# Trend Predictions of Tick-wise Stock Prices by Means of Technical Indices Selected by Genetic Algorithm

Seiji Tokuoka and Mieko Tanaka-Yamawaki

Department of Information and Knowledge Engineering, Tottori University, Tottori 680-8550 Japan

**Abstract**

We propose a systematic method for predicting the trend of the price time series at several ticks ahead of the current price by means of genetic algorithm, used to optimize the combination of the frequently used technical indices such as various moving averages, deviation index from the moving averages, and so on. We show that the proposed method gives good predictions on the directions of motion with the rate as high as 80% for multiple stocks of NYSE selected from four different .business types. We also show that the performance improves if we combine two or three indices, compared to the case of using a single index. However, the performance seems going down as we increase the number of the indices from the optimum value.

## 1. Introduction

Forecasts of financial time series is a fascinating game to most of us. This game became more attractive than ever with the recent progress of Internet. Tick-wise price motions are by themselves thrilling news for day-traders. Many of them use the price charts as well as other information to help their decisions. Still no comprehensive enough way is known to compile those elements of knowledge into a systematic procedure. In this paper, we attempt to establish such a systematic procedure to make us win the game by using evolutional computations to identify a combination of useful technical indices applied on the real-world tick data. Since the tick-wise prices have strong trends and features [1], unlike the daily close values of stock prices known to be the random walk, we expect such predictions quite possible. Indeed we have observed a long-term stability of the tick-wise price fluctuations in terms of the direction of move of the next tick [2]. For example, the conditional probability of the up motion after the down motion is almost like a universal constant for several years [3].

$$P(up|down) = 1 - P(down|down) = 0.7 \qquad (1)$$

Guided by this fact, we have attempted to predict the immediate future price by means of evolutional computations [4].

However, it is hard to use the information of the next tick in a practical situation. It would be more useful to know the trend of the price time series a few minutes, or few hours ahead of us, to win the forecast game, for example. We therefore talk about predicting, say, the price level at 10-tick ahead by using several technical indices.

## 2. Tick data of stock prices

We have selected 8 stocks from NYSE, by choosing two from each of the four different groups of business, including retail stores, computer business, oil, automobile. We use bid-prices of those stocks for a year of 1993 (January 1 to December 31). We show the symbols, business types, total ticks in the year of 1993, and the average interval between ticks in Table 1.

Table 1 : Tick-wise stock prices in 1993

| symbol | business | ticks | interval (s/tick) |
|---|---|---|---|
| BBY | retail | 54821 | 109 |
| SMRT | retail | 12525 | 473 |
| APC | oil | 23685 | 253 |
| BP | oil | 73562 | 83 |
| CA | computer | 65051 | 92 |
| IBM | computer | 455233 | 14 |
| F | automobile | 194561 | 32 |
| GM | automobile | 277241 | 23 |

## 3. Elements of Technical Analysis

There are two kinds of approaches for financial forecasts: technical analyses and fundamental analyses. In the technical analyses, investors use the price motions such as chart patterns to predict the future price, while in the fundamental analyses they use global information on the company such as financial statements and health, its management and competitive advantages, and its competitors and markets comprehensively, to judge the essential value of the stock under consideration.

In this paper, we use the technical analyses to predict the intra-day price motions. Although there are numbers of indices found in the literatures, we select the ones that are supposed to be effective for the intra-day analysis. For example, candlesticks are not included in our analysis. We focus on the following three types of indices:

(1) Price trends (MA, EMA, MAD, SMA-LMA)
(2) Price oscillations (MACD, RCI, RSI, MO,PHL)
(3) Volume oriented indices

The investors use their favorite combinations of those indices. Our purpose is to study the way of choosing the most suitable combination of the indices for each stock and their conditions by employing the evolutional computation, and establish the automatic prediction generator on the trend of the price movement at several ticks ahead.

## 4. Our Method

### 4.1 Quantization of indices

Each index is quantized into a number of finite states. For example, the moving average (MA) can take two different states in comparison to the price, PRICE > MA or PRICE < MA.. On the other hand, the rate correlation index (RCI) ranges between -100 and 100, which are quantized into five states of interval 40, such as [-100, -60], [-60, -20], [-20, 20], [20, 60], and [60, 100]. Table 2 shows the number of states we have chosen for each index.

Table 2 : Number of state for each index

| Label | Index | # of States |
|---|---|---|
| 0 | MO1(Momentum) | 3 |
| 1 | MO2(Two-step-Momentum) | 9 |
| 2 | MA1(Moving Average) | 2 |
| 3 | MA2(LongMA+ShortMA) | 6 |
| 4 | RSI(Relative Strength Index) | 6 |
| 5 | MACD(MA Converg/Diverg) | 2 |
| 6 | MAD(MA Deviation rate) | 4 |
| 7 | RCI(Rank Correlation Index) | 5 |
| 8 | PHL(Psychological Line) | 6 |
| 9 | EMA(Exponential MA) | 6 |

### 4.2 Combination of indices determined by GA

The total number of the possible combinations of 10 indices is 1023. By using the genetic algorithm, we attempt to search for the best combination for each stock under various conditions.

#### 4.2.1 Gene representation

A chromosome is a set of indices determined by means of genetic algorithm. The best indices are supposed to be chosen after learning is completed.

Example1）

| | | | |
|---|---|---|---|
| 3 | 5 | 8 | 9 |

Example2）

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 5 | 6 | 8 | 9 |

Fig. 1: Two examples of gene coding

#### 4.2.2 Genetic evolution based on adaptability

We have used the following parameters in the process of learning in genetic algorithm: Population=100, Number of generations repeated= 500, Crossover rate= 0.9, Mutation rate= 0.01. We use the elite preservation selection as well as the roulette selection.

#### 4.2.3 Resulting combination of indices

The optimum combinations of indices obtained by applying the above method on the eight stocks are shown in Table 3.

Table 3: Best combination of indices for eight stocks

| Symbol | Combination of indices |
|---|---|
| APC | MO1,MA1,MACD,RCI |
| BBY | MO2,RSI |
| BP | MO2,MA2,MAD |
| CA | MO1,MA2 |
| F | MO2,MA1,MA2 |
| GM | MO2,MAD,PHL |
| IBM | MO2,RSI,MAD |
| SMRT | MO2,MA2,MACD |

### 4.3 Pattern classifications

Once the indices are selected, the prediction strategies are determined for all the possible states of the set of multiple indices by using data. For the choice of MO1, MA, and RCI, there exist 30(=3×2×5) combination of the states as shown in Fig.2.

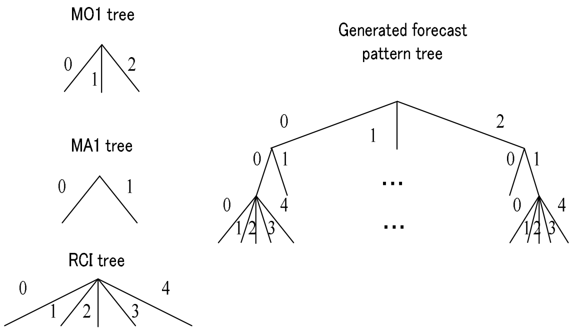


Fig. 2: A pattern represented in the tree structure

### 4.4 Method of Learning

The prediction strategy is identified as the majority of the direction of motion at X-ticks after the pattern of each combination of states appears. For example, the prediction strategy of X-ticks after the state (MO=0, MA=0, RCI=0) appears is learned to be “DOWN”, since there are more DOWN events than UP events after this pattern in the learning data as illustrated in Table 4.

Table 4: Strategy table

| (MO1,MA,RCI) | UP | DOWN | Strategy |
|---|---|---|---|
| 0,0,0 | 55 | 120 | DOWN |
| 0,0,1 | 23 | 6 | UP |
| 0,0,2 | 123 | 100 | UP |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 2,1,3 | 23 | 123 | DOWN |
| 2,1,4 | 8 | 4 | UP |

## 5. Testing new data

### 5.1.1 Setting

We test those prediction strategies on new data. The last half of the data set is served for this purpose. The prediction referred here is the UP/DOWN motion of the price at 10 ticks ahead compared to the current price.

### 5.1.2 Index used for the test

We examine the following three cases.

1. One for each index (total 9)
2. Combination of indices obtained in Section3.4
3. Combination of all the indices

Table 5: Parameter used for each index

| Index | | Period |
|---|---|---|
| MO1 | | 1 |
| MO2 | | 1 |
| MA1 | | 10 |
| MA2 | short | 5 |
| | long | 15 |
| RSI | | 10 |
| MACD | short | 5 |
| | long | 15 |
| | signal | 5 |
| MAD | | 10 |
| RCI | | 10 |
| PHL | | 25 |
| EMA | short | 5 |
| | long | 15 |

## 6. Result

Figure 3 shows the result of IBM and SMRT. The combination of indices selected by GA is the highest, though index 2, index 6, and index 9 are almost at the same level. This fact tells us that the combination of indices improves the result.

Indices related to moving averages, for example, MA, EMA, and MAD are proved effective for prediction. While MACD and RSI are not good alone, they improve the prediction rate if combined with other indices such as moving averages.

The combination of all the ten indices also performed well on the busy stocks of dense quot rates, although the result is poor for slow stocks like APC, BBY, and SMRT, which implies that the combination of all the indices requires heavy sized data for learning.

Figure 4 shows the prediction rate according to the number of indices. The best rate is shown for every size of the combined indices. For the case of APC, the combination of 3 indices performs best, and the rate decreases as the larger number of indices are combined. For GM, the rate of correct prediction increases as the number of indices increases up to 5, then starts decreasing after that.

The prediction rate improves when two or more indices are used. However, the combination of too many indices lowers the hitting rate.

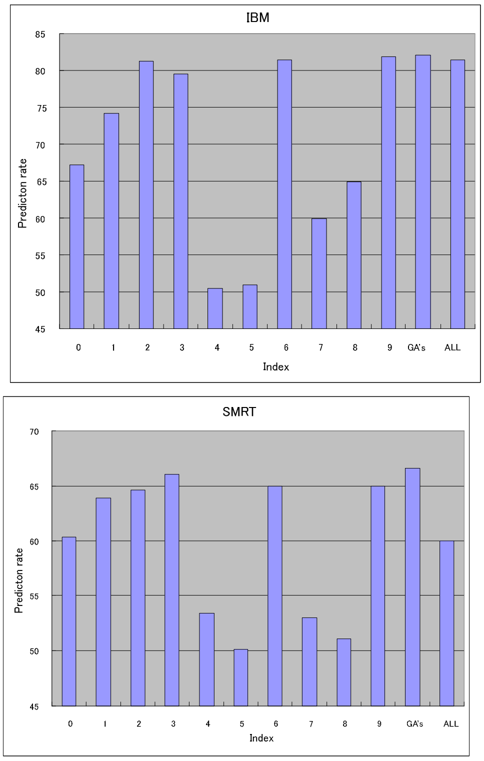


Fig.3: Prediction rate for IBM (above) and SMRT (below) according to each index

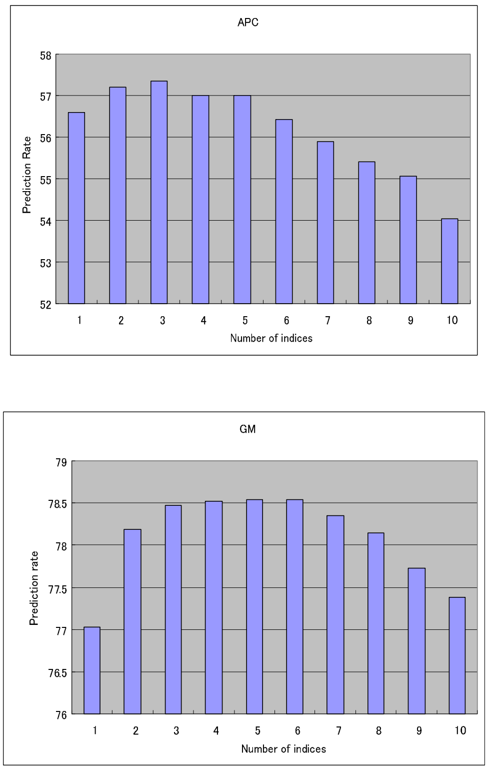


Figure 4:Comparison of prediction rate according to number of indices for APC(above) and GA (below).

Table 6 : Prediction rate according to stock symbols

| Brand | Coming true rate (%) | |
|---|---|---|
| | 1st | average |
| APC | 57.35 | 55.23 |
| BBY | 58.79 | 57.34 |
| BP | 65.36 | 63.44 |
| CA | 63.03 | 61.77 |
| F | 68.7 | 67.75 |
| GM | 78.54 | 77.37 |
| IBM | 82.25 | 81.38 |
| SMRT | 67.15 | 64.02 |

## 7. Conclusion

In this paper, we have extracted prediction strategies from the tick data of eight stocks of NYSE, by using the optimum combinations of technical indices chosen for each stock symbol based on genetic algorithm. As a result, we have obtained correct prediction rate as high as 81% for frequently exchanged stocks like IBM, or GM. The average prediction rate of eight stocks ended up to be about 65%.

We suppose that the reason of such a high rate of correct prediction lies in the characteristics of the tick-wise motion of stock prices. Especially, the strong correlation between the prices within the order of a few minutes (i.e., order of 10 ticks) resulted in the good performance of our method.

We have seen in this study that the moving averages are powerful tools for predicting the price range of the immediate future in the intraday time series of stock prices. We have also seen that the combination of multiple technical indices improves the prediction power compared to the case of using single index.

## References

[1] Takayasu H(Ed.) (2001), Empirical Science of Financial Fluctuantions, Springer

[2] Tanaka-Yamawaki M (2003), Stability of Markovian Structure Observed in High Frequency Foreign Exchange Data, Ann. Inst. Statist. Math, Vol. 55, pp.437-446

[3] Tanaka-Yamawaki M, Komaki S (2003), Characteristic Features of High Frequency Financial Time Series, Proc. AROB8th, pp.74-77

[4] Tanaka-Yamawaki M, Motoyama T (2004), Predicting the Tick-wise Price Fluctuations by Means of Evolutionary Computation, Proc. IEEE-CEC, pp.955-958