2006 International Joint Conference on Neural Networks
Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada
July 16-21, 2006

Classify Unexpected News Impacts to Stock Price by Incorporating Time Series Analysis into Support Vector Machine

Ting Yu, Tony Jan, John Debenham and Simeon Simoff

Abstract— **the paper discusses an approach of using traditional time series analysis, as domain knowledge, to help the data-preparation of support vector machine for classifying documents. Classifying unexpected news impacts to the stock prices is selected as a case study. As a result, we present a novel approach for providing approximate answers to classifying news events into simple three categories. The process of constructing training datasets is emphasized, and some time series analysis techniques are utilized to pre-process the dataset. A rule-base associated with the net-of-market return and piecewise linear fitting constructs the training data set. A classifier mainly built by support vector machine uses the training data set to extract the interrelationship between unexpected news events and the stock price movements.**

I. Introduction

Standard classification relies heavily on the given training datasets, and often ignores the existing prior domain knowledge. Some reasons are the imperfect quality and uncertain characteristics of the prior domain knowledge, as well as the difficulties of incorporating knowledge into (inductive) machine learning in systematic ways. These limitations introduce more uncertainty to the learning process rather than improving it. Some reviews on incorporation of domain knowledge in machine learning are presented in [1]. Recently, machine learning techniques and data mining techniques have found much promise in financial industry with their capacity to learn from available financial data. In financial industry, data has been accumulated over the last few decades through government regulation and financial auditing purpose. However, a large amount of domain knowledge has also been collected through research over the last few decades; however, they remain dormant to be used to improve machine learning system in financial industry.

This paper describes an approach of incorporating some financial domain knowledge (e.g. Time Series Features and Patterns) into a machine learning system such as Support Vector Machine (SVM) to realize a solution for a well-known difficult problem: classifying the impacts of broadcasted news to the stock price movements for selected companies.

Ting Yu is with the Institute for Information and Communication Technologies, Faculty of Information Technology, University of Technology, Sydney, PO Box 123, Broadway, NSW 2007, Australia, and Capital Markets Cooperative Research Centre, Australia

Tony Jan, John Debenham and Simeon Simoff are with the Institute for Information and Communication Technologies, Faculty of Information Technology, University of Technology, Sydney, PO Box 123, Broadway, NSW 2007, Australia.

In macroeconomic theories, the Rational Expectations Hypothesis (REH) assumes that all traders are rational and take as their subjective expectation of future variables the objective prediction by economic theory. In contrast, Keynes already questioned a completely rational valuation of assets, arguing those investors' sentiment and mass psychology play a significant role in financial markets. New classical economists have viewed these as being *irrational*, and therefore inconsistent with the REH. Hence, financial markets are viewed as evolutionary systems between different, competing trading strategies [2]. In this uncertain market, nobody really knows what exactly the fundamental value of each stock is: good news about economic fundamental reinforced by some evolutionary forces may lead to deviations from the fundamental values and even overvaluation of the fundamental value.

Hommes C.H. [2] specifies the Adaptive Belief System (ABS), which assumes that traders are *boundedly rational*, and implied a decomposition of return into two terms: one martingale difference sequence part according to the conventional EMH theory, and an extra speculative term added by the evolutionary theory. The phenomenon of volatility clustering occurs due to the interaction of heterogeneous traders. High volatility may be triggered by news about fundamental values and may be amplified by technical trading. As a non-linear stochastic system, ABS is:

$$X_{t+1} = F(X_t; n_{1t},...,n_{Ht}; \lambda; \delta_t; \varepsilon_t)$$

Where F is a nonlinear mapping, the noise term ε_t is the model approximation error representing the fact that a model can only be an approximation of the real world.

Maheu and McCurdy [3] specified a GARCH-Jump model for return series. They label the innovation to returns, which is directly measurable from price data, as the news impact from latent news innovations. The latent news process is postulated to have two separate components, *normal* and *unusual* news events. These news innovations are identified through their impact on return volatility. The unobservable normal news innovations are assumed to be captured by the return innovation component, $\varepsilon_{1,t}$. This component of the news process causes smoothly evolving changes in the conditional variance of returns. The second component of the latent news process causes infrequent large moves in returns, $\varepsilon_{2,t}$. The impacts of these unusual news events are labelled *jumps*. Given an information set at time t-1, which consists of the history of returns

0-7803-9490-9/06/$20.00/©2006 IEEE

$\Phi_{t-1} = \{r_{t-1}, ..., r_t\}$, the two stochastic innovations, $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ drive returns: $r_t = \mu + \varepsilon_{1,t} + \varepsilon_{2,t}$, $\varepsilon_{1,t}$ is a mean-zero innovation ($E[\varepsilon_{1,t} \mid \Phi_{t-1}] = 0$) with a normal stochastic forcing process, $\varepsilon_{1,t} = \sigma_t z_t, z_t \sim NID(0,1)$ and $\varepsilon_{2,t}$ is a jump innovation.

Both of the previous models provide general framework to incorporate the impacts from news articles, but with respect to thousands of news articles from all kinds of sources, these methods do not provide an approach to figure out the significant news for the given stocks. Therefore, these methods cannot make significant improvement in practice.

Literatures describe machine-learning researches that try to predict short-term movement of stock prices. However, very limited researches have been done to deal with unstructured data due to the difficulty of the combination of numerical data and textual data in this specific field. Marc-Andre Mittermayer developed a prototype NewsCATS [4], which provides a rather complete framework. Being different from this, the prototype developed in this paper, gives an automatic pre-processing approach to build training datasets and keyword sets. Within the NewsCATS, experts do these works manually, and this is very time consuming and lack of flexibility to dynamic environments of stock markets. A similar work has been done by B. Wuthrich and V. Cho et al [5]. The following part of this paper emphasizes the pre-processing approach and the combination of the rule-based clustering and nonparametric classifications.

II. Methodologies and System Design

Being different from interrelationships among multiple sequences of numerical observations, heterogeneous data e.g. price (or return) series and news event sequences are considered in this paper. Normally, the price (or return) series is numerical data, and the news events are textual data. In the previous GARCH-Jump model, the component $\varepsilon_{2,t}$ incorporates the impacts from events into price series. However, the model does not provide a clear approach to measure or quantify the impact, and to our best knowledge, the existing solutions are to employ some financial experts, who measure the value of impact, $\varepsilon_{2,t}$. Moreover, considering thousands of news from all over the world, it is almost impossible for one individual to pick up the significant news and make a rational estimation immediately after they happen.

A prototype proposed by this paper develops an "alignment" technique between time stamp data sequences throughout the combination of domain knowledge and non-parametric data-driven classification.

To initiate this prototype, news items from the archive of press release and a closing price series from the closing price data archive are fed into the news pre-processing engine, and the engine tries to "align" news items to the price (or return series). After the alignment, training news items are labelled as three types of news using a rule-based clustering. Further the labelled news items are fed into a keywords extraction engine within the news pre-processing engine [6], in order to extract keywords to construct an archive of keywords, which will later be used to convert the news items into term-frequency data understandable to the classification engine.

After the training process is completed, the inflow of news will be converted as a term-frequency format and fed into the classification engine to predict its impact to the current stock price.

The impact from unexpected news to the stock price movement could be viewed as a conditional probability with

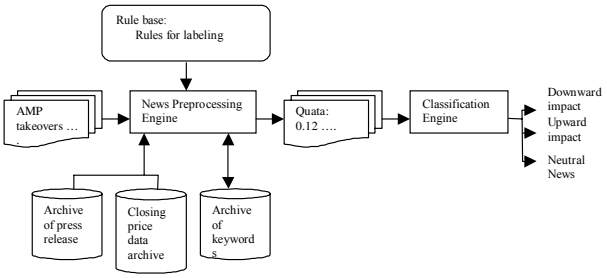

Fig. 1. Structure of the classifier

respect to the traders' current belief levels of company status and economical environments: Pr(Impact | Current Belief), where Impact $\subset$ {Positive, Negative, Neutral}. Due to lack of data about the current belief, this experiment assumes that the current belief levels are same.

III. Domain Knowledge Represented by Rule Bases

Supervised learning requires labelled training data for model learning. However it is a time-consuming task to label news items, which will construct the training dataset and be fed into the classifier. Here some domain knowledge can be utilized to facilitate this learning machine to prepare its training data. Ting Yu et al [1] has done some researches in incorporating domain knowledge into inductive machine learning. In this paper, the rule base represents domain knowledge, which integrates various discoveries from time series analysis to "align" the news items with the some time series patterns or features of stock price movements. In case of labelling unexpected news announcement, the causal links between the news arrival and short-range trend and unusual volatility are represented by knowledge about the subject area, and some time series analysis techniques are employed to discover these patterns or features from the stock price movements.

A. Domain Knowledge

Since the start of financial research, the correlation between information release and security price movement has been a prevailing topic. Charles Lee et al indicated that important information releases are already surrounded by dramatic price adjustment processes, e.g. extremely increase

of trading volume and volatility, and the process normally lasts up to one or two days [7]. These dramatic price discovery processes are often caused by unexpected news arrival, so-called "jump" in the GARCH-Jump model or "shock".

On the other hand, the previous ABS suggested that while high volatility may be triggered by news about fundamental values, the volatility may also be amplified by technical trading, and then the ABS implied a decomposition of return into two terms: one martingale difference sequence part according to the conventional EMH theory, and an extra speculative term added by the evolutionary theory. Borrowing some concepts from the electronic signal processing, volatility could be decomposed into two sets of disturbance: inherent structures of the process even without events, which are caused by traders' behaviours inside the market, and transient behavior reflecting the changes of flux after new event happens in the market. The transient problem may cause a shock at series of price (or return), or permanently change inherent structures of the stock, e.g. interrelationship between financial factors.

B. Using Domain Knowledge to help the data-preparation of Machine Learning

According to the previous financial domain knowledge, two time series analysis techniques, extreme volatilities detection and change point detection, are employed to find the desirable patterns. Then some rule-bases utilize these patterns to integrate two time series sequences together and align the news items sequence with the stock price movements. Here the rules are quite straightforward and consist of IF-THEN clauses, for example:

> **IF** a trading day is within downward trend and with large volatility, **THEN** this trading day has unanticipated news with negative impacts;
> **IF** a trading day is within upward trend and with large volatility, **THEN** this trading day has unanticipated news with positive impacts.

Collopy and Armstrong have developed some rule bases for time series forecasting. The objective of their rule base [8] are: to provide more accurate forecasts, and to provide a systematic summary of knowledge. The performance of their rule-based forecasting depends not only on the rule base, but also on the conditions of the series, where conditions mean a set of features that describes a series. The authors are inspired by their works, and further bridge the gap between the rule-based forecasting and numerical non-parametric machine learning.

The pseudo-code for an example of the algorithms:

```
/* Discovery Time Series Patterns*/
TrendDetection(ClosingPriceSquence);
VolatilityMeasure(ClosingPriceSquence);
/* Alignment between two time-series squences*/
While not finish the time series
    Lable=Rule_base(NewsItem, Date);
    Episode_array(Label)+=NewsItem;
End loop
Return Episode_array;

/**Rule-base**/
Rule_base(NewsItem, Date) {
    Rule 1: Date∈ {upward, neutral, downward};
    Rule 2: (Date == shock) ∈ {true, false};
    Switch {Rule 1, Rule 2}:
      case 1: label=upward impact;
      case 2: label=downward impact;
      case 3: label=neutral impact;
    End
    Return label;
}
```

In this paper, two time series analysis, net-of-market return and piecewise fitting, are employed to discover patterns and features: unusual high volatility and *a change in the basic trend* of a series. A piecewise regression line is fitted on the series to detect the level discontinuity and changes of basic trend. After detecting the change points, the next stage is to select an appropriate set of news stories. Victor Lavrenko et al [9] named this stage as "Aligning the trends with news stories".

C. Using Time Series Analysis to Discover Knowledge

John Roddick et al [10] described that time-stamped data can be scalar values, such as stock prices, or events, such as telecommunication signals. Time-stamped scalar values of an ordinal domain form curves, so-called "time series", and reveal trends. They listed several types of temporal knowledge discovery: A priori-like discovery of association rules, template-based mining for sequences, and classification of temporal data. In the case of trend discovery, a rationale is related to prediction: if one time series shows the same trend as another but with a known time delay, observing the trend of the latter allows assessments about the future behaviour of the former.

In financial research, the stock price (or return) is normally treated as a time series, in order to explore the autocorrelation between the current and previous observations. On the other hand, events, e.g. news arrival, may be treated as a sequence of observations, and it will be very significant to explore correlation between these two sequences of observations.

1) Unusual High Volatility of Stock Price Movement

In this paper, the observation beyond 3 standard derivations is treated as abnormal volatilities, and the news released within this day with abnormal volatilities will be labelled as shocking news.

Bing different from the often-used return, e.g. $R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$, the net-of-market return is the difference between absolute return and index return: $NR_t = R_t - IndexR_t$. This indicates the magnitude of information released and excludes the impact from the whole stock market.

2) Piecewise Linear Fitting to Detect the Trend of Price Movements

Piecewise linear fitting is to remove the inertial part of the series of return, i.e. the disturbance caused by traders' behaviours, which normally are around 70% total disturbances. Within the price series, a set of the piecewise linear models is fitted to the real price series and is used to detect the change of trend.

Eamonn Keogh et al [11] provides three major approaches to segment time series [12]: sliding windows, top-down and bottom-up. Here the bottom-up segmentation algorithm is used to fit piecewise linear functions to the price series. The piecewise segmented model M is given as in [13]:

$$\begin{aligned} Y &= f_1(t, w_1) + e_1(t), (1 < t < \theta_1) \\ &= f_2(t, w_2) + e_2(t), (\theta_1 < t < \theta_2) \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &= f_k(t, w_k) + e_k(t), (\theta_{k-1} < t < \theta_k) \end{aligned}$$

Where an $f_i(t, w_i)$ is the function that is fit in segment i. In case of the trend estimation, this function is a linear one between price and input date. The θ_i's are change points between successive segments, and $e_i(t)$'s are error terms.

In the piecewise fitting of a series of stock price, the connecting points of piecewise models represent the points of the significant changes in trends. In the statistics literature this is called the *change point detection* problem [13].

IV. Document Classification Using Support Vector Machine

The goal of text classification is the automatic assignment of documents, e.g. company announcements, to simple three categories. In this experiment, the commonly used Term Frequency-Inverse Document Frequency (TF-IDF) is utilized to calculate the frequency of predefined key words to represent documents as a set of term-vectors. The set of keywords is constructed by comparing general business articles from the website from the Australian Financial Reviews complemented by the company announcements collected and pre-processed by Dale [14]. Keywords are not restricted to single word but can be phrases. Therefore, the first step is to identify phrases in the target corpus. The phrases are extracted based on the assumption that two constituent words form a collocation if they co-occur a lot [6].

Documents are represented as a set of fields where each field is a term-vector. Fields could include the title of the document, the date of the document and the frequency of selected key words.

In corpus of documents, certain terms will occur in the most documents, while others will occur in just a few documents. The inverse document frequency (*IDF*) is a factor that enhances the terms that appear in fewer documents, while downgrading the terms occurring in many documents. The resulting effect is that the document-specific features get highlighted, while the collection-wide features are diminished in importance. TF-IDF assigns the term i in document k a weight computed as:

$$TF_{ik} * IDF(t_i) = \frac{f_k(t_i)}{\sqrt{\sum_{t_i \in D_k} f_k^2(t_i)}} * \log(\frac{n}{DF(t_i)})$$

Here DF (document frequency of the term (t_i)) – the number of documents in the corpus that the term appears; n – the number of documents in the corpus; TF_{ik} – the occurrence of term i at the document k [15]. As a result, each document is represented as a set of vectors $F^{dk} = < term_i, weight >$.

V. Experiments

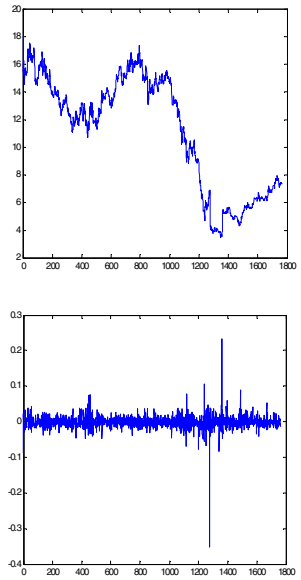

Fig. 2.1. Closing price and net-of-market return series of AMP

As a case study, the stock price and return series of an Australian insurance company, AMP, were studied. Figure 2.1 shows the closing price and net return series of AMP

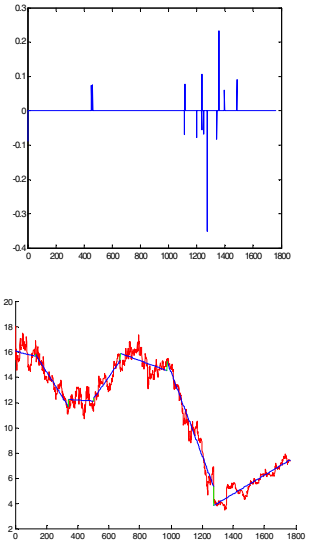

Fig. 2.2. Shocks (large volatilities) and Trend and changing points

from 15/06/1998 to 16/03/2005. At the same time, more than 2000 company announcements from AMP are collected as a series of news items, which covers the same period as the closing prices series.

Figure 2.2 indicates the shocks (large volatilities) and the trend changing points detected by the piecewise linear model fitting. After pre-processing, the training dataset consists of 464 upwards news items, 833 downward news items and 997 neutral news items. The keyword extraction algorithm constructs a keywords set consisting of 36 single or double terms, e.g. vote share, demerg, court, qanta, annexure, pacif, execut share, memorandum, cole etc. These keywords are stemmed from the Porter Stemming Algorithm [16].

The dataset is split into two parts: training and test data. The result of classification, e.g. upwards or downwards, is compared with the real trends of the stock price. Under LibSVM v2.8 [17], the accuracy of classification is 65.73%, which is significant higher than 46%, the average accuracy of Wuthrich's experiments which is the traditional industry standard approach [5].

VI. Conclusions and Further Work

This paper provides a framework of classifying the upcoming financial news into three categories: upward, neutral or downward. One of the main purposes of this research is to explore an approach of incorporating domain knowledge into some inductive machine learning, which traditionally is purely data-driven. Another main purpose is to provide financial participants and researchers an automatic and powerful tool to screen out influential news (information shocks) among thousand of news around this world everyday.

The current prototype has demonstrated promising results of this approach and further work will provide prototype that can analysis the impact of news on stock prices in more complex real environment.

In the further work, three major issues need to be concerned, which are suggested by Nikolaus Hautsch [18]: 1) Impact from the disclosure of inside information: if inside information has been disclosed at the market even before the announcement, the price discovery process will be different. 2) Anticipated vs. unanticipated information: if traders' belief has absorbed the information, so-called anticipated information, the impact must be expressed as a conditional probability with the brief as a prior condition. 3) Interactive effects between information: at the current experiment all news at one point are labelled as a set of upward impacts or other, but the real situation is much more complex. Even at one upward point, it is common that there is some news with downward impacts. It will be very challenging to distinguish the subset of minor news and measure the interrelationship between news.

Acknowledgment

The authors would like to thank Dr. Debbie Zhang and Paul Bogg for their invaluable comments and discussion, thank Prof Robert Dale for his company announcements as XML formats, and thank Dr Eamonn Keogh for his Matlab code of linear piecewise fitting

References

1. Yu, T., T. Jan, J. Debenham, and S. Simoff. *Incorporating Prior Domain Knowledge in Machine Learning: A Review*. in *AISTA 2004: International Conference on Advances in Intelligence Systems - Theory and Applications in cooperation with IEEE Computer Society*. 2004. Luxembourg.
2. Hommes, C.H., *Financial Markets as Nonlinear Adaptive Evolutionary Systems*, in *Tinbergen Institute Discussion Paper*. 2000, University of Amsterdam.
3. Maheu, J.M. and T.H. McCurdy, *News Arrival, Jump Dynamics and Volatility Components for Individual Stock Returns.* Journal of Finance, 2004. **59**(2): p. 755.
4. Mittermayer, M.-A. *Forecasting Intraday Stock Price Trends with Text Mining Techniques*. in *The 37th Hawaii International Conference on System Sciences*. 2004.
5. Wuthrich, B., et al., *Daily Stock Market Forecast from Textual Web Data*, in *IT Magazine*. 1998. p. 46-47.
6. Zhang, D., S.J. Simoff, and J. Debenham. *Exchange Rate Modelling using News Articles and Economic Data*. in *The 18th Australian Joint Conference on Artificial Intelligence*. 2005. Sydney Australia.
7. Lee, C.C., M.J. Ready, and P.J. Seguin, *Volume, volatility, and New York Stock Exchange trading halts.* Journal of Finance, 1994. **49**(1): p. 183-214.
8. Collopy, F. and J.S. Armstrong, *Rule-Based Forecasting: Development and Validation of an Expert Systems Approach to Combining Time Series Extrapolations.* Journal of Management Science, 1992. **38**(10): p. 1394-1414.
9. Lavrenko, V., et al., *Language Models for Financial News Recommendation*. 2000, Department of Computer Science, University of Massachusetts: Amhers, MA.
10. Roddick, J.F. and M. Spiliopoulou, *A survey of temporal knowledge discovery paradigms and methods.* Knowledge and Data Engineering, IEEE Transactions on, 2002. **14**(4): p. 750-767.
11. Keogh, E., S. Chu, D. Hart, and M. Pazzani. *An Online Algorithm for Segmenting Time Series.* in *In Proceedings of IEEE International Conference on Data Mining*. 2001.
12. Keogh, E., S. Chu, D. Hart, and M. Pazzani, *Segmenting Time Series: A Survey and Novel Approach*, in *Data Mining in Time Series Databases*. 2003, World Scientific Publishing Company.
13. Guralnik, V. and J. Srivastava. *Event Detection From Time Series Data*. in *KDD-99*. 1999. San Diego, CA USA.
14. Dale, R., R. Calvo, and M. Tilbrook. *Key Element Summarisation: Extracting Information from Company Announcements*. in *Proceedings of the 17th Australian Joint Conference on Artificial Intelligence*. 2004. Cairns, Queensland, Australia.
15. Losada, D.E. and A. Barreiro, *Embedding term similarity and inverse document frequency into a logical model of information retrieval.* Journal of the American Society for Information Science and Technology, 2003. **54**(4): p. 285 - 301.
16. Porter, M., *An algorithm for suffix stripping.* Program, 1980. **14**(3): p. 130-137.
17. Chang, C.-C. and C.-J. Lin, *LIBSVM: a Library for Support Vecter Machine*. 2004, Department of Computer Sicence and Information Engineering, National Taiwan University.

18. Hautsch, N. and D. Hess, *Bayesian Learning in Financial Markets - Testing for the Relevance of Information Precision in Price Discovery.* Journal of Financial and Quantitative Analysis, 2005.