Society of Korea Industrial and Systems Engineering Vol. 36, No. 4 : 123-129, December 2013 ISSN : 2005-0461(print)
http://dx.doi.org/10.11627/jkise.2013.36.4.123 ISSN : 2287-7975(online)

Conceptual Framework for Pattern-Based Real-Time Trading System using Genetic Algorithm

Suk-Jun Lee · Suk-Jae Jeong†

Business School, Kwangwoon University

유전알고리즘 활용한 실시간 패턴 트레이딩 시스템 프레임워크

이석준 · 정석재†

광운대학교 경영학부

The aim of this study is to design an intelligent pattern-based real-time trading system (PRTS) using rough set analysis of technical indicators, dynamic time warping (DTW), and genetic algorithm in stock futures market. Rough set is well known as a data-mining tool for extracting trading rules from huge data sets such as real-time data sets, and a technical indicator is used for the construction of the data sets. To measure similarity of patterns, DTW is used over a given period. Through an empirical study, we identify the ideal performances that were profitable in various market conditions.

Keywords : Stock Futures Market, Pattern-Based, Rough Set Analysis, Dynamic Time Warping, Genetic Algorithm

1. Introduction

The development of investment strategies for stable and lucrative profits from the stock futures market has been the subject of interest for many investors and professional analysts. A stock future is a financial derivative that is decided by the volatility in the underlying asset value and is known as one of the most successful financial innovations of the 1980s. Recent decades have experienced the development of pattern-based trading based on template matching techniques and pattern recognition in the stock futures market [1-4]. Many stock and futures traders mostly rely on various types of expert system including statistical and artificial intelligence methods for pattern trading [5-7]. A number of studies have shown that pattern-based trading using technical indicators can be successful and profitable. For that, researchers found various technical indicators to be particularly useful for identifying appropriate patterns [8-10].

Real-time trading system based on algorithmic or automated trading has become popular due to diverse sets of real-time data becoming available throughout a market [11]. The real-time data implies that the information is delivered immediately after collection, and is provided without any delay. In contrast, delayed or historical data are delivered after some time has elapsed (e.g., usually 10 to 30 minutes) due to the transaction. Therefore, the delayed or historical data may not be suitable for real-time trading because these data are usually adjusted after being combined with real-time data [12-14]. Developing pattern trading system with real-time data is not easy task because the data provides a large amount of information. Until now, most studies of futures markets have proposed trading rules on the basis of simple technical analysis without considering real-time data [15].

This study proposes a procedure for constructing a pattern-based real-time trading system (PRTS) that is based on rough

Received 29 August 2013; Finally Revised 10 December 2013; Accepted 18 December 2013
† Corresponding Author : sjjeong@kw.ac.kr

set analysis [16] of technical indicators and dynamic time warping (DTW) algorithm, originally introduced by Bellman [17]. The DTW algorithm is one of many techniques that can be used to measure similarity between two time series (or two patterns) particularly when the two are not aligned properly on time axis [17]. In fact, PRTS modifies the process of the RRTS (real-time rule-based trading system) developed by Lee et al. [18].

Notice that RRTS used the technical indicators consisting of trend following indicators and oscillators for generating the trading rules. Also, it employed a manual method for reduction and applied the Euclidian distance method that is static method when it finds a reference pattern close to the current movement. However, PRTS uses the oscillator that provides the signal (i.e. buy or sell) by recognizing reversal trend, and it uses a genetic algorithm (GA) method for the reduction and DTW for the recognition of dynamic stock pattern. The DTW is used as a core recognizer to identify similar patterns in the dynamic stock futures market. The algorithm is one of many pattern recognition techniques that can be used to measure the similarity between two time series (or two patterns) particularly when the two are not aligned properly on time axis. In empirical studies, the PRTS yielded profitable and stable earnings from the market.

2. Material and Methods

2.1 Data Sets and Initial Conditions

Traders require more powerful support in their investment decisions because their capability for analyzing enormous real-time data is limited. For a PRTS construction, we used the Korea Stock Price Index 200 (KOSPI 200) 30-minute interval as a dataset and considered the period from Jul. 1996 to Dec. 2006. The period is divided it into a pattern base-constructed period (Jul. 1996 to Dec. 2004) and a real-time trading period (Jan. 2005 to Dec. 2006). Then, the pattern base-constructed period was divided into a training period and a testing period. A system trading was applied to the real-time data from the pattern base-constructed period. As a default condition for the system trading, the initial capital was set to 1,000,000 won (equal to 1,000 dollar), open market interest rates were set to 5.00%, transaction cost was set to 10,000 won, and slippage was set to 25,000 won. To evaluate the trading system, the return rates were calculated for the underlying asset [18]. The return rates represent the yearly profit rates that are calculated from the ratio of the current capital value to the initial capital value after one year of trading. The yearly profit is defined as the yearly gross profit minus transaction costs and slippages in which the yearly gross profit is the yearly short position minus the yearly long position. The slippage cost is an additional amount set aside to prevent missed trading. These conditions were applied in this empirical study.

2.2 Application of Rough Set and DTW

This section provides a detailed description of the construction procedure for a PRTS. Real-time data is the stock futures market data which consist of the open, high, low, close prices, and trading volume. This study used nine oscillators that were created from the data. The definitions and formulas of the oscillators are listed in <Table 1>.

PRTS consists of three phases (see <Figure 1>). The first phase is pattern-wise rules generation using rough set analysis. This phase involves three steps, where trading rules are generated for each of the patterns using the rough set modeling process (quoted from Lee et al. [18]) with real-time data. Here, we used six patterns defined from Lee et al. [18]. Each of the patterns is characterized by a distinctive shape as follows : short-term ascending pattern (SAP), short-term descending pattern (SDP), long-term ascending pattern (LAP), long-term descending pattern (LDP), flat top pattern (FTP), and flat bottom pattern (FBP).

In the first step, data cleaning through the exploration is conducted. This includes the removal of obvious outliers and data completion. To improve the overall quality of the discovered information, further data transformation is typically conducted by using discretization, which makes the attributes smaller. For discretization, the equal frequency binning method is used for data transformation.

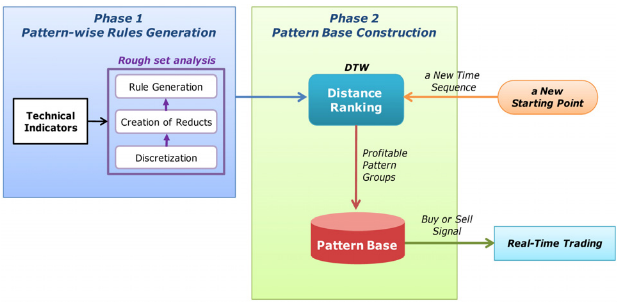
The second step is the creation of reducts, through the computation of the reducts. It is an important process because core information is extracted in the concrete rule from the transformed data. Each reduct that corresponds to six patterns is selected by the GA method in this study. Notice that RRTS used manual reducer that is trial and error. The generated reducts can be filtered according to criteria such as coverage and accuracy, attribute cost, advanced quality measures, or classificatory performance on external holdout data sets [19].

The final step is rule generation for a given pattern. Using

<Table 1> Nine Oscillators Used in This Study (adapted from Lee et al. [18])

Oscillators	Formulas
SMI (Stochastic Momentum Index)	$SMI_t(n) = 100 \times \left(\frac{2C}{(H_{t,\max}(n) - L_{t,\max}(n))} - 1 \right)$
William's %R	$William's\ \%R(n) = \frac{H_{t,\max}(n) - C_t}{H_{t,\max}(n) - L_{t,\max}(n)} \times 100$
NCO (Net Change Oscillator)	$NCO_t(n) = C_t - C_{t-n}$
ROC (Rate of Change)	$ROC_t(n) = \left(\frac{C_t}{C_{t-n}} - 1 \right) \times 100$
CCI (Commodity Channel Index)	$CCI_t(n) = \frac{M_t - \overline{M}_t(n)}{d_t(n) \times 0.015}$ where $M_t = (H_t + L_t + C_t)/3$ and $d_t(n) = \frac{1}{n}\sum_{i=1}^{n} \lvert M_{t-n+i} - \overline{M}_t(n) \rvert$.
PO (Price Oscillator)	$PO_t(m,n) = \frac{SMA_t(m) - SMA_t(n)}{SMA_t(m)}$
TRIX	$TRIX_t(n) = \frac{EMA_t^3(C, n) - EMA_{t-1}^3(C, n)}{EMA_{t-1}^3(C, n)}$
RSI (Relative Strength Index)	$RSI_t(n) = 100 - \frac{100}{1 + RS_t(n)}$ and $RS_t(n) = \sum_{i=1}^{n} U_{t-n+i} / \sum_{i=1}^{n} D_{t-n+i}$ where $U_t = \begin{cases} C_t - C_{t-1}, & C_t \geq C_{t-1}, \\ 0 & otherwise \end{cases}$ and $U_t = \begin{cases} D_{t-1} - D_t, & D_t \leq D_{t-1}, \\ 0 & otherwise \end{cases}$
MACD (Moving Average Convergence-Divergence)	$MACD_t(m, n) = EMA_t(C, m) - EMA_t(C, n)$

Note : In the above, the subscript t indicates that the related quantities are defined on the time interval $(t-1, t]$, unless stated otherwise. For example, C_t, H_t, L_t, and V_t denote the close price, the high price, the low price, and the trading volume recorded on $(t-1, t]$, respectively. Furthermore, we define $y_{t,\max}(n) = \max(y_t, \cdots, y_{t-n+1})$, $y_{t,\min}(n) = \min(y_t, \cdots, y_{t-n+1})$ and $\overline{y_t}(n) = \sum_{i=1}^{n} y_{t-n+1}/n$. Additionally, $m \leq n$ is always assumed.



<Figure 1> Architecture of RRTS

each reduct generated in the second step, rules can be written in the form of 'IF-THEN' statements to combine condition attributes (i.e., input variable) with decision attributes (i.e., output variable). The condition attributes are the values of nine oscillators, whereas a decision attribute is a BUY or SELL signal transformed from UP and DOWN of stock futures price. In general, the combination of condition attributes and decision attributes is based on conjunction. An example of the form of the generated decision rule is as follows :

IF value of 1st oscillator belongs to the range of the corresponding oscillator AND value of 2nd oscillator belongs to the range of the corresponding oscillator AND value of nth oscillator belongs to the range of the corresponding oscillator THEN BUY (or SELL)

In practice, one may apply rules sequentially on the number of positions to hold. For example, one may employ the following implementation (or trading) rule, to limit the number of positions.

IF signal at time t is BUY And IF signal at time $t-1$ is BUY, THEN HOLD ELSE SELL,
IF signal at time t is SELL And IF signal at time $t-1$ is SELL THEN HOLD ELSE BUY.

The second phase is a constructing pattern base using DTW. At this phase, the pattern base for real-time trading is developed, which consists of six pattern groups. Specifically, the pattern group can be defined as a set of profitable trading rules selected in SAP, SDP, LAP, LDP, FTP, and FBP. Here, each pattern group consists of trading rules selected in terms of return rates.

For pattern recognition in the pattern base, the best similar time sequence is identified for a given time sequence. Namely, the cumulative distance between a new time sequence and each reference time sequence corresponding to each group is calculated using DTW algorithm. The pattern groups are ranked by minimizing the cumulative distance for a new time sequence. The trading rules of pattern groups that correspond to the time sequence with the smallest distance will initially be used for a new time sequence in the testing period. In this study, all of the holdout trading intervals (i.e., length of time sequence) for 10 starting points and 6 patterns are fixed at 20 days, which represents the monthly effect in the futures market.

3. Experiments and Results

3.1 Profitable Rules Selection

For the construction of the PRTS, the training and the testing periods for each of the six patterns were used (see <Table 2>). Each period of patterns is different according to short –term pattern, long-term pattern, and flat pattern.

<Table 2> Training and Testing Periods Obtained from Jul. 1996 to Dec. 2004 for Each of the Six Distinctive Patterns [18]

Patterns	Training periods		Testing periods	
	Starting date	Ending date	Starting date	Ending date
SAP	Dec. 26, 1997	Mar. 05, 1998	Dec. 26, 2000	Jan. 22, 2001
SDP	Mar. 06, 1998	Jun. 15, 1998	Jul. 14, 2000	Sep. 22, 2000
LAP	Oct. 01, 1998	Jul. 09, 1999	Apr. 02, 2003	Apr. 23, 2004
LDP	Sep. 09, 1996	Dec. 24, 1997	Apr. 24, 2002	Apr. 01, 2003
FTP	Jul. 12, 1999	Feb. 03, 2000	Apr. 28, 2004	Dec. 30, 2004
FBP	Jun. 16, 1998	Sep. 30, 1998	Sep. 25, 2000	Dec. 22, 2000

At first phase, reducts were generated, selected by the genetic algorithm (GA) programming from each pattern by the rough set modeling process in the training periods.

As discussed above, we use a GA to select appropriate reducts by each pattern with the follow form.

{Rule}{condition}{connector}

{Rule}{condition}{connector}⋯{Action}

Where the individual rules are based on technical indicators, conditions are TRUE or FALSE, the connectors are Boolean operators (in this case AND, OR, and XOR) and the action is one of BUY, SELL, and HOLD.

At each trial of GA performs the following stages in each iteration until the algorithm converges : construction of population and evaluation of fitness.

We initialize the population using ancillary uniform random integers which are translated to the strings that constitute the initial population. We also allow strings to be of various lengths but, for case of representation, achieve this by generating strings of a fixed length and ignoring parts of the generated string. This can be thought of as generating a string containing all available indicators and their associated units and then 'switching some on.'

At each trail over a user-defined period, each of reducts is tested by simulating their trading performance over historical data. The fitness evaluation of every string is calculated at each iteration using criteria : the return rates discussed in section 2.1. It is calculated by essentially excess return minus slippage cost, occurring due to missed trading.

The generated reducts have decision rules that are able to convert trading rules. The return rate of the reducts was produced by a combination of these trading rules. <Table 3> lists reducts with profitable trading rules for each of the

<Table 3> Reducts with Trading Rules of the Six Pattern Groups and Their Return Rates (%) in the Training and Testing Periods.

Pattern group	Reducts	Return rates(%)[a]	
		Training period	Testing period
SAP	{SMI, RSI, CCI, PO}	262.05	107.63
	{SMI, ROC, PO}	134.17	34.64
	{RSI, William's %R, PO}	312.37	194.23
SDP	{RSI, William's %R, MACD, TRIX}	152.10	101.62
	{RSI, NCO, MACD, TRIX}	167.56	121.69
	{SMI, RSI, TRIX}	117.00	24.59
	{RSI, MACD, CCI, TRIX}	118.67	42.15
	{SMI, RSI, William's %R, MACD, PO}	113.66	23.33
	{RSI, ROC, MACD, TRIX}	141.65	34.63
	{RSI, MACD, TRIX, PO}	213.11	109.65
	{RSI, NCO, William's %R, MACD, CCI, PO}	124.52	34.88
LAP	{ROC, NCO, William's %R, MACD, CCI, TRIX, PO}	77.07	14.28
LDP	{RSI, NCO, MACD, TRIX}	39.36	12.75
	{RSI, MACD, CCI, TRIX}	10.28	12.29
	{RSI, William's %R, MACD, TRIX}	13.95	16.16
FTP	{William's %R, MACD, CCI, TRIX, PO}	21.59	43.38
FBP	{RSI, ROC, PO}	38.18	55.62

[a] Return rates are calculated for the training and the testing period of each of the six pattern groups.

six patterns. The reducts in <Table 3> are twice as good as the open market interest rate in the training and testing periods. The reducts constitute the pattern groups. The number of the pattern group for each of the SAP, SDP, LAP, LDT, FTP, and FBP is 3, 8, 1, 3, 1 and 1, respectively, for the trading rules that consist the pattern base.

3.2 Patterns Recognition

For the evaluation of PRTS, 10 starting points were randomly selected from the real-time trading period. All of the trading intervals were fixed at 20 days. If a starting point is determined, then the pattern base was activated. Similarity is calculated between the historical time sequences corresponding to each of the six patterns at the pattern base and a new time sequence by the DTW. It notes that time sequences are oscillators, and historical oscillators of new time sequence have to be generated and calculated corresponding to oscillators of each patterns. Average of distance value of each oscillator is similarity. <Table 4> shows the distance between the 6 pattern groups and 10 new time sequences and the distance of pattern in increasing order from left to right. It means that the shortest distance implies the most similar pattern.

3.3 Trading Simulations

<Table 5> shows that as the number of combination of patterns increases, the "absolute" value of the average return rates tend to decrease. For instance, the return rate means average return rate of LAP, LDP, and FTP when number of combinations of pattern group is 3 at first starting date (Mar. 03, 2005). Theoretically, this is due to the small value of patterns increasing the dependence of the PRTS on the particular pattern(s) among the patterns, whereas a large value of patterns yields a robust PRTS (in the sense that it performs reasonably in various situations).

For the practical selection of patterns, the estimated Sharpe ratio was used to evaluate the performance of the pattern base as a rule portfolio. The Sharpe ratio is defined as a ratio of the expected difference between the return rates of a given portfolio and the risk-free asset over the standard deviation of the difference [20]. In <Table 5>, the average return rates of the pattern base were compared for 10 new time sequences with the Sharpe ratios for each pattern. Here, the return rates of the risk-free asset used for the Sharpe ratio calculation are those of Treasury bills with a 3-year maturity (average return rate from 2005 to 2006 was 4.55%). The Sharpe ratio increases to 0.28, 0.33, and 0.36

<Table 4> Results of Calculations of Distance between Reference Time Sequences (i.e., Pattern Groups) and 10 New Time Sequences on the Basis of Each Starting Point

Starting points	Distance of pattern groups in increasing order from left to right					
Mar. 03, 2005	LAP group (189.06)	LDP group (200.17)	FTP group (272.71)	SAP group (283.03)	FBP group (285.08)	SDP group (305.27)
May 31, 2005	FTP group (235.56)	LDP group (246.59)	LAP group (248.40)	FBP group (297.64)	SDP group (319.63)	SAP group (328.13)
Jul. 04, 2005	LDP group (223.06)	LAP group (227.04)	FTP group (260.46)	FBP group (317.42)	SDP group (333.56)	SAP group (349.06)
Aug. 01, 2005	LAP group (241.64)	LDP group (241.81)	FTP group (268.61)	FBP group (301.15)	SAP group (331.61)	SDP group (341.87)
Sep. 01, 2005	LDP group (220.92)	LAP group (246.96)	SAP group (267.57)	FTP group (274.64)	FBP group (298.60)	SDP group (323.06)
Dec. 16, 2005	LDP group (244.36)	LAP group (246.53)	FTP group (273.73)	FBP group (273.73)	SAP group (313.67)	SDP group (326.83)
May 04, 2006	LAP group (24468)	LDP group (245.77)	FTP group (248.71)	SAP group (307.36)	FBP group (311.28)	SDP group (313.21)
Jun. 21, 2006	LDP group (218.96)	FBP group (236.76)	SAP group (238.86)	LAP group (258.09)	SDP group (273.49)	FTP group (292.61)
Sep. 06, 2006	FBP group (204.41)	LAP group (215.14)	LDP group (234.33)	FTP group (245.38)	SAP group (317.20)	SDP group (325.71)
Nov. 23, 2006	LDP group (243.83)	LAP group (252.28)	FTP group (275.86)	FBP group (297.38)	SAP group (340.45)	SDP group (356.82)

<Table 5> Average Return Rates (%) of the Specifically Determined Combinations of Pattern Groups, in Order of Shortest Distance

Starting date	Number of combinations of pattern groups					
	1	2	3	4	5	6
Mar. 03, 2005	21.21	32.52	33.48	-3.69	-19.96	2.63
May 31, 2005	20.01	43.69	39.52	-4.15	-4.02	-11.24
Jul. 04, 2005	-1.84	-5.93	-6.11	-32.89	-68.53	-25.74
Aug. 01, 2005	4.41	37.28	16.07	-6.76	9.77	-56.29
Sep. 01, 2005	8.44	-6.92	-1.34	-16.35	-19.01	-38.73
Dec. 16, 2005	3.71	-12.41	17.59	-43.10	-27.13	-43.34
May 04, 2006	8.65	2.86	12.84	-8.29	-77.73	14.21
Jun. 21, 2006	-2.26	-6.54	-8.45	-42.10	-47.83	11.66
Sep. 06, 2006	40.8	3.28	23.6	13.06	-22.95	-15.10
Nov. 23, 2006	17.52	49.52	33.15	-1.69	2.26	-12.53
Average return rate	12.06	13.73	16.03	2.58	-3.44	-5.78
Sharpe ratio	0.28	0.33	0.36	-0.25	-1.44	-1.90

by the third pattern group, and then plunges dramatically from the fourth pattern group. Namely, if the number of pattern groups is three (i.e. LAP, LDP, and FTP at first starting point), then the rule portfolio exhibits the best performance. Despite an average return rate of 16.03% in the third combination of pattern groups, the investment using rough set analysis can be profitable compared with the average 5% of the open market interest rate and the return overcomes the performance of RRTS.

4. Conclusions

This study demonstrates the usefulness of the rough set and DTW for constructing PRTS in the Korean stock futures market. In particular, the PRTS using the rough set and the DTW was useful in various market situations. Because real-time data or trading causes the market to be more volatile and unpredictable than before, the PRTS provides a desirable, robust solution compared with various situations. In the meantime, we envision that this study will encourage further studies of the PRTS because it incorporates basic tools within the rough set and DTW. A more elaborate procedure for the PRTS might be developed with the help of other information reduction techniques such as manual reducer, dynamic reducts, and approximate hitting set approaches.

References

[1] Lo, A.W., Mamaysky, H., and Wang, J., Foundation of technical analysis : computations, algorithms, statistical inference, and empirical implementation. *Journal of Finance*, 2000, Vol. 55, p 1705-1765.

[2] Leigh, W., Purvis, R., and Ragusa, J.M., Forecasting the NYSE Composite Index with technical analysis, pattern recognizer, neural network, and genetic algorithm : a case study in romantic decision support. *Decision Support Systems*, 2002, Vol. 32, p 161-174.

[3] Chen, G., Wu, X., and Zhu, X., Sequential pattern mining in multiple streams, in Proceedings of the Fifth International Conference on Data Mining. *IEEE Computer Society*, 2005, p 585-588.

[4] Dorr, D.H. and Denton, A.M., Establishing relationships among patterns in stock market data. *Data and Knowledge Engineering*, 2009, Vol. 68, p 318-337.

[5] Armano, G., Marchesi, M., and Murru, A., A hybrid genetic-neural architecture for stock indexes forecasting. *Information Science*, 2005, Vol. 170, p 3-33.

[6] Chang, P.C., Wang, Y.W., and Yang, W.N., An investigation of the hybrid forecasting models for stock price variation in Taiwan. *Journal of the Chinese Institute of Industrial Engineering*, 2004, Vol. 21, p 358-368.

[7] David, E. and Suraphan, T., The use of data mining and neural networks for forecasting stock market returns. *Expert System Application*, 2005, Vol. 29, p 927-940.

[8] Chavarnakul, T. and Enke, D., Intelligent technical analysis based equivolume charting for stock trading using neural networks. *Expert Systems with Applications*, 2008, Vol. 34, p 1004-1017.

[9] Sevastianov, P. and Dymova, L., Synthesis of fuzzy logic and Dempster-Shafer Theory for the simulation of the decision-making process in stock trading systems. *Mathematics and Computers in Simulation*, 2009, Vol. 80, p 506-521.

[10] Yudong, Z. and Lenan, W., Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert Systems with Applications*, 2009, Vol. 36, p 8849-8854.

[11] Kim, K., Electronic and algorithmic trading technology : The complete guide(Complete technology guides for financial services), Burlington : Academic Press, 2007.

[12] Christoffersen, P., Ghysels, E., and Swanson, N.R., Let's get "real" about using economic data. *Journal of Empi-*

rical Finance, 2009, Vol. 9, p 343-360.

[13] Anderson, T.G., Bollerslev, T. Diebold, F.X., and Vega, C., Micro effects of macro announcements : Real-time price discovery in foreign exchange. *American Economic Review*, 2003, Vol. 93, p 38-62.

[14] Clark, T.E. and Kozicki, S., Estimating equilibrium real interest rates in real-time. *North American Journal of Economics and Finance*, 2005, Vol. 16, p 395-413.

[15] Ben, R.M. and Rochester, H.C., Is technical analysis profitable on a stock market which has characteristics that suggest it may be inefficient?. *Research in International Business and Finance*, 2005, Vol. 19, p 384-398.

[16] Pawlak, Z., Rough set approach to knowledge-based decision support. *European Journal of Operational Research*, 1997, Vol. 99, p 48-57.

[17] Bellman, R., Dynamic Programming, NJ : Princeton University Press, 1957.

[18] Lee, S.J., Ahn, J.J., Oh, K.J., and Kim, T.Y., Using rough set to support investment strategies of real-time trading in futures market. *Applied Intelligence*, 2010, Vol. 32, p 364-377.

[19] Øhrn, A., Discernibility and rough sets in medicine : Tools and applications, Ph.D. Thesis, Trondheim : Norwegian University of Science and Technology, 1999.

[20] Sharpe, W.F., The Sharpe ratio. *Journal of Portfolio Management*, 1994, Vol. 21, p 49-58.