SAILING THROUGH BUSINESS INFORMATION

Radu Adrian CIORA [1], Carmen Mihaela SIMION [2]

[1, 2] Lucian Blaga University of Sibiu, Romania

Abstract
Business information is a key aspect in any business. The evolution and transformations through which business information undergoes in time are impressive. These changes start from information sources, continue with inside information and end with customer information. Although we have specified distinct groups of information, these three groups are not independent, but rather tightly coupled.
In an interconnected world, information is vital. The movement from product to product as a service, or product with services, has made a significant change in the way companies handle their relationships with customers as well as with employees. Information technology use in a global economy reduces the barrier of space between businesses, thus allowing better interactions among partner businesses even when they are in different geographical areas. Thus, loads of information can instantly be sent to a server, which is in cloud nowadays, where this information can be analysed in almost real-time and decisions can be made on the fly, permitting businesses to adapt to the shifts of the market. In this paper, based on the developments in the field of business information, we present a general-purpose flexible business information model, which can be useful for business owners and top-level management for information management in order to develop a more context-aware business plan.

Keywords: business information, information sources, business plan
JEL classification: M21, M31

1. Introduction

A good business administration and management require access to last minute information in order to take correct business decisions. These decisions need to be based on elaborated information, timely received – which, nowadays cannot be achieved without the use of computer science.

[1] *PhD student, Faculty of Engineering, e-mail: radu.ciora@ulbsibiu.ro*
[2] *Professor, Faculty of Engineering, e-mail: carmen.simion@ulbsibiu.ro*

For managers, building an information management system might seem a waste of resources; because it is a common believe that the benefit/cost ratio is subunit (less than 1). If we analyse the damage generated by an erroneous decision, caused by the lack of information, it turns out that an information system is indispensable.

In the digital era, it is impossible to process the amount of information that is generated. Thus, an automated, on-demand information processing system is a must. This is why we came up with the idea of an automated information processing system in the financial domain, that can process data retrieved from various information sources. We focused our research on analysing RSS feeds from financial websites (e.g. Investopedia), in order to determine and predict the market evolution, but also the evolution of specific assets that a certain business owner might be interested in.

2. Previous Work

The prediction of stock market consists of the process of determining the future values of certain assets based on previous price evolution observations. Although it is a common belief, that the volatility of stock market cannot be precisely predicted, there are several approaches in the literature regarding this matter, which have certain degrees of reliability, starting with the use of fractal theory for market prediction by Edgar Peters, continuing with price prediction based on past prices interpreted as a time series and ending with data mining techniques.

Nowadays methods imply machine learning techniques, like support vector machines (SVM), which can have prediction rates around 55% accuracy as in the work of Kyoung-jae Kim. A similar work is one of Chen et all where their implementation introduced a regression based SVM learning techniques.

As we did not find evidence of data augmentation in any of the approaches presented in the literature. Although, it might not be obvious that sentiments have anything to do with stocks, we cannot forget that we are dealing with people, which are the end buyers and sellers, who are by nature sentimentalists. Thus, we consider that introduction of sentiment analysis must be introduced in the equation of stock prediction.

3. Methodology

We propose a method for data analysis that aims to establish the stock market price evolution, by extracting features from economic RSS feeds and processing them with a pretrained machine learning algorithm. Word extraction is based on a clustered category lists of words which need to be weighted in order to have better prediction results.

The weighting algorithm is based on the word frequency and it is characterized by the following equations:

(1)

where σ^2 is defined as:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} \tag{2}$$

where $\mathbf{x_i}$ is the current occurrence of a word,  is the weighted mean value and **n** represents the number of iterations, where the mean is defined by:

(3)

Thus,  gives an interpretation of the analysed text, trying to infer its trend.

In order to process the information from various sources it first needs to be trimmed. In order to achieve this, we propose 32 unweighted word categories (achievements, anger, anxiety, certain, determine, family, feel, financial, friends, hear, home, inhibition, intuition, leisure, nationalism, negative, people, perceptive, positive, rational, religion, sadness, see, sexual, social, work, emotional, failures, success, swear, uncertain, slang). Each of these classes have on average 10 words. Trimming refers to extracting only the words from RSS feeds that belong to these classes.

In the next step we augment the words extracted with contextual information like the company name, owner, and other existing information.

We then proposed a weighed support vector machine learning algorithm, which predicts the future values, based on previous $n-1$ values. In order to train the machine learning network in the learning phase we used a set of documents manually labelled as having positive influence on stock price and another negatively labelled. In the training phase use used 10 classes, which enable the SVM to adjust its weights. We then retrained the network against another set of 10 classes different from the first set and so the SVM readjusted the weights again and after the we tested the network against the remaining classes.

4. Results and Conclusions

The results seem promising but are directly depend on the words we used in the dictionary and the classes we attached them to. The method used for analysing the various data sources aims to establish the stock future price evolution by extracting features based on a dictionary composed of classes that contain words.

In order to validate our initial results, a new approach to text categorization was designed, by using learning algorithms. From these algorithms augmented Bayesian support vector machines proved to be the most suitable for text categorization. As the training set is augmented, there is no need for parameters adjustment as they are predetermined in the training phase.

Although support vector machines proved feasible, on long term predictions, their prediction rates prove to deprecate, which is normal as they cannot readjust its parameters learnt information. Thus, a future research direction would be the use of deep learning, which support unsupervised learning and can adjust its weights based on newly acquired knowledge.

5. References

- *Barbara, B., Executives Are "Ego Surfing" The Web for Business Information, Available at: http://career-intelligence.com/ego-surfing-for-business-information*
- *Chen, W.-H., Shih, J.-Y., Wu, S., Comparison of support-vector machines and back propagation neural networks in forecasting the six major Asian stock markets, International Journal of Electron Finance, 1(1), pp. 49–67, 2006, DOI:10.1504/ IJEF.2006.008837*
- *Gîfu, D., Teodorescu, M., Communication concepts vs. Sciences concepts, International Letters of Social and Humanistic Sciences (ILSHS) 18, 48-57, 2014*
- *Gîfu, D., Teodorescu, M., Neutrosophic routes in multiverse of communication, Neutrosophic Sets and Systems, book series 6, 81, 2014*
- *Kim, K., Financial time series forecasting using support vector machines, Neurocomputing, vol. 55, 2003*
- *Krishna, K.S., Priti, D., Krishna, N.R., The Implications of Big Data in Indian Stock Market, International Journal of Computer Applications 99 (2), 8-11, 2014*
- *Mesnil, G. He, X. Deng, L., Bengio, Y., Investigation of Recurrent-Neural-Network Architectures and Learning Methods for Spoken Language Understanding, Interspeech, 2013*
- *Modrak, V., Teodorescu, M., Gîfu, D., Highlights on mental processes in management, International Letters of Social and Humanistic Sciences (ILSHS) 24, 56-65, 2014*
- *Novak-Marcincin, J., Gîfu, D., Nicolescu, A., The standard of axes in ontology of communication, International Letters of Social and Humanistic Sciences 30 (2), 176-183, 2014, DOI:10.18052/www.scipress.com/ILSHS.41.176*
- *Peters, E., Fractal Market Analysis: Applying Chaos Theory to Investment and Economics, pp.336, Wiley & Sons, 1994*
- *Porter, M.E., "Competitive Advantage", 1, 11-15, The Free Press, New York, 1985*