Sentiment Analysis of News Headlines for Stock Price Prediction

Research · April 2016

DOI: 10.13140/RG.2.1.4606.3765

2 authors:

D. K. Kirange
J. T. Mahajan College of Engineering Faizpur
19 PUBLICATIONS **94** CITATIONS
SEE PROFILE

Ratnadeep R. Deshmukh
Dr. Babasaheb Ambedkar Marathwada University
227 PUBLICATIONS **577** CITATIONS
SEE PROFILE

ISSN:2320-0790

An International Journal of Advanced Computer Technology

Sentiment Analysis of News Headlines for Stock Price Prediction

Mr. D. K. Kirange[1], Dr. Ratnadeep R. Deshmukh[2]

[1]Dept. of Computer and IT, J T Mahajan College of Engineering, Faizpur, Tal, Yawal, Dist. Jalgaon

[2]Dept. of Computer Science and IT Dr. Babasaheb Ambedkar Marathwada University, Aurangabaad

Abstract: Stock market data analysis needs the help of artificial intelligence and data mining techniques. The volatility of stock prices depends on gains or losses of certain companies. News articles are one of the most important factors which influence the stock market. This study basically shows the effect of emotion classification of financial news to the prediction of stock market prices. In order to find correlation between sentiment predicted from news and original stock price and to test efficient market hypothesis, we plot the sentiments of two companies (Infosys and Wipro) over a period of 10 years. For emotion classification, various classifiers such as Naive Bayes, Knn and SVM are evaluated. The comparison between positive sentiment curve and stock price trends reveals co-relation between them.

Keywords: Sentiment Analysis, Naive Bayes, SVM, KNN.

I. INTRODUCTION

Stock market is the bone of fast emerging economies such as India. Major of capital infusion for companies across the country was made possible only thru shares sold to people. So our country growth is tightly bounded with the performance of our stock market. Considering the fact of lack of knowledge and awareness across the people stock market prediction techniques plays a very crucial role in bringing more people into market as well as to retain the existing investors. There is a strong yet complicated relation between the market and the information available in the form of news. The arrival of news at every moment changes the perception or sentiment towards a particular company or their adopted business strategies. These days due to the bliss of internet, the traders, and investors have constant access to the updated news, and the news constantly mould their sentiments and influences their decision to invest in a particular company.

News has always been an important source of information to build perception of market investments. As the volumes of news and news sources are increasing rapidly, it's becoming impossible for an investor or even a group of investors to find out relevant news from the big chunk of news available. But, it's important to make a rational choice of investing timely in order to make maximum profit out of the investment plan. And having this limitation, computation comes into the place which automatically extracts news from all possible news sources, filter and aggregate relevant ones, analyze them in order to give them real time sentiments. In this paper we have collected, classified and analyzed relevant real time news from widely accepted and trusted news sources available in the public domain to forecast the perception around a particular company which helps the investors and traders to come up with informed, rational investment plan timely which ensures maximized profit. In this study we also have tracked the sentiments of the companies over a period of 10 years in order to find out correlation between the emotions predicted from the relevant news and the original stock price curve's ups and downs.

The main contributions of the paper are as follows:

Implementation of an efficient prediction model that scores emotions from all relevant real time news available in public domain.

This model classifies all of the relevant real time news from the large database of news.

This forecasts real time news sentiment that reflects stock price movement trends.

Compares and finds out a correlation between positive sentiment movement and original stock price curve of 2 companies over a period of 10 years.

The rest of the paper is organized as follows. Section II outlines the existing work in this field of emotion classification and concept of stock price prediction. Section III proposes and describes our predictive model. Section IV we evaluate the prediction model using different accuracy measurement techniques. In this section, we also compare and assess the co-relation of the sentiment trends of the proposed model with the original stock price movement. Finally in Section V, we conclude our paper and discuss the future work related to it.

II. RELATED WORK

Prediction of stock price variation is a very challenging task and the price movement behaves more like a random walk and time varying. In recent times, researchers have used various types of AI techniques to make trading decisions. Here, we present a brief review of some of the significant researches.

A Sheta [1] has used Takagi-Sugeno (TS) technique to develop fuzzy models for two nonlinear processes. They were the software effort estimation for a NASA software projects and the prediction of the next week S&P 500 for stock market. The development of the TS fuzzy model can be achieved in two steps 1) the determination of the membership functions in the rule antecedents using the model input data; 2) the estimation of the consequence parameters. They used least-square estimation to estimate these parameters. The results were promising.

Ching Long Su *et al.* [2] have developed a self-organized, five-layer neuro-fuzzy model to model the dynamics of stock market by using technical indicators. The model effectiveness in prediction and forecasting was validated by a set of data containing four indicators: the stochastic oscillator (%K and %D), volume adjusted moving average (VAMA) and ease of movement (EMV) from TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock Index). A modified moving average method can be proposed to predict the input set for the neuro-fuzzy model in forecasting stock price. Simulation results have shown that the model was effective in prediction and accurate in forecasting. The input error from the prediction of the modified moving average method attenuated significantly by the neuro-fuzzy model to yield better forecasting results.

M.H. FazelZarandi*et al.* [3] have developed a type-2 fuzzy rule based expert system for stock price analysis. Interval type-2 fuzzy logic system permitted to model rule uncertainties and every membership value of an element was interval itself. The proposed type-2 fuzzy model applied the technical and fundamental indexes as the input variables. The model can be tested on stock price prediction of an automotive manufactory in Asia. Through the intensive experimental tests, the model had successfully forecasted the price variation for stocks from different sectors. The results were very encouraging and implemented in a real-time trading system for stock price prediction during the trading period.

Robert K. Lai *et al.* [4] have established a financial time series-forecasting model by evolving and clustering fuzzy decision tree for stocks in Taiwan Stock Exchange Corporation (TSEC). The forecasting model integrated a data clustering technique, a fuzzy decision tree (FDT), and genetic algorithms (GA) to construct a decision-making system based on historical data and technical indexes. The set of historical data can be divided into k sub-clusters by adopting K-means algorithm. GA was then applied to evolve the number of fuzzy terms for each input index in FDT so the forecasting accuracy of the model can be further improved. A different forecasting model was generated for each sub-cluster. In other words, the number of fuzzy terms in each sub cluster was different. Hit rate applied as a performance measure and the proposed GAFDT model had the best performance when compared with other approaches on various stocks in TSEC.

Shyi-Ming Chen and Yu-Chuan Chang [5] have presented a method for multi-variable fuzzy forecasting based on fuzzy clustering and fuzzy rule interpolation techniques. First, the proposed method constructed training samples based on the variation rates of the training data set and then used the training samples to construct fuzzy rules by making use of the fuzzy C-means clustering algorithm, where each fuzzy rule corresponded to a given cluster. Then, they determined the weight of each fuzzy rule with respect to the input observations and used such weights to determine the predicted output, based on the multiple fuzzy rules interpolation scheme. They applied the proposed method to the temperature prediction problem and the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) data. The experimental results have shown that the proposed method can be produced better forecasting results than several existing methods.

WengLuen Ho *et al.* [6] have proposed an interday financial trading system with a predictive model empowered by a novel brain-inspired evolving Mamdani-Takagi-Sugeno Neural-Fuzzy Inference System (eMTSFIS). The eMTSFIS predictive model possessed synaptic mechanisms and information processing capabilities of the human hippocampus, resulted in a more robust and adaptive forecasting model as compared to existing econometric and neural-fuzzy techniques. The trading strategy of the proposed system was based on the moving-averages-convergence/divergence (MACD) principle to generate buy-sell trading signals. By introducing forecasting capabilities to the computation of the MACD trend signals, the lagging nature of the MACD trading rule can be addressed. Experimental results based on the S&P500 Index confirmed that eMTSFIS was able to provide highly accurate predictions and the resultant

system was able to identify timely trading opportunities while avoiding unnecessary trading transactions. These attributes enabled the eMTSFISbased trading system to yield higher multiplicative returns for an investor.

EsmaeilHadavandi*et al.* [7] have presented an integrated approach based on genetic fuzzy systems (GFS) and artificial neural networks (ANN) for constructing a stock price forecasting expert system. At first, they used stepwise regression analysis (SRA) to determine factors which have most influence on stock prices. At the next stage, they divided their raw data into k clusters by means of self-organizing map (SOM) neural networks. Finally, all clusters were fed into independent GFS models with the ability of rule base extraction and data base tuning. They evaluated capability of the proposed approach by applying it on stock price data gathered from IT and Airlines sectors, and compared the outcomes with previous stock price forecasting methods using mean absolute percentage error (MAPE). Results have shown that the proposed approach outperformed all previous methods, so it can be considered as a suitable tool for stock price forecasting problems.

Kelvin Sim*et al.* [8] have proposed a method, 3D subspace clustering to generate rules to pick potential undervalued stocks. 3D subspace clustering is effective in handling high dimensional financial data and is adaptive to new data. The obtained results are not influenced by human's biases and emotions, and are easily interpretable. The conducted extensive experimentation in the stock market over a period of 28 years (from 1980 to 2007), and found that using rules generated by 3D subspace clustering algorithms, CAT Seeker and MIC, results in 60% more profits than using Graham's rules alone.

Anil Rajput [9] have proposed a rule based classification method which generates rules with preprocessed data. It focus on each class separately and increases the probability of exact classification. PRISM algorithm is applied to generate rules by examining the training data and identifying suitable rules that covers majority of class. Also it measures the success of each rules. It serves as a suitable tool for generating rules for buy and sell.

III. PROPOSED SYSTEM

The Figure 1 shows the overall outline of our system. The first step consists of news gathering. In this step, we got a scrapping of the online news from Indian Companies such as Infosys an d Wipro over period of 10 years and stored the data in a database. The next step consisted of natural language processing in order to extract emotion from unstructured news texts. We removed stop words. The remaining useful words were then used to analyze the emotions. For emotion classification we have evaluated the performance using Naïve bayes, SVM classifiers. We conducted supervised learning experiments aimed at predicting rises and falls in stock prices.

A. Gathering Text News

In this study the data have been collected from the press releases by the respective companies and the regular financial news sources – Wall Street Journal, Financial Times, Forbes, Reuters, CNBC, NDTV, Economic Times, Hindustan Times, Times of India, India Times, Telegraph, CNN Money, Market Watch, and Fortune

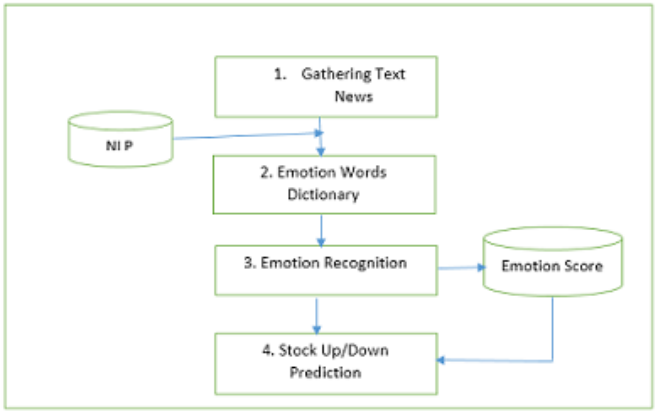

Figure 1: Proposed System

B. Emotion Words Dictionary

Analyzing the emotion of the news involves recognizing and defining the 'emotional state' expressed in the text. The emotion word dictionary plays a crucial role in opinion mining to build linguistic resources, which classify emotion polarity and discriminate between emotions. We have used the dictionary of positive and negative keywords i.e. dictionary of LIWC [11].

C. *Emotion Recognition*

Figure 2 shows the emotion classification for text news. The raw news data collected from various Indian companies websites is preprocessed for stop words removal. The emotion words dictionary of LIWC is used for construction of the feature matrix. The appropriate classifier is trained using the feature matrix and human annotated emotion prediction. The responsibility of the classifier is to trim the corpus offline. The system is tested using various classifiers such as SVM, Knn and Naïve Bayes.

Figure 2: Emotion Classification

Sentiment Term Polarity	Accuracy		
	SVM	KNN	Naïve Bayes
Positive	71.9212	46.58	48.59
Negative	96.4123	81.7734	52.76
Neutral	90.6404	71.0660	88.42
Conflict	99.5074	69.13	65.89
Average Accuracy	90.5172	67.1373	63.91

Table I: Accuracy for Infosys Dataset

D. Stock Up/Down Prediction

The sentiment score should be changed an opinion to predict a rise or fall in the stock price. An opinion for prediction is decided to be positive, or not on a certain point. A next trading day's stock price is predicted to rise, or, conversely, predicted to fall. The emotion classification of the news headlines over the period of 10 years is correlated with the stock price up/down.

IV. EVALUATION

We have identified the various sentiment categories as positive, negative, neutral and conflict etc. The well known Naïve Bayes, Knn and SVM classifiers are used for classification of a stock news headlines in appropriate category. The result is compared with the human annotations and the accuracy is computed as Accuracy = sum (abs (Expected output – actual output))/2

Sentiment Term Polarity	Accuracy		
	SVM	***KNN***	***Naïve Bayes***
Positive	62.7273	60.00	56.88
Negative	63.6364	49.29	69.73
Neutral	78.1818	24.5098	76.45
Conflict	97.2727	58.16	87.52
Average Accuracy	75.4545	47.9899	72.64

Table II: Accuracy for Wipro Dataset

The dictionary of LIWC is used for positive and negative keywords. The SVM [10] classifier is trained with the term document matrix prepared using the LIWC dictionary of keywords. The result is compared with the human annotations provided and the accuracy is computed as above. We have tested the performance of SVM classification with KNN and Naïve Bayes as shown in table I and II. The emotions classified using SVM are correlated with the BSE stock prices as bellow [12]. Evaluation of the result was done in order to measure accuracy of the prediction model and also to compare and find out the co-relation between predicted sentiment trends and original stock price movement, if at all any.

V. CONCLUSIONS

In this paper, we have proposed a predictive model to predict sentiment around stock price. First the relevant real time news headlines and press-releases have been filtered from the large set of business news sources, and then they have been analyzed to predict the sentiment around companies. In order to find correlation between sentiment predicted from news and original stock price and to test efficient market hypothesis, we plot the sentiments Infosys over a period of 10 years.

Although the accuracy level of this predictive model is satisfactory, it can further be developed by incorporating more complex classifiers, and analysis techniques of machine learning or data mining. The future scope of this paper can be varied. We can compare the sentiment of several competitive companies and show how the sentiment around one company impacts on the other companies over the time. Also we can also study other factors ranging from social to legal issues to check whether at all they have any impact on overall market scenario. Using this proposed model, we can do several financial modeling like portfolio management, risk estimation models, and several other strategic modeling etc. where perception plays an important role.

YEAR	CORRELATION			
	% of Positive News Using SVM	Stock Price (High)	***Stock Price (Low)***	***Up/Down***
2005	52.56%	379.97	231.25	Down
2006	78.85%	600.25	306.25	Up
2007	80.02%	609.75	377.79	Up
2008	75.42%	504.25	260.00	Down
2009	83.29%	652.48	275.25	Up
2010	89.85%	861.75	583.25	Up
2011	91.72%	873.49	542.25	Up
2012	87.56%	747.50	525.42	Down
2013	93.78%	893.75	547.50	Up
2014	94.44%	1100.25	723.50	Up

ACKNOWLEDGMENT

The authors would like to welcome the reviewers for their critical and constructive comments and suggestions. The authors would like to thank the Department of Computer Science of Dr. Babasaheb Ambedkar Marathwada

University, Aurangabaad for providing the infrastructure to carry out the research.

VI. REFERENCES

[1] A Sheta, "Software Effort Estimation and Stock Market Prediction Using Takagi-Sugeno Fuzzy Models", In Proceedings of the IEEE International Conference on Fuzzy Systems, pp.171-178, Vancouver, BC, 2006.

[2] Ching Long Su, ChuenJyh Chen and Shih Ming Yang, "A self-organized neuro-fuzzy system for stock market dynamics modeling and forecasting", WSEAS Transactions on Information Science and Applications,Vol7.8,No.9, September 2010 .

[3] M.H. FazelZarandi, B. Rezaee, I.B. Turksen and E. Neshat, "A type-2 fuzzy rule-based expert system model for stock price analysis", Expert Systems with Applications,Vol.36, No.1, pp. 139-154,January 2009.

[4] Robert K. Lai, Chin-Yuan Fan, Wei-Hsiu Huang and Pei-Chann Chang, "Evolving and clustering fuzzy decision tree for financial time series data forecasting", An International Journal of Expert Systems with Applications, Vol.36,No.2, pp. 3761-3773, March 2009.

[5] Shyi-Ming Chen and Yu-Chuan Chang, "Multi-Variable Fuzzy Forecasting Based On Fuzzy Clustering and Fuzzy Rule Interpolation Techniques", Information Sciences, Vol.180, No.24, pp. 4772-4783,2010.

[6] WengLuen Ho, Whye Loon Tung and Chai Quek, "Brain-Inspired Evolving Neuro-Fuzzy System for Financial Forecasting and Trading of the S&P500 Index", Lecture Notes in Computer Science, Vol.6230, pp.601-607,2010

[7] EsmaeilHadavandi, Hassan Shavandi and ArashGhanbari, "Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting", Knowledge-Based Systems, Vol.23, No.8, pp. 800-808,December 2010.

[8] Kelvin Sim, VivekanandGopalkrishnan, Clifton Phua and Gao Cong, "3D Subspace Clustering for Value Investing", IEEE Intelligent Systems, Vol. PP, No.99, pp. 1, 2012.

[9] Anil Rajput , S.P. Saxena , Ramesh Prasad Aharwal and RituSoni, "Rule based Classification of BSE Stock Data with Data Mining", International Journal of Information Sciences and Application, Volume 4, Number 1 (2012), pp. 1-9.

[10] XianggaoCai, Su Hu, XiaolaLin,"Feature Extraction Using Restricted Boltzmann Machine for Stock Price Prediction", 978-1-4673- 0089-6/12/$26.00 ©2012 IEEE.

[11] http://www.liwc.net

[12]http://www.moneycontrol.com/stocks/companydetails/hist_graph.php