting the stock price return from January 1st to March 30th, 2009.

5.1 Number of Clusters

We will discuss the effect of the cluster number to the prediction accuracy. The stock price return data are clustered into 2 to 10 clusters by Ward method. The AIC values of the discrete value sets are compared in Table 1. Table shows that the AIC is minimized at the discrete number $L = 5$.

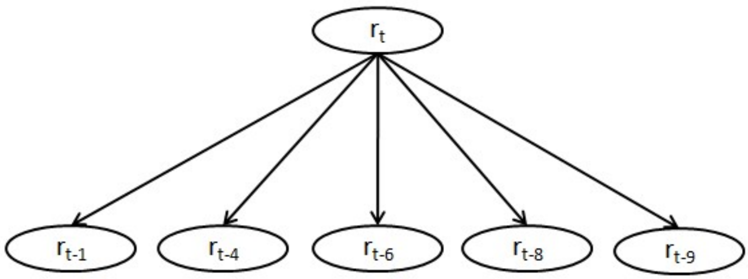

Fig. 4 Bayesian network

5.2 Network Determination

The discrete value sets at $L = 5$ are indicated in Table 2. The parameter $c_l(r^l)$ is the discrete value and the parameters $(C_l)_{\min}$ and $(C_l)_{\max}$ denote the maximum and minimum values among the samples in the cluster C_l. The Bayesian network, which is determined from the data of Table 2, is shown in Fig.4. The number of samples means the number of stock price returns included in the cluster. The notation $c_l(r^l)$ denotes the cluster center; the discrete value. We notice that the return r_t has the dependency on the one-day prior return r_{t-1}, the 4-days prior return r_{t-4}, the 6-days prior return r_{t-6}, the 8-days prior return r_{t-8}, and the 9-days prior return r_{t-9}.

5.3 Comparison of Prediction Accuracy

The prediction accuracy is compared in Table 3. The row label AR(9), MA(6), ARMA(9,6) and ARCH(9,9) denote the results by AR model with $p = 9$, MA model with $q = 6$, ARMA model with $p = 9$ and $q = 6$ and ARCH model with $p = 9$ and

Table 2 Discrete value set of TOYOTA Motor Corp. stock return

Cluster	$(C_l)_{\min}$, $(C_l)_{\max}$	$c_l(r^l)$
C_1	[-21.146%, -3.900%)	-5.55%
C_2	[-3.900%, -1.285%)	-2.10%
C_3	[-1.285%, 0.000%]	-0.47%
C_4	(0.000%, 2.530%]	1.08%
C_5	(2.530%, 16.264%]	4.12%

Table 3 Comparison of prediction accuracy

	Max. error	Min. error	Ave. error	Correlation coefficient
AR(9)	7.5091	0.0615	0.7448	2.6657
MA(6)	7.6259	0.0319	0.7417	2.6859
ARMA(9,6)	7.1204	0.0401	0.7427	2.6739
ARCH(9,9)	8.0839	0.0597	0.7527	2.6992
Bayesian Network	7.8579	0.0415	0.7122	3.1494

$q = 9$, respectively. Besides, the row label "Bayesian network" means the result by the present method.

Table 3 shows that, in the present algorithm, the average error is about 5%-smaller than the others and the correlation coefficient is improved by 10 to 20% against the others.

6 Conclusions

The stock price return prediction algorithm using Bayesian network was presented in this study. The stock price return distribution is discretized by the Ward method. Bayesian network models the stochastic dependency of the stock price returns. The present algorithm was compared with the traditional time-series prediction algorithms such as AR, MA, ARMA and ARCH models in the TOYOTA motor corporation stock price prediction. The present algorithm show, in the average error, about 5% better than the time-series prediction algorithms.

References

1. Box, G., Jenkins, G.M., Reinsel, G.C.: Time Series Analysis: Forecasting and Control. Prentice-Hall (1994)
2. Brockwell, P.J., Davis, R.A.: Time Series: Theory and Methods. Springer (2009)
3. Mills, T.C.: Time Series Techniques for Economists. Cambridge University Press (1990)
4. Box, G., Jenkins, G.M.: Time Series Analysis: Forecasting and Control, 2nd edn., Oakland (1976)
5. Engle, R.F., Ng, V.K.: Measuring and testing the impact of news on volatility. Journal of Finance 48(5), 1749–1778 (1993)
6. Bollerslevb, T.: Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327 (1986)
7. Nelson, D.B.: Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59, 347–370 (1991)
8. Takayasu, H.: Practical Fruits of Econophysics. Springer (2006)
9. Pearl, J., Russell, S.: Bayesian networks. In: Arbib, M.A. (ed.) Handbook of Brain Theory and Neural Networks, pp. 157–160. MIT Press (2002)
10. Ben-Gal, I.: Bayesian networks. In: Ruggeri, F., Kennett, R.S., Faltin, F.W. (eds.) Encyclopedia of Statistics in Quality and Reliability, vol. 1, John Wiley & Sons (2007)
11. Akaike, H.: Information theory and an extention of the maximum likelihood principle. In: Proc. 2nd International Symposium on Information Theory, pp. 267–281 (1973)

12. Heckerman, D., Geiger, D., Chickering, D.: Learning Bayesian networks: The combination of knowledge and statistical data. Machine Learning 20(3), 197–243 (1995)
13. Cooper, G.F., Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data. Machine Learning 9(4), 309–347 (1992)
14. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Representation and Reasoning Series. Morgan Kaufmann (1988)
15. Ward, J.H.: Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58, 236–244 (1963)