INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
www.ijrcar.com
Vol.2 Issue.11, Pg.: 59-69
November 2014

INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

AN AUTOMATED FRAMEWORK FOR INCORPORATING NEWS INTO STOCK TRADING STRATEGIES

J.Nithiya Devi [1], G.Vijayabharathi[2]

[1]*M.Phil Research Scholar, PG & Research Dept. of Computer Science*
Kaamadhenu Arts & Science College, Sathayamangalam - 638 503
Jennifercs2011@gmail.com
[2]*Associate Professor of Comp. Science, PG & Research Dept. of Computer Science*
Kaamadhenu Arts & Science College, Sathayamangalam - 638 503
Viji_vijee@yahoo.com

Abstract:

A stock market or equity market is the aggregation of buyers and sellers of stocks (shares); these are securities listed on a stock exchange as well as those only traded privately. With the increasing number of information sources, resulting in high volumes of news, manual processing of the knowledge being conveyed becomes a highly difficult task. So, in the existing work a framework is presented for automatic exploitation of news in stock trading strategies.

1. Introduction

Data mining is the practice of automatically searching large stores of data to discover patterns and trends that go beyond simple analysis. Data mining uses sophisticated mathematical algorithms to segment the data and evaluate the probability of future events. Data mining is also known as Knowledge Discovery in Data (KDD).

The key properties of data mining are:

- Automatic discovery of patterns
- Prediction of likely outcomes
- Creation of actionable information

- Focus on large data sets and databases

Data mining is the process of discovering actionable information from large sets of data. Data mining uses mathematical analysis to derive patterns and trends that exist in data. Data mining is one of the most important research fields that are due to the expansion of both computer hardware and software technologies, which has imposed organizations to depend heavily on these technologies. Data is considered as the number one asset of any organization, it is obvious that this asset should be used to predict future decisions sequent, and since organizations are continuously growing, their relative databases will grow as well; as a result their current data mining techniques will fail to cope up with large databases which are dynamic by nature.

Data mining is the way to help organization make full use of the data stored in their databases and when it comes to decision making, this is true in all fields, and is also true in all different types of organizations .Data mining is the task of discovering interesting and hidden patterns from large amounts of data where the data can be stored in databases, data warehouses, OLAP (online analytical process) or other repository information. It is also defined as knowledge discovery in databases (KDD).

Data mining involves integration of techniques from multiple disciplines such as database technology, statistics, machine learning, neural networks, information retrieval, etc. "Data mining is the process of discovering meaningful patterns and relationships that lie hidden within very large databases. Data mining is a part of a process called KDD-knowledge discovery in databases. This process consists basically of steps that are performed before carrying out data mining, such as data selection, data cleaning, pre-processing, and data transformation. Typically, these patterns cannot be discovered by traditional data exploration because the relationships are too complex or because there is too much data.

Knowledge discovery in databases process or KDD is relatively young and interdisciplinary field of computer science is the process of discovering new patterns from large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems. The goal of data mining is to extract knowledge from a data set in a human-understandable structure. Data mining is the entire process of applying computer-based methodology, including new techniques for knowledge discovery, from data. Databases, Text Documents, Computer Simulations, and Social Networks are the Sources of Data for Mining. Knowledge extraction, the process of finding interesting, interpreted, useful and novel data from a large set of data is known as Knowledge Discovery in Databases (KDD). The steps involved in mining the data are as follows: Pre-processing, mine the data and interpret the results.

2. Literature Review

2.1 Semi-Automatic Financial Events Discovery Based on Lexico-Semantic Patterns-Jethro Borsje, Frederik Hogenboom, Flavius Frasincar, 2010.

The Semantic Web provides the right technologies to classify the information in news items and make it available for both human and machine consumption. Being able to identify financial events from news items would help the trader to make a decision whether to react on the financial market.

In this work, investigate how a user acting as a trader can identify the financial events of interest in titles extracted from RSS news feeds. The only requirement that have for the user is that he should be familiar with the financial domain as captured in ontology. Due to his interest in buying and selling stocks of certain companies assume that the user has a minimum knowledge of the financial markets. The user does not have to be familiar with Semantic Web technologies, the domain ontology will be presented in a graphical manner as a tree of concepts and concept relationships. Such an approach should allow the user to describe the events of interest, extract the event instances from news items, and update the domain ontology based on the effects of the discovered event instances. During the design of this approach special attention is given to the user interface that should allow a simple interaction between the user and the system. Such an interface should enable a simple specification of the events of

Vol.2 Issue.11, Pg.: 59-69
November 2014

interest and event triggered-updates for the domain ontology. For this purpose exploit the triple paradigm due to its intuitiveness and simplicity. Triples are used for defining lexico-semantic information extraction patterns that resemble simple sentences in natural language. In addition, triples are also used to express the event-triggered ontology updates.

2.2 Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support William Leigh a,*, Russell Purvis b, James M. Ragusa, 2002.

This work exemplifies the potential that lies in the novel application and combination of methods, in this case to evaluating stock market purchasing opportunities using the ‘‘technical analysis’’ school of stock market prediction. Members of the technical analysis school predict market prices and movements based on the dynamics of market price and volume, rather than on economic fundamentals such as earnings and market share. The results of this work support the effectiveness of the technical analysis approach through use of the ‘‘bull flag’’ price and volume pattern heuristic. The romantic approach to decision support exemplified in this work is made possible by the recent development of: (1) high-performance desktop computing, (2) the methods and techniques of machine learning and soft computing, including neural networks and genetic algorithms, and (3) approaches recently developed that combine diverse classification and forecasting systems.

3. MODULE DESCRIPTION

List of modules

- Extraction of Event Information
- Pre-processing
- Relationship between News and Share Prices
- Fine-grained analysis
- Technical trading indicators
- Genetic algorithm
- Performance Evaluation

Extraction of Event Information

The event information extraction from the news messages is based on recognizing a predefined set of events as well as the affiliated entities. Viewer Pro is an application created by SemLab that enables the identification of events in news messages. These events can be used to determine the impact of a news item on equity. Viewer Pro turns enormous amounts of unstructured news into structured trading information. Once the unstructured news information is fed in the Viewer Pro system, it undergoes several (proprietary) processing steps in order to filter out unwanted information and select solely that which is relevant. Large amounts of news messages are filtered for equity specific news and the semantic analysis system of Viewer- Pro interprets the impact of every individual news message.

Pre-processing

In this module, does not include articles issued on days when the stock exchange is closed as the events contained in these messages will not have an immediately quantifiable impact on the stock price. Since several events can occur during the period when the stock exchange is closed, associating these events with changes in price over this period will introduce additional variance with regard to which event precisely influences the change in price. At times, news messages may be repeated to provide updates on an event described in a previous message. This result in events that are on the same day, concern the same company, and are identical to another event previously extracted on the same day. Since these news messages describe the same events, it suffices to only consider them once and thus incorporate the associated impact for the event in the stock price projection only once.

Relationship between News and Share Prices

The impact of news on stock prices is assessed using relative returns, based on end-of-day data, i.e., closing prices P. For a single asset, a return is computed as:

$$r_i = \frac{P_{i+1} - P_i}{P_i} \times 100,$$

Where i represents the day before the event and n represents the number of days over which the return is calculated, with $n > 0.$ In case multiple events of the same type appear in different days, regarding the same asset, the return is averaged for the number of days, as follows:

$$R_i = \frac{\sum_{j=1}^{N} r_j}{N},$$

Where N is the number of days where events of this type occurred. To correct the returns for the general market sentiment, focus on excess returns. The excess return is calculated as the individual return of an asset that is achieved in excess of the market return, i.e., the return of the main index in which the asset is included:

$$a_i = r_i - r_i^I,$$

Where r_i^I denotes the return of the index employed as benchmark. When dealing with multiple events of a certain type appearing in different days and with excess returns, correct these returns for the number of days:

$$A_i = \frac{\sum_{j=1}^{N} a_j}{N},$$

Where N is the number of days where events of this type occurred.

Fine-grained analysis

In this section, the fine-grained analysis is considered in order to provide better trading strategies. In this method, temporal patterns are considered in which the interaction between events are considered within the same day or within the fine-grained time intervals. So, it should provide more information for generating trading strategies. This will provide the deeper understanding of the way that news impact stock prices and may lead to more profitable trading strategies.

Technical trading indicators

This section focuses on the technical trading indicators used in trading strategies generated through genetic programming. The indicators included in the study are: the simple moving average (SMA), the Bollinger band (BB), the exponential moving average (EMA), the rate of change (RoC), momentum (MOM), and moving average convergence divergence (MACD). The choice for these indicators is based on their widespread use in technical trading.

Simple Moving Average

The SMA averages the last 20 days of the price of a stock, and is computed as:

$$M_i = \frac{\sum_{i=1}^{N} P_i}{N}$$

Where P_i represents the price on day i. The average is calculated over a fixed period of 20 days prior to the day for which the average is calculated, i.e., N = 20, which is standard for this indicator. A buy signal is generated when the price crosses the moving average in an upward movement, while a sell signal is generated when the price crosses the moving average in a downward movement.

Bollinger Bands

The Bollinger band is a technical indicator which creates two ‘bands’ around a moving average. These bands are based on the standard deviation of the price. It is assumed that the price will move within these bands, around the moving average. If the volatility is high, the bands are wide and when there is little volatility the bands are narrow. The lower and upper Bollinger bands can be calculated as:

$$L = M - 2 \times \sigma_M,$$

$$U = M + 2 \times \sigma_M,$$

Where σ_M stands for the volatility of moving average M. A buy signal is generated when the price is below the lower band, which is regarded as an oversold situation. A sell signal is generated at an overbought situation, when the price is above the upper band.

Exponential Moving Average

The exponential moving average aims to identify trends by using a short and a long term average. When the averages cross each other, it is the start of a new trend. The short term average is set at five days and the long term average at 20 days:

$$E_i = \frac{2}{N+1} \times (P_i - E_{i-1}) + E_{i-1}$$

Where P_i represents the price on day i, and N is the number of days. The initial EMA is calculated using the SMA, in this case for five and 20 days respectively starting from the first observation, as previously described. When the short term average crosses the long term average upwards, a buy signal is generated. A sell signal is generated when the short term average crosses the long term average downwards.

Rate of Change

The rate of change is an indicator that calculates the difference between the closing price P_i of the current day i and the closing price P_{i-10} of 10 days earlier, according to the following equation:

$$C_i = \frac{P_i - P_{i-10}}{P_{i-10}}$$

If the RoC starts decreasing above 0 (a peak was reached), a sell signal is generated. If it starts increasing below 0, a buy signal is generated.

Momentum

The momentum indicator uses exactly the same formula as the RoC. Instead of creating a buy signal after a peak, it creates a buy signal when the momentum crosses the zero level upwards. A sell signal is generated when the RoC crosses the zero level downwards.

Moving Average Convergence Divergence

The moving average convergence divergence is a technical indicator that subtracts two exponential averages from each other, namely the 12 and the 26 day exponential average. The mathematical formula for the MACD is:

$$D_i = E[12]_i - E[26]_i$$

A buy signal is generated when the MACD reaches the zero level in an upward motion. A sell signal is generated when the MACD breaks through the zero level in a downward motion.

Genetic algorithm

Next, a framework is introduced for incorporating news in stock trading strategies. The framework assumes that events have been extracted from news messages and are available together with the date on which the events took place. Additionally, a predefined impact should be assigned to each event, allowing the news variable to be included in the trading strategies. The trading strategies determine take the form of trees that, when evaluated, return a Boolean value: true, when a trading signal is generated, or false, when no signal is generated and thus no action has to be taken. The trading strategies include at least one technical indicator or a news variable. Most often, the trading strategies include multiple variables, that may be either technical indicators or the news variable, connected by the logical operators ‘and’ and ‘or’. This rule generates a trading signal when the simple moving average generates a trading signal simultaneously with at least one of the exponential moving average and rate of change indicators. The fitness of a trading strategy is computed based on the return that it generates on the data set that used.

The employed genetic programming algorithm for determining the optimal trading strategies is presented. Start from a random initial population of trees, and generate new populations of trading strategies by applying crossover and mutation on the population from the previous iteration. Crossover consists of selecting two trading strategies, and determining two random crossover points, i.e., one for each tree. Next, the sub trees generated under the crossover point are exchanged between the two trading strategies, thus resulting in two new rules that are added to the new population. Mutation only relates to the technical indicators included in a trading strategy, and consists of a slight change in the parameters of the randomly selected technical indicator, e.g., changing the number of days used by the simple moving average from 5 to 7. The stopping condition for the algorithm relates to the improvement in the best solution found, i.e., when the optimal solution cannot be improved in a number of generations, the algorithm stops.

4. Experimental Setup

Require: $\alpha \geq 0$: minimum improvement

$\beta > 0$: Maximum times of no improvement

$\gamma > 0$: Population size

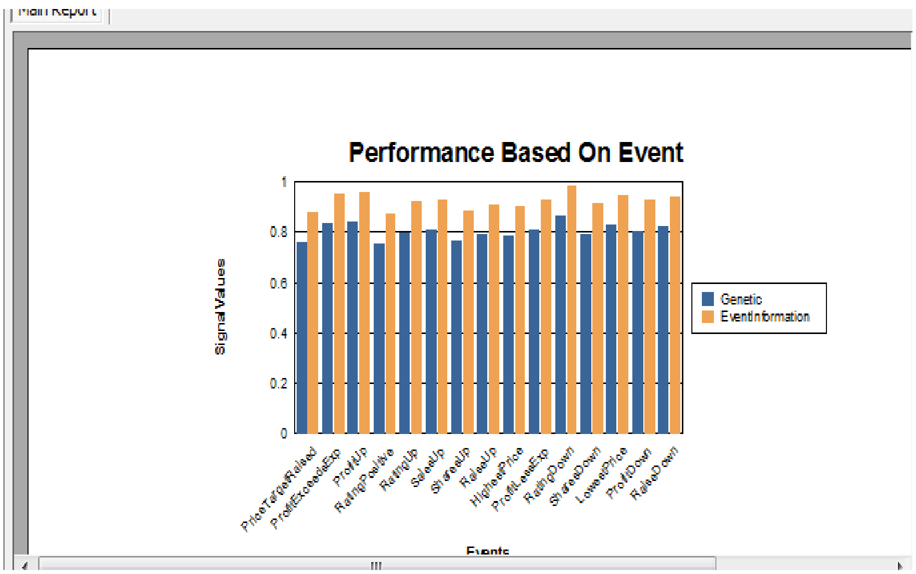
$0 < \rho \leq \gamma$: Number of parents

Vol.2 Issue.11, Pg.: 59-69
November 2014

$0 \le \mu \le 1$: Mutation of probability

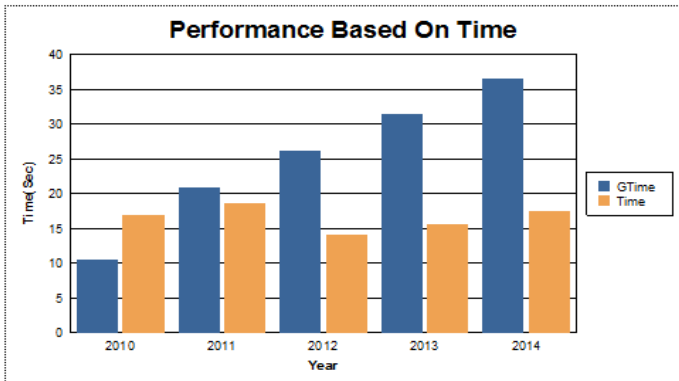
1. $\pi = generaterandomPopulation(\gamma)$
2. $\sigma_{old} = -\infty, \sigma_{new} = calcFitness(\pi), b = 0$
3. While $b < \beta$ do
4. $addindividual\left(\pi', getbest(\pi, \sigma_{new})\right)$
5. While $|\pi'| < |\pi|$
6. $\theta = SelectRandomParents(\pi, \sigma_{new}, \rho)$
7. $\vartheta = crossover(\theta)$
8. $\vartheta' = mutate(\vartheta, \mu)$
9. $addindividual(\pi', \vartheta')$
10. $end\ while$
11. $\pi = \pi', \sigma_{old} = \sigma_{new}, \sigma_{new} = calcFitness(\pi)$
12. $if \sigma_{new} - \sigma_{old} \le \alpha$ then
13. $b = b + 1$
14. Else if $b > 0$
15. B=0
16. End if
17. End while
18. Return π

PERFORMANCE BASED ON EVENT



The performance based on the event. In the existing system, a framework is presented for automatic exploitation of news in stock trading strategies. In the proposed system, fine-grained analysis of for automatic exploitation of news in stock trading strategies is used. In the X-axis events are taken. In the Y-axis the signal values are taken. When compared to the existing system there is high performance in the proposed system.

PERFORMANCE BASED ON TIME



The performance based on the time. In the existing system, a framework is presented for automatic exploitation of news in stock trading strategies. In the proposed system, fine-grained analysis of for automatic exploitation of news in stock trading strategies is used. In the X-axis year is taken. In the Y-axis the time in seconds are taken. When compared to the existing system there is less time taken for finding the optimal strategies in the proposed system.

5. EXISTING SYSTEM

In the existing system, a framework is presented for automatic exploitation of news in stock trading strategies. A three step approach is presented consisting of: (i) extracting the relevant events, as well as the involved entities, from the text of the news messages, (ii) associating an impact with each of the extracted events, and (iii) making use of the impact of news events in trading strategies. Upon extracting the events and associating these with a predefined impact, trading rules based on news can be derived. The technical trading indicators are only considered as part of these trading rules, but the approach can be easily extended to incorporate other indicators, e.g., those initiating from fundamental analysis. Hypothesize that, if the proposed framework is valid, news will be included in the trading strategies generated through genetic programming. Additionally, the trading strategies that derive in this way should generate positive returns. The first hypothesis comes from the idea that, when providing a genetic program with a pool of variables without the restriction that all these variables should be included in a trading strategy, only the variables that are maximizing the returns will be selected. Trading strategies including a news variable will thus indicate that the content of the news messages has been quantified in a way that enables generation of profit beyond the ability of trading rules based solely on technical analysis. The second hypothesis states that, next to generating trading strategies based on news, the resulting rules should also be able to obtain a positive return.

PROPOSED SYSTEM

In the proposed research, a fine-grained analysis is used of the news messages, e.g., identification of event-related information such as the involved actors, should provide more information for generating trading strategies. Last, considering the interaction between events occurring within the same day, or within finer-grained time intervals, will provide a deeper understanding of the way that news impact stock prices and may lead to more profitable trading strategies.

Last, a genetic program is used to discover complex trading rules based on technical indicators and news-based signals. The employed genetic programming algorithm for determining the optimal trading strategies is presented. Start from a random initial population of trees, and generate new populations of trading strategies by applying crossover and mutation on the population from the previous iteration. Crossover consists of selecting two trading strategies, and determining two random crossover points, i.e., one for each tree. Next, the sub trees generated under the crossover point are exchanged between the two trading strategies, thus resulting in two new rules that are added to the new population. Mutation only relates to the technical indicators included in a trading strategy, and consists of a slight change in the parameters of the randomly selected technical indicator, e.g., changing the number of days used by the simple moving average from 5 to 7. The stopping condition for the algorithm relates to the improvement in the best solution found, i.e., when the optimal solution cannot be improved in a number of generations, the algorithm stops.

6. Conclusion

A framework is presented for incorporating news into stock trading strategies. The trading strategies that considers may include (in addition to the news variable) any number of technical trading indicators. The news variable is quantified based on the events extracted from the text of news messages and the assignment of an expert-defined impact to each of these events.

The selected technical indicators are also tested, and the individual performance of each indicator is reported. Additionally, combinations of individual technical indicators and the news variable are investigated. Last, a genetic program is used to discover complex trading rules based on technical indicators and news-based signals. But only achieve less profitable trading strategies in the existing system. So, in the proposed work, the interaction between events is considered occurring within the same day, or within finer-grained time interval for better trading strategies.

7. Acknowledgement

For future work, a method is suggested to investigate accounting for the type of news events. Additionally accounting the general stock market events instead of only just the company specific news.

REFERENCES

[1] J. Borsje, F. Hogenboom, and F. Frasincar, “Semi-Automatic Financial Events Discovery Based on Lexico-Semantic Patterns,” Int’l J. Web Eng. and Technology, vol. 6, no. 2, pp. 115-140, 2010.

[2] W. IJntema, J. Sangers, F. Hogenboom, and F. Frasincar, “A Lexico-Semantic Pattern Language for Learning Ontology Instances From Text,” J. Web Semantics: Science, Services and Agents on the World Wide Web, vol. 15, no. 1, pp. 37-50, 2012.

[3] [3] S.R. Das and M.Y. Chen, “Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web,” Management Science, vol. 53, no. 9, pp. 1375-1388, 2007.

[4] M. Cecchini, H. Aytug, G.J. Koehler, and P. Pathak, “Making Words Work: Using Financial Text as a Predictor of Financial Events,” Decision Support Systems, vol. 50, no. 1, pp. 64-175, 2010.

[5] P.C. Tetlock, “Giving Content to Investor Sentiment: The Role of Media in the Stock Market,” J. Finance, vol. 62, no. 3, pp. 1139- 1168, 2007.

[6] P.C. Tetlock, “More than Words: Quantifying Language to Measure Firms’ Fundamentals,” J. Finance, vol. 63, no. 3, pp. 1437- 1467, 2008.

[7] W. Leigh, R. Purvis, and J.M. Ragusa, “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, vol. 32, no. 4, pp. 361-377, 2002.

[8] K. Mehta and S. Bhattacharyya, “Adequacy of Training Data for Evolutionary Mining of Trading Rules,” Decision Support Systems, vol. 37, no. 4, pp. 461-474, 2004.

[9] M.-A. Mittermayer and G.F. Knolmayer, “Text Mining Systems for Market Response to News: A Survey,” technical report, Institute of Information Systems University of Bern, http:// www.ie.iwi.unibe.ch/publikationen/berichte/resource/WP- 184, 2006.

[10] M.L. Mitchell and J.H. Mulherin, “The Impact of Public Information on the Stock Market,” J. Finance, vol. 49, no. 3, pp. 923-950, 1994.

[11] W.S. Chan, “Stock Price Reaction to News and No-News: Drift and Reversal After Headlines,” J. Financial Economics, vol. 70, no. 2, pp. 223-260, 2003.

[12] S.T. Kim, J.C. Lin, and M.B. Slovin, “Market Structure, Informed Trading, and Analyst’ Recommendations,” J. Financial and Quantitative Analysis, vol. 32, no. 4, pp. 507-524, 1997.

[13] S.G. Ewalds, M.B.J. Schauten, and O.W. Steenbeek, “De Informatiewaarde Van Kwartaalcijfers,” Maandblad voor Accountancy en Bedrijfseconomie, no. 7/8, pp. 333-341, 2000.

[14] J.B. Warner, R.L. Watts, and K.H. Wruck, “Stock Prices and Top Management Changes,” J. Financial Economics, vol. 20, no. 1, pp. 461-492, 1988.

[15] K.A. Bonnier and R.F. Bruner, “An Analysis of Stock Price Reaction to Management Change in Distressed Firms,” J. Accounting and Economics, vol. 11, no. 1, pp. 95-106, 1989.

[16] A.J. Keown and J.M. Pinkerton, “Merger Announcements and Insider Trading Activity: An Empirical Investigation,” J. Finance, vol. 36, no. 4, pp. 855-869, 1981.

[17] R.J. Rosen, “Merger Momentum and Investor Sentiment: The Stock Market Reaction to Merger Announcements,” J. Business, vol. 79, no. 2, pp. 987-1017, 2006.

[18] D.L. Ikenberry and S. Ramnath, “Under reaction to Self-Selected News Events: The Case of Stock Splits,” Rev. Financial Studies, vol. 15, no. 2, pp. 489-526, 2002.

[19] R. Michaely, R.H. Thaler, and K.L. Womack, “Price Reactions to Dividend Initiations and Omissions: Overreaction or Drift,” J. Finance, vol. 50, no. 2, pp. 573-608, 1995.

[20] X.F. Zhang, “Information Uncertainty and Stock Returns,” J. Finance, vol. 61, no. 1, pp. 105-137, 2006.

[21] J. van Bommel, “Rumors,” J. Finance, vol. 58, no. 4, pp. 1499-1520, 2003.

[22] D.L. McGuinness and F. van Harmelen, “OWL Web Ontology Language: W3C Recommendation 10 February 2004,” http:// www.w3.org/TR/owl-features/, 2004.

[23] V. Milea, F. Frasincar, and U. Kaymak, “Knowledge Engineering in a Temporal Semantic Web Context,” Proc. Eighth Int’l Conf. Web Eng. (ICWE ’08), pp. 65-74, 2008.

[24] F. Frasincar, V. Milea, and U. Kaymak, “tOWL: Integrating Time in OWL,” Semantic Web Information Management: A Model-Based Perspective, R. De Virgilio, F. Giunchiglia, and L. Tanca, eds., pp. 225-246, Springer, 2010.

Vol.2 Issue.11, Pg.: 59-69
November 2014

[25] V. Milea, F. Frasincar, U. Kaymak, and G. Houben, "Temporal Optimizations and Temporal Cardinality in the Towl Language," Int'l J. Web Eng. and Technology, vol. 7, no. 1, pp. 45-64, 2012.

[26] [V. Milea, F. Frasincar, and U. Kaymak, "tOWL: A Temporal Web Ontology Language," IEEE Trans. Systems, Man and Cybernetics, Part B, Cybernetics, vol. 42, no. 1, pp. 268-281, Feb. 2012.

[27] F. Allen and R. Karjalainen, "Using Genetic Algorithms to Find Technical Trading Rules," J. Economics, vol. 51, no. 2, pp. 245-271, 1999.

[28] F. Hogenboom, M. de Winter, M. Jansen, A. Hogenboom, F. Frasincar, and U. Kaymak, "Event-Based Historical Value-at- Risk," Proc. IEEE Conf. Computational Intelligence for Financial Eng. Economics (CIFEr '12), 2012.

[29] S.B. Achelis, Technical Analysis from A to Z. McGraw-Hill, 2000. [30] J. Koza, Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press, 1992.