© 2015 IJSRSET | Volume 1 | Issue 2 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

A Genetic Programming Approach for Optimal Trading Strategies

R. Divya, M. Divya, K. Surya
Dhanalakshmi College of Engineering, Chennai, Tamilnadu, India

ABSTRACT

This paper is about an outline for computerized development of information in stock trade strategy. Actions are extracted from information post offered in open content with no explanation. We study the introduced plan by deriving trade strategy based on scientific indicator and impact of the extract actions. The strategy take the structure of policy that merge scientific trade indicator with a consecutively adaptable, and are exposed throughout the utilize of genetic programming. We discovery that the information changeable is frequently incorporated in the best possible trading policy, representing the further charge of information for projecting purpose and validate our future structure for consequentially incorporate reports in stock trade strategy.

Keywords: Computer applications, evolutionary computing and genetic algorithms, learning, natural language processing, web text analysis.

I. INTRODUCTION

Financial markets are driven by information. An important source of information is news communicated by different media agencies through a variety of channels. With the growing number of information sources, resulting in high volumes of news, manual processing of the knowledge being conveyed becomes a highly hard task. Moreover, given that this information is time-sensitive, especially in the context of financial markets, selecting and processing all the relevant information in a decision-making process, such as the decision whether to buy, hold, or sell an asset is an especially challenging task. This environment motivates a need for automation in the dispensation of information, to the extent that investment decisions where the news factor plays an important role can be based on an automatically generated recommendation that takes into account all news messages relevant to a certain financial asset. In previous work we have devised lexico-semantic patterns for information mining from news that extend the well-known lexico-syntactic patterns with semantic aspects [1], [2]. Using information extracted from text in a financial context recently enjoys increasing attention. In [3] the authors extract investor sentiment from stock message boards. The prediction of bankruptcy of firms, as well as fraud, based on textual data from the management discussion and analysis sections (MD&A) of 10-K reports is investigated in [4].

A popular Wall Street Journal column is used for investigating asset prices and trading volumes in [5]. Financial news stories are used for the prediction of stock returns and firms' future cash flows in [6]. Thus, the qualitative data may appear from different sources, and can be used for the prediction of different financial aspects of firms' performance. We focus on information presented in textual format, i.e., financial news messages with a particular focus on Company listed under the FTSE350 stock index. The Research question addressed is how the information communicated through textual news messages can be automatically incorporated into trading strategies. We use a three step approach consisting of: (i) extracting the related 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

II. METHODS AND MATERIAL

A. Related Work

Regarding the relationship between news and the stock market, we consider three key aspects: (i) there is evidence that a relationship exists between new announcements and financial markets, (ii) the contact of events on financial markets can be quantified, and a list of relevant events can be recognized, and (iii) the relationship between information in the form of news and financial markets is not a trivial one. One aspect left aside relates to mining news messages for assessing market retort. For a survey of the different methods employed for this purpose, we refer the reader to [7].

- The information effect (negative): the management performance is worse than expected by the market.
- The real effect (positive): the change is in shareholders' interest. If a company performs very badly, a management change could mean
- A new vision, strategy, etc., so the expectations about the companies' future results could be revised. The news is received positively.

B. A Preliminary Analysis of the Relationship between News and The Stock Market

Our analysis of the relationship between news and the stock market, as apparent from the collected data set is inattentive on discovering the manipulate that news have on the share price of the concerned companies, as well as on whether this influence can be captured through the extraction of events from news messages and employing a predefined impact for determining the direction of this influence on prices.

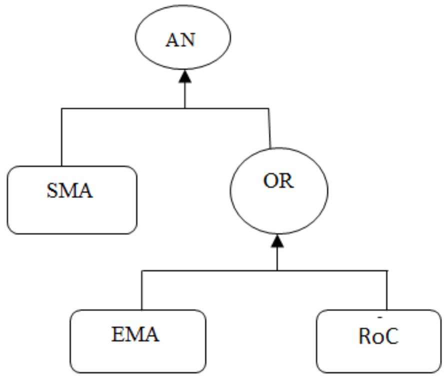
C. Technical Trading

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 [8].

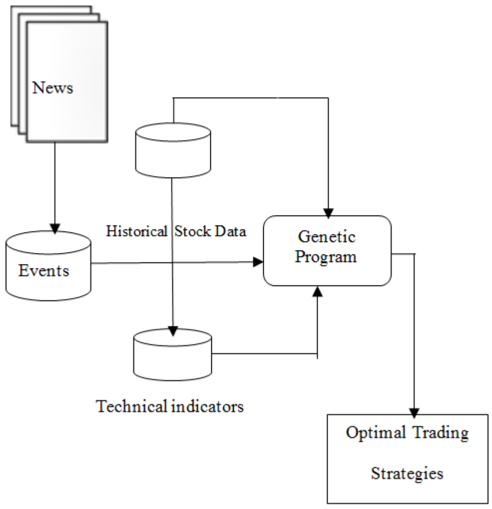
D. A News-Based Trading Framework

Genetic programming [9] is a technique where the potential solutions are represented as computer programs rather than numerical values encoded in some manner. Starting from a (usually randomly generated) initial population, genetic programs attempt to improve the fitness of individuals over successive generations through a process inspired by natural evolution.

During this process, individuals are altered, usually based happening their fitness values, by combining them with other individuals (crossover), or by slightly modifying some parts of the individual with a predefined probability (mutation). In this paper, genetic programming is used for finding optimal trading strategies based on technical indicators and news. Genetic programming has previously been used in the design of decision support systems, e.g., in [10], [11].



An example of a trading strategy that may be generated is given above.



We summarize the proposed framework in above diagram. As illustrated in the figure, the events are extracted from the news messages represented in free text format, and constitute the input to the algorithm. The historical price data constitutes an individual input to the search algorithm, used for computing the performance of the trading strategies, but is simultaneously used to derive the values for technical trading indicators, another input to the algorithm. Finally, the optimal trading strategies are determined through genetic programming.

E. Analysis Methodology

News citations per month have weak positive correlation coefficients of 0.01 with returns. There is large dispersion over time in the cross sectional correlation, although the relation is still statistically significant. The occurrence of headlines is more strongly related to turnover. The correlations range from -0.1 to 0.6, and average about 0.15. I conclude that headlines do not seem to favour good news (denoted by high returns). Also, depending on one's interpretation of turnover, one could infer that highly liquid and/or more controversial stocks attract more media attention [12]. The news includes many corporate actions such as mergers and tender offers, as well as earnings announcements. These are and often accompanied by other news. However, my news set misses some earnings announcements.

Furthermore, many news "events" are not company actions or pre-scheduled earnings releases. These Include analyst ratings changes, capital spending announcements, block holder sales and purchases, and new contracts. Second, there are some months when a reading of the headline does not reveal if the news was good or bad. For example, acquisitions and ratings changes are accompanied by both positive and negative returns. This suggests that it may be wise to rely on the market reaction to filter "good" and "bad" news. Third, the "winner" and "loser" categories are broad because I use thirds to divide firms by returns. For instance, some stocks display zero or slightly positive returns in some months, but may be classified as "losers", based on relative performance. Finally, news does not appear auto correlated, since a single stock can switch from being a news winner to a news loser several times in a year.

F. Financial News Article Sources

In real-world trading applications, the amount of textual data available to stock market traders is staggering. This data can come in the form of required shareholder reports, government-mandated forms, or news articles concerning a company's outlook. Reports of an unexpected nature can lead to wildly significant changes in the price of a security. Textual data itself can arise from two sources; company generated and independently generated sources. Company generated sources such as quarterly and annual reports can provide a rich linguistic structure that if properly read can indicate how the company will perform in the future (Kloptchenko, Eklund et al. 2004). This textual wealth of information may not be explicitly shown in the financial ratios but encapsulated in forward-looking statements or other textual locations. Independent sources such as analyst recommendations, news outlets, and wire services can provide a more balanced look at the company and have a lesser potential to bias news reports [13].

G. Post-Merger Performance

Here, a brief and selective review of prior research on long-run post merger underperformance is presented. A comprehensive review has been carried out by Agrawal and Jaffe (2000). Langetieg (1978) reported significant cumulative abnormal returns (CARs) between -2.23% and -2.62% over a six-year period after a merger. Asquith (1983) found that acquiring firms' CAR decreases by 7.2% in one year following the completion of mergers. Malatesta (1983) found a statistically significant CAR of -7.6% one-year after the merger announcement [14]. Management changes occur for different reasons, and differ in their a priori relation through prior share performance. Some could follow good performance, and others could have no relation to prior performance. In either case, inclusion of such changes biases tests against finding an inverse relation between share perfol-mance and top management changes. To address this potential problem, several types of management changes are studied using information on the details of observed management changes [15].

H. Temporal Coalescing and Temporal Cardinality

In the context of the Semantic Web, a number of Approaches have already been designed, addressing Different temporal aspects in relation to ontology languages. A rather wide approach towards extending ontology languages with a temporal dimension is Temporal RDF (Gutierrez et al., 2007). This work is similar to the towel language as it concerns the ability to represent temporal information in ontologies, but differs in that the language considered is the Resource Description Framework (RDF). Another approach is OWL-Time, which focuses on OWL rather than RDF. The initial purpose behind the design of time ontology (OWL-Time) (Hobbs et al., 2004) was to represent the temporal content of Web pages and the temporal properties of Web Services. This approach is rather extensive in describing quantitative time and the

Qualitative relations that may exist among instants and intervals [16].

I. Research On Qualitative Information

In addition to the deadlock (2007) study discussed earlier, several new research projects investigate the importance of qualitative information in finance. Our study is most closely related to contemporaneous work by Li (2006) and Davis, Piker, and Sedor (2006), who analyze the tone of qualitative information using objective word counts from corporate annual reports and earnings press releases, respectively. Whereas Davis, Piker, and Seder (2006) examine the contemporaneous relationship between earnings, returns, and qualitative information, Li (2006) focuses on the predictive ability of qualitative information as we do. Here is also some prior and contemporaneous research that analyses qualitative information using sophisticated subjective measures, rather than simple word counts.[17]. Neural networks are nonparametric, nonlinear models that can be trained to map past values of a time series, for purposes of classification or function estimation. We use a feed forward neural network with back propagation learning, which is the most conventional sort of neural network [18].

J. Announcement Return Measurement Error and Anticipation

Although the evidence thus far agrees with the conclusion that revisions are information-free, on average, a natural concern is that revisions provide analyst information, yet because of measurement error or investor anticipation of that information, the information is impounded in the large pre-returns, leaving the announcement return information-free. Here, we examine whether the 40-min review announcement window is too short, biasing the measured announcement returns toward zero. We assess the shortness bias concern by expanding the window to one hour and then to two hours, while controlling for confounding events. Designing a temporal extension of OWL-DL− begins with a clarification of what is understood under the general, common denominator time. We consider a couple of fundamental aspects hereof, namely: I) temporal infrastructure, and ii) change. The first aspect, temporal infrastructure, regards the representation of time in the form of instants and/or intervals. From this perspective, we aim for an approach that incorporates both a point-based as well as an interval-based time representation. Such an approach should provide not only the temporal entities that constitute the temporal infrastructure of the language, but also the relations that may hold between these entities, e.g., the before relation that may hold between intervals. [19][20].

K. Genetic Programming

In traditional genetic algorithms, genetic structures are represented as character strings of fixed length. This representation is adequate for many problems, but it is restrictive when the size or the form of the solution cannot be assessed beforehand. Genetic programming, developed by Koza (1992), is an extension of genetic algorithms that partly alleviates the restrictions of the fixed-length representation of genetic structures. For the purposes of this paper, the main advantage of genetic programming is the ability to represent different trading rules in a natural way. In genetic programming, solution candidates are represented as hierarchical compositions of functions. In these tree-like structures, the successors of each node provide the arguments for the function identified with the node. The deadly nodes (i.e., nodes with no successors) correspond to the input data. The entire tree is also interpreted as a function, which is evaluated recursively by simply evaluating the root node of the tree. The structure of the solution candidates is not specified a priori. Instead, a set of functions is defined as building blocks to be recombined by the genetic algorithm [21].

L. Multivariate Model

VaR is nearly always measured in a multivariate context. In this subsection, I present the multivariate generalization of the model given above. I will now use X to refer to an n × 1 vector of market risk factors:

$$X = (X1\ X2\ ...\ X_{n-1}\ X_n)'$$

The issue of correlation is crucial to multivariate modelling of market risk factors. In a multivariate context, we care about the risk of correlated jumps. For measuring market risk, we care about correlated jumps over a short time horizon (e.g., one day or ten days). The characteristics of the portfolio will dictate the correlation assumption that is appropriate for that portfolio. Because correlation is crucial, it will be important to validate whatever correlation assumption is made [22].

M. Applying Genetic Programming to Finding Trading Rules

This paper uses genetic programming to find technical trading rules for a composite stock index. The goal of the algorithm is to find decision rules that divide days into two distinct categories, either 'in' the market (earning the market rate of return) or 'out' of the market (earning the risk-free rate of return). Each genetic construction represents a particular technical trading rule. A trading rule is a function that returns either a 'buy' or a 'sell' signal for any given price history. The trading strategy specifies the position to be taken the next day, given the current position and the trading rule signal.

The trading strategy is implemented as a simple automaton, which works as follows: If the current state is 'in' (i.e., holding the market) and the trading rule signals 'sell', switch to 'out' (move out of the market); if the current state is 'out' and the trading rule signals 'buy', switch back to 'in'. In the other two cases, the current state is preserved. [23].

N. Finding and Evaluating Trading Rules

Most of the previous trading rule studies have sought to test whether particular kinds of technical rules have forecasting ability. A machine learning approach to finding testable trading rules sheds light on a question that is subtly different. Instead of asking whether specific rules work, we seek to find out whether in a certain sense optimal rules can be used to forecast future returns. Allowing a genetic algorithm to compose the trading rules avoids much of the arbitrariness involved in choosing which rules to test. Were we to use a more traditional search method, we would have to decide the form of the trading rules before searching through the parameter space for the chosen class of rules[24].

III. 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 event 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 variables are investigated.

IV. REFERENCES

[1] 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.

[2] S.B. Achelis, Technical Analysis from A to Z. McGraw-Hill, 2000.

[3] J. Koza, Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press,1992.

[4] H. Zhao, "A Multi-Objective Genetic Programming Approach to Developing Pareto Optimal Decision Trees," Decision Support Systems, vol. 43, no. 3, pp. 809-826, 2007.

[5] W. Fan, P. Pathak, and L. Wallace, "Nonlinear Ranking Function Representations in Genetic Pro-gramming-Based Ranking Discovery for Personal-ized Search," Decision Support Systems, vol. 42, no. 3, pp. 1338-1349, 2006.

[6] W.S. Chan, "Stock Price Reaction to News and No-News: Drift and Reversal After Headlines," J. Fi-nancial Economics, vol. 70, no. 2, pp. 223-260, 2003.

[7] 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.

[8] R.J. Rosen, "Merger Momentum and Investor Sen-timent: The Stock Market Reaction to Merger An-nouncements," J. Business, vol. 79, no. 2, pp. 987-1017, 2006.

[9] 9. 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.

[10] 10. A.J. Keown and J.M. Pinkerton, "Merger An-nouncements and Insider Trading Activity: An Em-pirical Investigation," J. Finance, vol. 36, no. 4, pp. 855-869, 1981.

[11] P.C. Tetlock, "More than Words: Quantifying Language to Measure Firms' Fundamentals," J. Fi-nance, vol. 63, no. 3, pp. 1437-1467, 2008.

[12] W. Leigh, R. Purvis, and J.M. Ragusa, "Forecasting the NYSE Composite Index with Technical Analy-sis, Pattern Recognizer, Neural Network, and Genet-ic Algorithm: A Case Study in Romantic Decision Support," Decision Support Systems, vol. 32, no. 4,pp. 361-377, 2002.

[13] 13. On the information role of stock recommendation revisions$ Oya Altınkılıc- a , Robert S. Hansen b, a Joseph M. Katz Graduate School of Business, Uni-versity of Pittsburgh, Pittsburgh, PA 15260, USA b A.B. Freeman School of Business, Tulane Universi-ty, New Orleans, LA 70118, USA.

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

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

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

[17] F. Allen and R. Karjalainen, "Using Genetic Algo-rithms to Find Technical Trading Rules," J. Eco-nomics, vol. 51, no. 2, pp. 245-271,1999.

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