2014 International Symposium on Nonlinear Theory and its Applications
NOLTA2014, Luzern, Switzerland, September 14-18, 2014

Combinatorial Optimization of Financial Technical Indicators Based on Bayesian Network

Haruaki Sakaki† and Tomoya Suzuki†

†Graduate School of Science and Engineering, Ibaraki University,
Nakanarusawa-cho 4-12-1, Hitachi-shi, Ibaraki, 316-8511, Japan.
Email: 14nm909n@vc.ibaraki.ac.jp, tsuzuki@mx.ibaraki.ac.jp

Abstract—In the field of technical analysis, some technical indicators have been proposed to decide investment decisions and their timings, and are often used together to improve the decision-making accuracy. However, there is no specific guideline to select and combine technical indicators, and this selection is one of combinatorial optimization problems. Therefore, we applied the Bayesian network to estimate the causality among technical indicators, and identified the best combination of them. To confirm the validity of our proposed method, we performed some investment simulations based on real stock data, and confirmed to be able to make investments more efficient and safer, simultaneously.

1. Introduction

It is considered that financial markets have three different regimes: bullish trend, bearish trend, and range mode. If a market is in bullish trend, then we buy a long position, and if a market is in bearish trend, then we sell a short position. If a market is in range mode, we wait for a new trend breaking out. However, it is quite difficult to decide which kind of regime the present market is and how the regime will change in the near future. For this reason, some technical analysis indictors have been proposed [1]. These indictors are classified into trend-following, trend-reversal, volatility, momentum, and volume indicators, etc. Because each indictor focuses on different aspects of a market, it is said that we need to use some indictors simultaneously to improve the detection accuracy of trading chances. However, there is no specific guideline to use them together, and therefore it is often decided by our subjective intuition or experience.

In this point, behavioral finance points out that human judgment, emotion, and behavior are biased by unconscious or uncontrollable mental processing. For this reason, we are better to make a trading decision only by the specific strategy based on objective historical experiences. One of the possible methods is to calculate the conditional probability of how the past historical prices changed. As the conditions to clarify this probability, we can apply technical indicators in this study. However, it is difficult to use many indicators because it causes overfitting to the historical data or makes the conditional probability unstable especially when the number of historical samples is small. To avoid these problems, we have to increase the number of historical samples or have to select only essential indicators. However, the market structure is often changed by external accidents such as the Lehman Brothers bankruptcy and even politician's speeches. Therefore, we cannot use long historical data, and have to identify the essential indicators through the small historical samples. For this reason, we apply the Bayesian network [2] to solve the above problems, and improve trading performance by this specific numerical strategy based on objective historical experiences.

2. Technical Indicators

In the present study, we use major 10 technical indicators together as follows: SMA(Simple moving average), Bollinger bands, RSI(Relative strength index), Difference from moving average, Psychological line, Volume moving average, Wako volume ratio, DMI(Directional movement index), and MFI(Money flow index).

As an example, we show how to use the SMA, which works as a filter to reduce stochastic noises and to extract a market trend hidden in noisy price movements. If we denote $s(t)$ as the stock price at the time of t, this indicator is defined by

$$m(t,n) = \frac{1}{n}\sum_{a=0}^{n-1} s(t-a). \tag{1}$$

Here, if we suppose $n_1 < n_2$, when the short-term SMA $m(t,n_1)$ becomes over the long-term SMA $m(t,n_2)$, it is called as the Golden cross, which means that the market changed to an up trend and is considered as a buy signal. On the other hand, when SMA $m(t,n_1)$ becomes under SMA $m(t,n_2)$, it is called as the Dead cross, which means that the market changed to a down trend and is considered as a sell signal. Therefore, this SMA works as a trend-follow indicator, and we set $n_1 = 5$ and $n_2 = 20$ as the most popular days. The other technical indicators we used are categorized as trend-reversal (mean reversion) indicator, volatility indicator, momentum indicator, and volume indicator. Then, we set the most popular parameters for each indicator.

3. Application of Bayesian Network

The Bayesian Network is the graphical probabilistic model, which is defined by stochastic variables, conditional dependences among these variables, and their conditional probabilities [2]. Stochastic variables are expressed by nodes, and conditional dependencies between these variables are expressed by links. Bayesian network is modeled as a non-cycling directed network, the tip of a node is connected to a child node, and the end of it is connected to a parent node. Here, the conditional probability means the possibility of a child node after getting outputs from its parent nodes.

In the present study, we deal with the alternative question: whether tomorrow's stock price will rise or fall. First, we convert stock prices $s(t)$ into return rates $r(t)$ by

$$r(t) = \frac{s(t) - s(t-1)}{s(t-1)}, \tag{2}$$

and express the rise and drop of stock prices by the following stochastic variable R:

$$R(t+1) = \begin{cases} 1 & \text{if } r(t+1) \geq 0, \\ -1 & \text{if } r(t+1) < 0. \end{cases} \tag{3}$$

On the other hand, technical indicators introduced in Sec.2 are expressed by $\{X_1, \ldots, X_i, \ldots, X_{10}\}$. These are also considered as stochastic variables. Here, $X_i = 1$ means that ith technical indicator shows a buy signal, $X_i = -1$ means that it shows a sell signal, and $X_i = 0$ means that it shows no signal.

To compose a Bayesian network, we have to optimize its network topology on the basis of historical learning data. For this reason, we use the AIC to evaluate the likelihood of each network topology, and maximize this criterion. However, it takes a huge calculation time according to the number of nodes, that is, technical indicators. To solve this combinational optimization problem, we use the greedy search algorithm [3] where we find out the most reasonable parents of each child node by a steepest descent method.

Then, we consider the way of a probabilistic inference of $P(R|X_1, \ldots, X_{10})$. If we do not apply the Bayesian network, we find out the same situations as $\{X_1(t), \ldots, X_{10}(t)\}$ from the historical data, and empirically calculate the conditional probability of $R(t+1)$ as follows:

$$\begin{aligned} &P(R(t+1)|X_1(t), \ldots, X_{10}(t)) \\ &\quad = \frac{N(R(t+1), X_1(t), \ldots, X_{10}(t))}{N(X_1(t), \ldots, X_{10}(t))}, \end{aligned} \tag{4}$$

where $N(\cdot)$ means the function to count the occurrence of its inputs. For example, when the 1st technical indicator shows a buy signal ($X_1(t) = 1$), and the 4th indicator shows a sell signal ($X_4(t) = -1$), and the others show no signal ($X_2(t) = 0$, $X_3(t) = 0$, $X_5(t) = 0, \ldots, X_{10}(t) = 0$). In this case, there are two methods we can consider to calculate Eq.(4). The first method supposes that all variables $\{X_i(t)\}$ are important to infer the target variable $R(t+1)$, and calculates Eq.(4) by $P(R(t+1)|X_1(t) = 1, X_4(t) = -1, X_2(t) = 0, \ldots, X_{10}(t) = 0)$, following the historical data. The second method supposes that the only variables showing signals are important, and therefore Eq.(4) can be calculated by $P(R(t+1)|X_1(t) = 1, X_4(t) = -1)$. These methods are simple and low technique, and so we consider them as previous methods. Especially, we name the fist method "previous method 1" and the second method "previous method 2."

In terms of the amount of information, previous method 1 has the advantage over previous method 2. However, if this method includes independent variables from the target variable, their information becomes noise that causes the overfitting problem and disturbs the accurate estimation of Eq.(4). Moreover, as the number of historical data

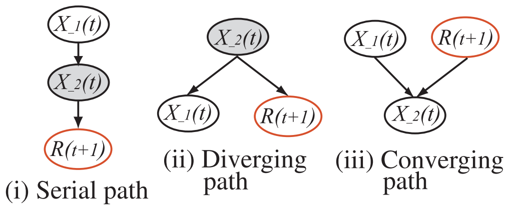

Figure 1: Connection patterns between nodes

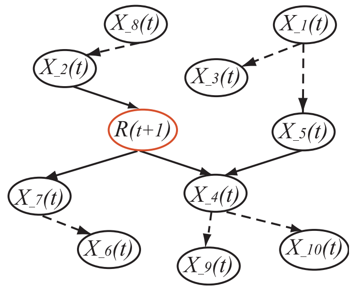

Figure 2: Example of an estimated network where solid arrows mean the dependance on node $R(t+1)$, and dashed arrows mean the independence from it by the d-separation.

is smaller or the number of technical indicators is larger, the number of the events that the function $N(\cdot)$ can count becomes so smaller. Therefore, the probability of Eq.(4) is often unstable. On the other hand, previous method 2 can make it more stable because technical indicators are limited to ones showing signals. However, the causality with the target variable is not considered, and therefore there is a danger that we ignore the essential variables for the probabilistic inference of Eq.(4). Besides, this method might include independent variables from the target variable as well as previous method 1.

Next, as the third method to evaluate Eq.(4), we apply the Bayesian network. At the beginning, we consider the d-separation, which is an important property of the Bayesian network for probabilistic inference. If $X_1(t)$ and $R(t+1)$ are d-separated by $X_2(t)$ on a Bayesian network composed of three nodes, it can be written by $(X_1(t) \perp\!\!\!\perp R(t+1)|X_2(t))$, which means that node $X_1(t)$ and node $R(t+1)$ are conditionally independent variables. As shown in Fig.1, there are three kinds of condition patterns. Here, $X_1(t)$ and $R(t+1)$ are d-separated by the following conditions:

$$\begin{cases} \cdot\ X_2(t) \text{ on a serial path shows a signal,} \\ \cdot\ X_2(t) \text{ on a diverging path shows a signal,} \\ \cdot\ X_2(t) \text{ on a converging path shows no signal.} \end{cases}$$

As an example, if we obtain a Bayesian network shown in Fig.2, the nodes affecting the target node $R(t+1)$ are only $X_2(t)$, $X_7(t)$, $X_4(t)$, and $X_5(t)$. It is because

$$\begin{cases} \cdot\ X_2(t) \text{ is the parent node of } R(t+1), \\ \cdot\ X_7(t) \text{ and } X_4(t) \text{ are the children nodes of } R(t+1), \\ \cdot\ X_5(t) \text{ is the parent node of } X_4(t). \end{cases}$$

Therefore, we do not need to use the other variables independent from $R(t+1)$, and can consider Eq.(4) as $P(R(t+

Table 1: Names of each method

Timing of checking $R(t+1)$'s probability	Everyday	Any indicators show signals	Any indicators affecting $R(t+1)$ show signals
Previous method 1	1-A	1-B	N/A
Previous method 2	2-A	2-B	N/A
Proposed method	3-A	3-B	3-C

$1)|X_2(t) = 0, X_7(t) = 0, X_4(t) = -1, X_5(t) = 0)$. We name this method "proposed method" in this study. Because our proposed method estimates the causality of the target variable $R(t+1)$, it can solve the overfitting problem that previous method 1 has, and does not ignore the essential nodes differently from previous method 2.

4. Investment Simulation

4.1. Outline

To confirm the validity of our proposed method, we perform some invest simulations with daily data (closing price) of 200 stocks listed in the first section of Tokyo Stock Exchange. Then, the learning data was set about four years from 1995 to 1999, and the test data for investment simulation was set about 4 years just after the learning data.

In the case of estimating a Bayesian network for our proposed method, 23 out of 200 stocks showed that the target node $R(t+1)$ is completely independent of the other nodes $\{X_i(t)\}$. In this case, we can consider that any technical indicators are meaningless, and namely $P(R(t+1)|X_1(t), \ldots, X_{10}(t)) = P(R(t+1))$. For this reason, we omitted these stocks when we applied the proposed method. On the other hand, because previous methods 1 and 2 do not estimate the causality of $R(t+1)$ at all, we applied all 200 stocks to these previous methods.

Next, we make an investment decision as follows: we take a long position if the probability of $R(t+1) = 1$ is more than θ[%], or we take a short position if the probability of $R(t+1) = -1$ is more than θ[%]. Here, θ[%] is a threshold to be set freely. On the next day, we close the position by the opposite trade.

Here, there are three different ideas about the timing of checking $R(t+1)$'s probability as shown in Table 1. The first idea is that we have to check it everyday because the entry condition mentioned above might be satisfied even when any technical indicators show no signals. These cases are also considered as good chances based on the historical data. On the other hand, it might be able to considered that these cases are not good chances because any technical indicators show no signals, and this means that the present market is quite stable, not bullish or bearish. Therefore, the second idea is that we check the $R(t+1)$'s probability only when any indicators show signals. Finally, if we can learn the causality properly by the Bayesian network, it might be the best to check it only when any indicators affecting $R(t+1)$ show signals because it is meaningless to focus on the nodes d-separated with $R(t+1)$. This is the third idea. The seven methods in total are named as shown in Table 1.

4.2. Measures to evaluate trading performance

We evaluate the investment performance by following measures: the number of executed trades T, the asset growth rate M, the maximum draw down rate R_d, the profit factor P_f, and the winning rate R_ω.

First, $T(t)$ means the total number of executed trades until the present time t. If each method can detect good trading chances, larger $T(t)$ is better. Otherwise, larger $T(t)$ is worse because it is risky. Then, the asset growth rate is calculated by

$$M(t) = \frac{A(t)}{A(1)}, \tag{5}$$

where $A(t)$ means the present total asset, and $A(1)$ means the initial asset. Of course, large $M(t)$ is better, and we can get profits if $M(t) > 1$. Next, a draw down rate[%] means

$$r_d(t) = \left[1 - \frac{A(t)}{\max\limits_{1\le t'\le t}\{A(t')\}}\right] \times 100, \tag{6}$$

and the maximum draw down rate[%] is calculated by

$$R_d(t) = \max_{1\le t'\le t}\{r_d(t')\}. \tag{7}$$

This shows how much we lost our asset so far, that is, the degree of possible danger of an investment strategy. Namely, smaller $R_d(t)$ is better and safer. Then, the profit factor is calculated by

$$P(t) = \frac{\sum\limits_{t'=1}^{t}\{\Delta A(t')|\Delta A(t') \ge 0\}}{\sum\limits_{t'=1}^{t}\{\Delta A(t')|\Delta A(t') < 0\}}, \tag{8}$$

where $\Delta A(t) = A(t) - A(t-1) = r(t)M(t-1)$. If $P_f(t) > 1$, the total profit is larger than the total loss, which means that this investment strategy works well. Finally, the winning rate[%] is calculated by

$$R_w(t) = \frac{\sum\limits_{t'=1}^{t} W(t')}{T(t)} \times 100, \tag{9}$$

where $W(t)$ is the function to distinguish whether the inference method of Eq.(4) can get the correct answer or not.

$$W(t) = \begin{cases} 1 & \text{if taking a buy position and } r(t+1) \ge 0, \\ 1 & \text{if taking a sell position and } r(t+1) < 0, \\ 0 & \text{otherwise.} \end{cases}$$

Because R_w corresponds to the inference accuracy, $R_w >$ 50[%] means that this inference works well.

4.3. Results

The results of investment simulations are shown in Figs.3 and 4. According to Fig.3, the number of executed trades T of the proposed method is smaller than that of the previous methods 1 and 2, which means that the proposed method takes positions more carefully. Then, it is natural that the number of invested stocks becomes smaller as θ is larger. However, previous methods 1 and 2 are not the case, which means that Eq.(4) was sometimes unstable in the counting process of $N(\cdot)$ and these outliers created in Eq.(4) were over even strict threshold θ. That is why most stocks are traded by previous methods 1 and 2. On the other hand, we can say that our proposed method can estimate Eq.(4) stably. That is why the number of invested stocks becomes smaller as θ is larger.

Next, as shown in Fig.3, in terms of the timing of checking $R(t+1)$'s probability, there is no difference between the first idea A (Everyday) and the second idea B (Any indicators show signals). However, only proposed method could improve trading performance by using the third idea C in 50[%] $\leq \theta \leq$ 60[%]. It might be because we focused on the indicators affecting $R(t+1)$ to omit noisy information of the other indicators.

Finally, as shown in Fig.4, the asset growth rate M of our proposed method was smaller than that of previous methods 1 and 2 because T of our proposed method is smaller compared with the previous methods. However, if we select a suitable θ, M can be larger than 1 by any methods. Then, in terms of the winning rate R_w, the profit factor P_f, and the draw down rate R_d, the propose method shows the best performance. Namely, we can say that our propose method can predict $R(t+1)$ more accurately, can improve the efficiency of investment, and can reduce the risk of investment.

5. Conclusion

We applied the Bayesian network to estimate the causality among a future return and technical indicators. This estimation is one of combinatorial optimization problems. Then, we inferred the probability of future returns with the estimated Bayesian network. Moreover, we discussed the suitable timing of this probabilistic inference. As a result, we could improve the trading performance by focussing on the technical indicators affecting the future return. Moreover, our propose method could enlarge the wining rate and the profit factor, and could reduce the draw down rate. Therefore, our propose method can not only improve the efficiency of investment but also reduce the risk of it. As a future work, we need to experiment with defferent markets and periods.

This was partially supported by a Grant-in-Aid for Scientific Research (C) (No.25330280) from JSPS.

References

[1] Robert W. Colby: "The Encyclopedia of Technical Market Indicators," McGraw-Hill, (2012).

[2] D.Heckerman: "A Tutorial on Learning With Bayesian Networks," Technical Report, no.MSR-TR-95-06, Microsoft Research, (1995).

[3] D.M.Chickering: "Optimal Structure Identification with Greedy Search," Journal of Machine Learning Research, vol.3, pp.507–554, (2002).

[4] D. Geiger, T.S. Verma and J. Pearl: "Identifying Independence in Bayesian Networks," Networks vol.20, pp.507–534, (1990).

[5] J. Pearl: "Probabilistic Reasoning in Intelligent Systems," Morgan Kaufmann Publishers Inc., (1988).

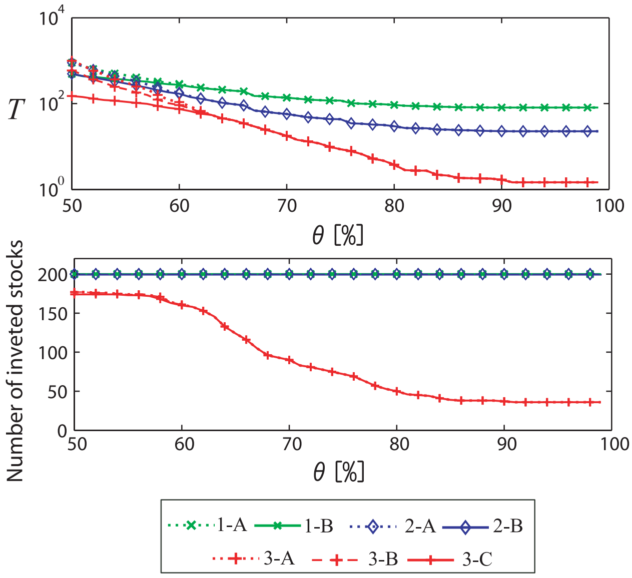

Figure 3: The number of executed trades T is shown as the average value of all stocks. Then, the number of the stocks traded more than once (i.e. $T > 0$) is shown in the bottom figure. Each legend is the same as Table 1.

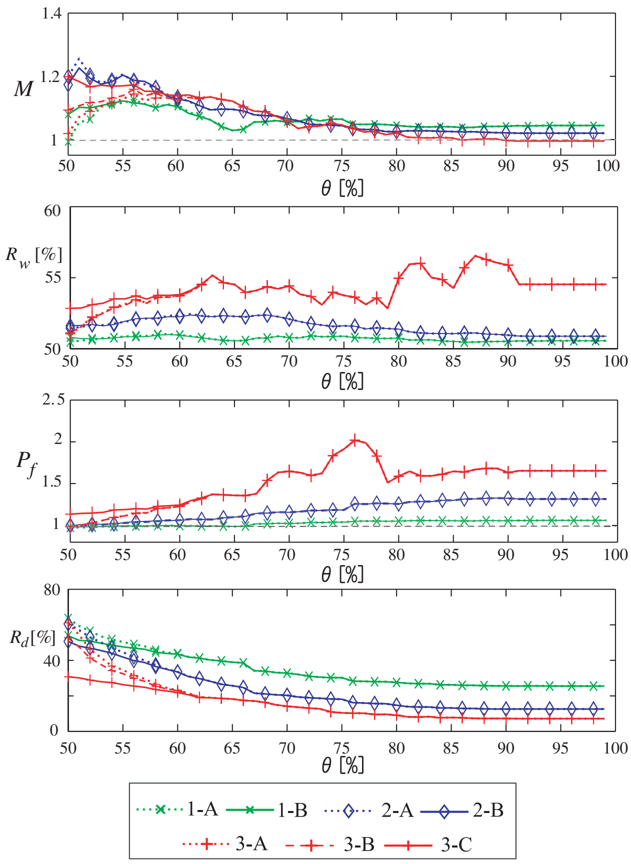

Figure 4: Each trading performance is shown as the average value of all stocks, but the stocks that had never been traded (i.e.T=0) were omitted. Each legend is the same as Table 1.