

Up/Down Analysis of Stock Index by Using Bayesian Network

Yi Zuo¹ & Eisuke Kita^{1,2}

¹ Graduate School of Information Science, Nagoya University, Nagoya, Japan

² Graduate School of System Informatics, Kobe University, Kobe, Japan

Correspondence: Eisuke Kita, Graduate School of Information Science, Nagoya University, Nagoya 464-8601, Japan. Tel: 81-52-789-3521. E-mail: kita@is.nagoya-u.ac.jp

Received: May 9, 2012 Accepted: June 26, 2012 Online Published: July 11, 2012

doi:10.5539/emr.v1n2p46

URL: <http://dx.doi.org/10.5539/emr.v1n2p46>

Abstract

Bayesian network is the graphical model which can represent the stochastic dependency of the random variables via the acyclic directed graph. In this study, Bayesian network is applied for the up/down analysis of the stock index. The up/down rates of the daily stock indexes in three major markets are taken as the network nodes and then, the network is determined by K2 algorithm with the K2 metric as the prediction accuracy of the network. The present algorithm is applied for predicting the up/down analysis of the daily stock indexes in 2007 and the results are compared with the traditional algorithms; Psychological line and trend estimation, which are popular algorithms which are well-known by the traders. Their accuracy comparison shows that the average correction rate of the present algorithm is almost 60%, which is almost equal or higher than them of the traditional algorithms such as the psychological line (50-59%) and the trend estimation (50-52%). Moreover, the vertical trading results reveal that the profit of the present algorithm is much greater than the others.

Keywords: Bayesian network, stock index, K2 algorithm, K2 metric, psychological line, trend estimation

1. Introduction

Day trading refers to the practice of buying and selling financial instruments within the same trading day such that all positions are usually closed before the market close for the trading day. Traders who participate in day trading are simply called day traders. Some of the more commonly day-traded financial instruments are stocks, stock options, currencies, and a host of futures contracts such as equity index futures, interest rate futures, and commodity futures. In the day trading of the stock indexes, the most important information for the day trader is not the exact future stock price but the up/down information of the future stock price. In this study, Bayesian network is applied for the up/down analysis of the stock index.

Bayesian network is the graphical model which can represent the stochastic dependency of the random variables via the acyclic directed graph (Pearl & Russell, 2002; Ben-Gal, 2007; Heckerman, Geiger & Chickering, 1995). The network takes the random variables as the nodes and can represent their dependency as the arrows. The strength of the dependency between the random variables is estimated according to their concurrent occurrence probability.

In this study, the Bayesian network is applied for predicting the up/down analysis of the stock indexes in three major markets; Dow Jones Industrial Average (Dow, 30), Financial Times Stock Exchange 100 (FTSE100) and Nikkei Stock Average (Nikkei, 225). The up or down of the stock index on the next day is taken as the random variables for the Bayesian network. The network is determined by using K2 algorithm with the K2 metric as the prediction accuracy of the network (Ben-Gal, 2007; Heckerman, Geiger & Chickering, 1995). The network is applied for predicting the up/down analysis of the stock index FTSE100 in 2007 and then, the results are compared with the traditional day-trading strategies such as psychological line and trend estimation.

The stock price forecast by using Bayesian network was presented in the reference (Zu & Kita, 2012). The aim of the study was to predict the daily stock prices by using Bayesian network. The daily stock prices were digitized by the clustering algorithm in order to define the random variables for the network. Since the prediction accuracy depends on the clusters and the data size, huge data are necessary for the good prediction accuracy. Therefore, the daily stock prices from 1985 to 2008 were used in the study. On the other hand, the up and down of the daily stock price are taken as the random variables in this study. Therefore, necessary data set size is relatively small, which is the daily stock prices from 2005 to 2006 in this study.

The remaining part of this paper is organized as follows. In section 2, the study background is described. The Bayesian network algorithm and the present algorithm are explained in section 3 and 4, respectively. Numerical results are shown in section 5. The conclusions are summarized in section 6.

2. Background

2.1 Stock Exchange

A stock exchange is a form of exchange which provides services for stock brokers and traders to trade stocks, bonds, and other securities. Stock exchanges also provide facilities for issue and redemption of securities and other financial instruments, and capital events including the payment of income and dividends. Securities traded on a stock exchange include shares issued by companies, unit trusts, derivatives, pooled investment products and bonds.

Three traditional major stock exchanges are New York Stock Exchange (NYSE), London Stock exchange and Tokyo Stock Exchange.

The New York Stock Exchange is a stock exchange located at New York City, USA. It is by far the world's largest stock exchange and its average daily trading value was approximately US\$153 billion in 2008. One of the popular stock market indexes in NYSE is Dow Jones Industrial Average (Dow30).

The London Stock Exchange is a stock exchange located in the City of London in the United Kingdom. As of December 2011, the Exchange had a market capitalization of US\$3.266 trillion (short scale), making it the fourth-largest stock exchange in the world by this measurement (and the largest in Europe). The main stock market index is Financial Times Stock Exchange 100 index (FTSE100).

The Tokyo Stock Exchange is a stock exchange located in Tokyo, Japan. It is the third largest stock exchange in the world by aggregate market capitalization of its listed companies. It had 2,292 listed companies with a combined market capitalization of US\$3.8 trillion as of Dec 2011. The main stock market index is the Nikkei225 index of companies selected by the Nihon Keizai Shimbun (Japan's largest business newspaper).

2.2 Stock Index Prediction

Traditional algorithms for the up and down analysis of the stock indexes are briefly classified into fundamental and technical analyses (Bianchi, Boyle & Hollingsworth, 1999).

Fundamental analysis involves analyzing financial statements, health, management and competitive advantages, and competitors of the business. When it is applied for predicting a stock, it focuses on the overall state of the economy, interest rates, production, earnings, and management. Since large amount data are necessary for the fundamental analysis, it may not be adequate for the stock index prediction in the short term.

A fundamental principle of technical analysis is that a market's price reflects all relevant information, so their analysis looks at the history of a security's trading pattern rather than external drivers such as economic, fundamental and news events. Price action also tends to repeat itself because investors collectively tend toward patterned behavior. Technical analysis employs models and trading rules based on price and volume transformations, such as the relative strength index, moving averages, regressions, inter-market and intra-market price correlations, business cycles, stock market cycles or, classically, through recognition of chart patterns.

In this study, the Bayesian network is applied for the up and down analysis of the stock indexes and then, the results are compared with psychological line and trend estimation, which are popular technical analysis algorithms.

2.2.1 Psychological Line

The reference value of the psychological line on the day t is defined as follows.

$$\eta_1^t = \frac{\sum_{i=1}^N f(r_{t-i+1} - r_{t-i})}{N} \times 100 \quad (1)$$

$$f(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases} \quad (2)$$

where the parameter N denotes the number of days for estimating psychological line. The term $\sum_{i=1}^N f(r_{t-i+1} - r_{t-i})$ in Eq.(1) estimates the number of days of which closing stock price is higher than that of the previous day. For example, the parameter $\eta_1^t = 80\%$ in case of $N=10$ indicates that the closing stock price rose on eight days of past ten days. If $\eta_1^t = 100\%$ in case of $N = 10$, stock price every days rose during last ten days and thus, investors must consider that the stock price will decrease on the next day.

The psychological line, in this study, gives the following predictions.

- 1) If $\eta_1 < 75$, it is predicted that the stock price rises on the next day.
- 2) If $\eta_1 > 75$, it is predicted that the stock price declines on the next day.

The parameter N is taken as follows.

Case P1) $N = 8$

Case P2) $N = 12$

2.2.2 Trend Estimation

The trend estimation is focusing on their ability to pick up turning points quickly at the end of a series. In this study, the base and the deflector lines are used for picking up turning points.

When the base and the deflector lines are referred to as η_{2b} and η_{2d} , respectively, the trend estimation leads to the following predictions.

- 1) If $\eta_{2b} < \eta_{2d}$, it is predicted that the stock price rises on the next day.
- 2) If $\eta_{2b} > \eta_{2d}$, it is predicted that the stock price declines on the next day.

The values η_{2b} and η_{2d} are defined as follows.

Case T1) The value η_{2b} and η_{2d} are defined as the average values of the highest and the lowest stock prices on the past 26 days and the past 9 days, respectively.

Case T2) The value η_{2b} and η_{2d} are defined as the average values of the highest and the lowest stock prices on the past 42 days and the past 9 days, respectively.

3. Bayesian Network

3.1 Conditional Probability

When the probabilistic variable x_i depends on the probabilistic variable x_j , the relation is represented as

$$x_j \rightarrow x_i \quad (3)$$

where the node x_j and x_i are named as the parent node and the child node, respectively. If the node x_i has multiple parent nodes, the class of the parent nodes is defined as the class $Pa(x_i)$

$$Pa(x_i) = \{x_1, x_2, \dots, x_M\} \quad (4)$$

where the notation x_i and M denote the parent nodes of the node x_i and the total number of the parent nodes, respectively.

The dependency of the child node x_i to the class of parent nodes $Pa(x_i)$ is quantified by the conditional probability $P(x_i|Pa(x_i))$, which is given as

$$P(x_i|Pa(x_i)) = \frac{P(x_i)P(Pa(x_i)|x_i)}{P(Pa(x_i))} \quad (5)$$

where

$$P(Pa(x_0)|x_0) = \prod_{i=1}^M P(x_i, x_0). \quad (6)$$

3.2 K2Metric

The validity of the network is evaluated by K2Metric (Cooper & Herskovits, 1992; Pearl, 1988), which is defined as follows.

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

and

$$N_{ij} = \sum_{k=1}^L N_{ijk} \quad (8)$$

where the parameter 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 $Pa(x_i) = Y^j$.

3.3 K2 Algorithm

The graph structure is determined by K2 algorithm (Cooper & Herskovits, 1992; Pearl, 1988). The K2

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

1. Set $i = 1$.
2. Set a parent node class $Pa(x_i)$ as an empty class ϕ .
3. Estimate K2Metric of the network composed of the node x_i and the class $Pa(x_i)$, which is referred to as S_{best} .
4. Set $j = i + 1$.
5. Add the node x_j to the class $Pa(x_i)$.
6. Estimate K2Metric of the network composed of the node x_i and the class $Pa(x_i)$, which is referred to as S .
7. If $S \leq S_{best}$, remove the node x_j from the class $Pa(x_i)$.
8. If $S > S_{best}$, set $S_{best} = S$.
9. Set $j = j + 1$. If $j \leq N$, goto step 5.
10. Set $i = i + 1$. If $i \leq N$, goto step 3.
11. Define the Bayesian network B as the network composed of the node x_i and the class $Pa(x_i)$.

3.4 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 (Cooper & Herskovits, 1992).

The marginalization algorithm gives the probability $P(x_i = X^l|e)$ as follows:

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

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

4. Present Prediction Algorithm

4.1 Process

The object of this algorithm is to estimate the probability of the up/down of the stock index FTSE100 on the next day by using Bayesian Network. The prediction process is summarized as follows.

1. Estimate the up/down rates in the past stock indexes between the day t and the day $t - 1$.
2. Determine the Bayesian network from the up/down rates in the past stock indexes.
3. Predict the stock price on the next day so that the probability of the up/down rate in the stock index is maximized.

4.2 Stock Price UP/Down Rate

Assume FTSE100 on the day t is referred to as P_t^F , the up/down rate r_t^F on the day t is defined as follows.

$$r_t^F = \begin{cases} 0, & P_t^F \leq P_{t-1}^F \\ 1, & P_t^F > P_{t-1}^F \end{cases} \quad (10)$$

When Dow30 and Nikkei225 on the day t are given as P_t^D and P_t^N , respectively, the related up/down rates r_t^D and r_t^N on the day t are defined as follows.

$$r_t^D = \begin{cases} 0, & P_t^D \leq P_{t-1}^D \\ 1, & P_t^D > P_{t-1}^D \end{cases} \quad (11)$$

$$r_t^N = \begin{cases} 0, & P_t^N \leq P_{t-1}^N \\ 1, & P_t^N > P_{t-1}^N \end{cases} \quad (12)$$

4.3 Prediction of Up/Down Rate

Bayesian network is determined according to the K2 algorithm from the set of the up/down rates of the indexes. The total order relationship of the random variables is defined according to the time order (Figure 2).

Once the Bayesian network B is determined, the probability of the up/down rate in FTSE100 on the next day is defined as $P(r_{t+1}^F|B)$. Therefore, the decision on the FTSE100 improvement on the next day is decided as follows.

$$r_{t+1}^F = \begin{cases} 0, & P(r_{t+1}^F|B) < 0.5 \\ 1, & P(r_{t+1}^F|B) \geq 0.5 \end{cases} \quad (13)$$

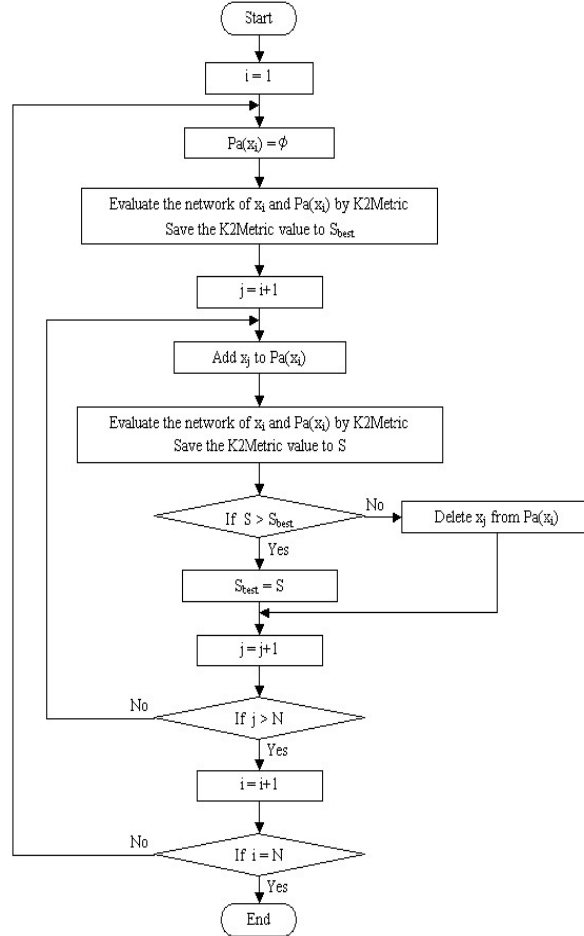


Figure 1. Graph search algorithm

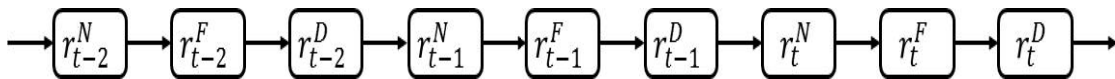


Figure 2. Total order relationship of up/down rates of stock indexes

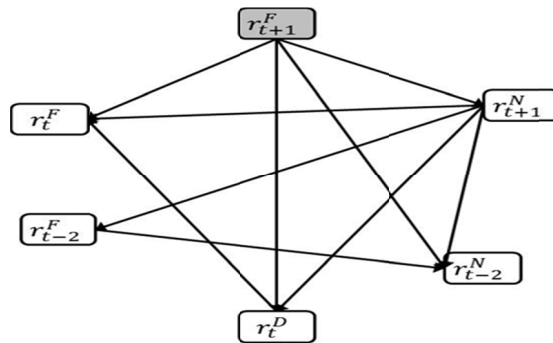


Figure 3. Bayesian network for FTSE100 up/down rate prediction

5. Numerical Examples

5.1 Problem Setting

The stock index data from January 2005 to December 2006 are used for determining the Bayesian network. The stock index from January 2007 to December 2007 is predicted and then, the results are compared with the psychological line and the trend estimation.

The vertical investments are performed on the data from January 2007 to December 2007 according to three algorithms and then, the accuracy and the total profit are compared.

The vertical investment strategies are given as follows.

- (1) If it is predicted on the day $t - 1$ that the stock index on the day t will rise, the specific amount of stocks are bought on the closing time of the day $t - 1$. Then, the same amount of stocks are sold on the closing time of the day t .
- (2) If it is predicted on the day $t - 1$ that the stock index on the day t will decline, the specific amount of stocks are sold on the closing time of the day $t - 1$. Then, the same amount of stocks are bought on the closing time of the day t .

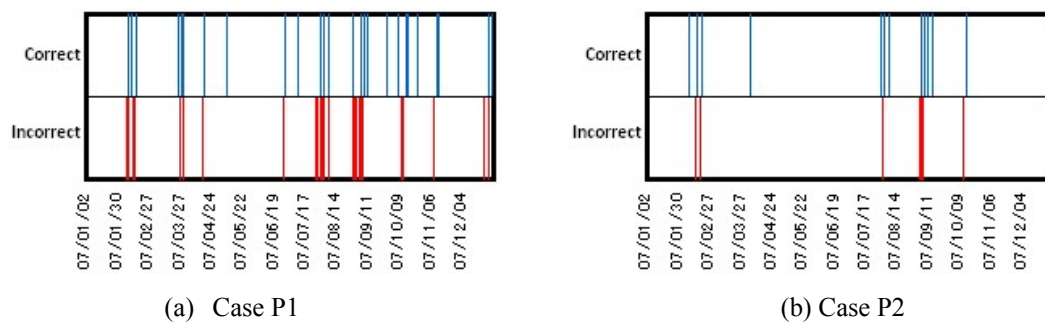


Figure 4. Results of psychological line

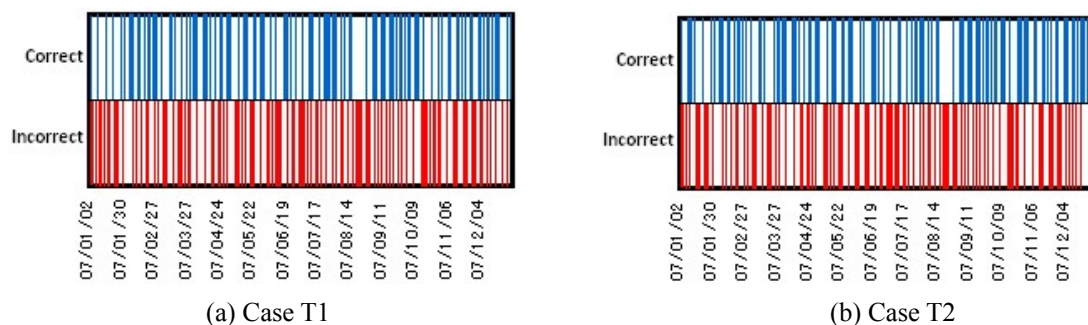


Figure 5. Results of trend estimation

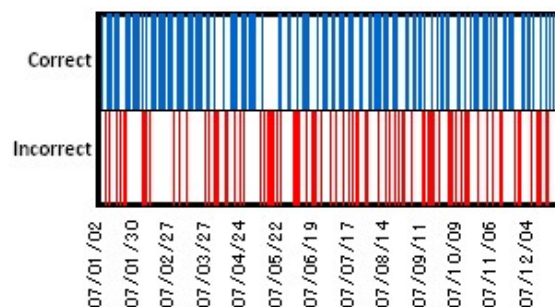


Figure 6. Result of Bayesian network

Table 1. Accuracy comparison

		Number of tradings	Correct answer rate	Total profit
Psychological line	Case P1	54	51.85 %	282.3 GB £
	Case P2	22	59.09 %	534.1 GB £
Trend estimation	Case T1	253	51.38 %	19.9 GB £
	Case T2	253	50.59 %	-91.7 GB £
Bayesian network		249	61.44 %	5139.7 GB £

5.2 Results

The Bayesian network, which is determined from the stock indexes from January 2005 to December 2006, is shown in Fig.3. The figure shows that the up/down rate of FTSE100 r_{t+1}^F is directly related to the FTSE100 r_t^F , the Nikkei225 r_t^N and r_{t-2}^N , and the DOW30 r_t^D . Then, the Nikkei225 up/down rates r_t^N and r_{t-2}^N are related with the FTSE100 up/down rate r_{t-2}^F .

The prediction results are shown in Figs.4, 5 and 6, respectively. The figures are plotted with the date as the horizontal axis and the correct/incorrect of the prediction as the vertical axis, respectively. The number of prediction, correct answer rate and total profit are shown in Table 1. We notice from Table 1 that the correct answer rate of the Bayesian network, which is 61.44%, is the highest among them. Specially, it is almost 10%-higher than the others except for Case P2. Although the correct answer rate of the case P2 (psychological line with $N = 12$), which is 59.09%, is almost similar to the of the Bayesian network, the number of predictions in case P2 is much smaller than the other algorithms. As a result, total profit of the Bayesian network is much greater than the others.

6. Conclusions

The application of Bayesian network to the up/down rate analysis of the stock index was presented in this study. The network was determined according to the K2 algorithm with K2 metric as the network score from the up/down rates of the stock indexes; FTSE100, DOW30 and Nikkei225. The network was applied for predicting the improvement in the FTSE100 in 2007. The present algorithm showed almost 60% correct answer rate, which is higher than the results by the traditional algorithms such as psychological line and the trend estimation. Although the correct answer rate of the psychological line with $N = 12$ showed the similar accuracy, the number of investments is much smaller than that of the present algorithm. Therefore, the vertical investment revealed that total profit of the Bayesian network was much greater than the others.

References

- Ben-Gal, I. (2007). Bayesian networks. *Encyclopedia of Statistics in Quality and Reliability*, 1. John Wiley and Sons. <http://dx.doi.org/10.1002/9780470061572.eqr089>
- Bianchi, M., Boyle, M., & Hollingsworth, D. (1999). A comparison of methods for trend estimation. *Applied Economics Letters*, 6(2), 103-109. <http://dx.doi.org/10.1080/135048599353726>
- Cooper, G. F., & Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9(4), 309-347. <http://dx.doi.org/10.1007/BF00994110>
- Heckerman, D., Geiger, D., & Chickering, D. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20(3), 197-243. <http://dx.doi.org/10.1007/BF00994016>
- Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. *Representation and Reasoning Series*. Morgan Kaufmann.
- Pearl, J., & Russell, S. (2002). Bayesian networks. *Handbook of Brain Theory and Neural Networks* (pp. 157-160). MIT Press.
- Zuo, Y., & Kita, E. (2012). Stock price forecast using Bayesian network. *Expert Systems with Applications*, 39(8), 6729-6737. <http://dx.doi.org/10.1016/j.eswa.2011.12.035>