# Text mining current news to predict short term stock share price movement

Michael Ka Ming Cheung[1] and Simon KC Cheung[2]

[1] The Shatin College
[2] The University of Hong Kong
simonkc@hku.hk

**Abstract.** Due to the advance in technology, the web has become an important source of information. For example, news articles and opinion blogs reveal sentiments on many events. Financial news contain useful information on publicly traded companies. Since prediction of a stock share price movement is an important activity in finance, data mining and analysis of financial information can aid such prediction. This article explores the application of an existing text mining techniques and an adjusted Bayesian classifier on current news downloaded from CNBC website to predict the direction of the underlying stock share price movements of the following week after the news is released. The share price movement is determined by the volume weighted average of the share prices of the following 10 days after the news are released. In our experiment, we find that our proposed method is able to predict the direction of the short term share price movement of the underlying stock at a high rate of accuracy.

**Keywords:** Stock Prediction, News, Text mining.

## 1 Introduction

Stock share price movement reflects collectively the change in future expectation of the financial performance of the underlying company. Besides the change in prevailing market sentiment, usually driven by the volatility of the world markets, a stock price can be fluctuated for many reasons. Without a prior knowledge, a stock price fluctuation is often event-driven. Nowadays, events are best summarized by news scripts. As web technologies continue to be embraced by the financial sector, these news scripts are delivered by financial news websites and are posted within a very short time after the event happened. There are many financial news providers such as the Bloomberg, Reuters and the Wall Street Journal. Although a stock price will eventually move to reflect the impact by its news, there is a reaction time lag.

An example is a news downloaded on 17 May 2016 10:19:55:

"Corvex disclosed a 9.9 percent stake in Pandora on Monday, making it the largest single shareholder in the company. Pandora's stock rose more than 6 percent Tuesday as the music streaming company faces pressure from an activist investor to sell some of its assets. We have become increasingly concerned that the company may be pursuing a costly and uncertain business plan, without a thorough evaluation of all shareholder value-maximizing alternatives, Corvex Management, a hedge fund run by Keith Meister, said in a Monday letter to Pandora shareholders. Despite its many strengths, the company has been unable to date to translate its great product into a great business with an attractive public market valuation."

The following table displays the share prices and volume traded for Pandora (ticker: P) on the following 10 days:

**Table 1.** Share Prices and Volume traded of P after the news is released.

| | **+0** | **+1** | **+2** | **+3** | **+4** | **+5** |
|---|---|---|---|---|---|---|
| **SHARE PRICE** | 11.65 | 12.08 | 12.21 | 12.49 | 12.25 | 12.08 |
| **VOLUME** | | 158373 | 173093 | 136052 | 222729 | 152887 |
| | +6 | +7 | +8 | +9 | +10 | |
| **SHARE PRICE** | 12.21 | 11.84 | 11.67 | 11.25 | 11.12 | |
| **VOLUME** | 156829 | 184566 | 243454 | 193061 | 171841 | |

As the news suggests a downward pressure on the stock Pandora, there are tremendous volume traded every time share price of the stock are peaked. This pushes the share price down to below the price when the news was released.

A second example is downloaded on 15 Nov 2016 10:21:16:

"Cramer: What Home Depot earnings tell us about consumer spending Home Depot's earnings report emerged as one of few positive reports in the home-improvement sector this quarter, a development that gave CNBC's Jim Cramer faith in consumers spending on their homes. Sherwin-Williams, Whirlpool, and Masco had some of the most disappointing reports, and were the only three Cramer said clued him in on any slowdown in home spending. Now, you can say, 'Well, isn't that everything? The paint aisle, the appliance aisle, and the cabinet aisle,' but ??those companies aren't necessarily indicative of how Home Depot did, Cramer said Tuesday. Instead, Cramer said Home Depot's leadership has been very, very constructive on the longer term, and it has urged its shareholders to have a longer-term outlook on demographic shifts and birth rates that could improve spending. 'There's a lot of money still pent up to spend on homes,' Cramer said in a nod to a theory voiced by Home Depot CFO Carol B. Tome. The election has caused a pause in consumer spending overall, and retailers like Macy's and J.C. Penney are eager to calm shareholder worries by blaming any dips on pre-election jitters, Cramer said. 'When you listen to the CEOs in retail, ??they actually think the election's a sea change and they definitely think that now [that] it's behind us, people are going to start going out again,' Cramer said. "

The following table displays the share prices and volume traded for Home Depot (ticker: HD) on the following 10 days:

**Table 2.** Share Prices and Volume traded of HD after the news is released.

| | +0 | +1 | +2 | +3 | +4 | +5 |
|---|---|---|---|---|---|---|
| **SHARE PRICE** | 124.4 | 125.33 | 128.93 | 128.33 | 128.22 | 130.98 |
| **VOLUME** | | 5067816 | 4016438 | 2743820 | 2525716 | 2341552 |
| | +6 | +7 | +8 | +9 | +10 | |
| **SHARE PRICE** | 131.21 | 131.57 | 130.64 | 129.62 | 129.40 | |
| **VOLUME** | 3411366 | 967160 | 2852506 | 1889058 | 3336388 | |

It clearly shows that the price jumps with volume support. We learn from these examples that news and subsequent volume traded are important determinants of the direction of a share price movements. It shed light on the possibility that a machine can learn from a pattern of words written in a news to predict the direction of a subsequent share price movement. However, the information contained in the news are unstructured. Unstructured information makes this a challenging machine learning problem. A news can be bullish, bearish, spam, rumor, or simply unrelated to a stock. We need actually to find ways to filter out important news from the noise.

In this paper, we propose to select those news related only to a single stock. With the help of a NLP tools, we suggest to convert the news into a 1 to 3 gram words news-terms matrix. We measure the performance of a stock by the volume weighted average prices of the following 10 days after the news is released. An adjustment to the classical Bayesian classifier is proposed to predict the direction of the movement of the stock prices. Experimental results have shown that news and volume are very useful to significantly improve the prediction accuracy.

## 2 Related Work

Many past studies used text to better predict market trends. [1] introduced tree structure in news, [2] used Twitter data, [3] identified expert investors, [4] studied the effect of news in event-driven trading, [5] proposed a method that predicts risk based on financial reports, [6] demonstrated that text can predict the long term asset prices better than other quantitative variables and [7] applied NLP to find that events reported in news are important to stock price movements. [8,9,10] used various lexical and syntactic contraints to extract event features for stock forecasting. Linear classifiers and deep neural networks were applied as predictive models.

## 3 Data

We collected over 2190 financial news scripts for 460 stocks from [12] between Jan 2016 and Feb 2018. At the end of each news script, [12] provides a list of tickers that may be potentially affected. We suggest to only use those news scripts with a single ticker. Since each of those news focus on a single stock, the content of the news are easier for the machine to understand and the impact to the share prices are more or less uni-directional. To prepare the news for the further processing, misspelled words are replaced by correctly spelled words.

## 4 Feature representation

After the news are prepared, an NLP processor converts every news scripts to lower case and removes numbers, stopwords, punctuation and spaces from those news. A stemmer is used to convert all words to their roots. Then a ngram tokenizer is invoked to convert each news to a vector of 1 to 3 gram words. The entries in the vector are the frequency of the corresponding words in the news. Combining all these vectors creates a news-terms matrix. The news-terms matrix is the basis for calculating empirical probabilities. The important of a term increases proportionally to the number of times the term appears in the news.

## 5 Stock Prices

For each collected news, we record the share prices and volumes of the related ticker at the date the news was collected and that of the following 10 days. Let $\{(p_1, v_1), \ldots, (p_{10}, v_{10})\}$ be the share prices and volumes of the following 10 days after the news is released. The volume weighted average of the share prices is defined as

$$\bar{p} = \frac{\sum_{i=1}^{10} v_i p_i}{\sum_{i=1}^{10} v_i}.$$

We define that a news is a positive news if $\bar{p} > 1.03p_0$, where $p_0$ is the share price at the date the news is released. Conversely, a news is negative if $\bar{p} < 0.97p_0$.

## 6 Theory of an adjusted Bayesian classifier

Let $W$ be the set of words which can be a single word, a 2 gram word or a 3 gram word. A row of the news-terms matrix is of the form $D = (n(w_1), n(w_2), \ldots), w_i \in W$, where $n(w)$ is the number of occurrence of $w$ in $D$. The rows are a representation of the news. Define $Y = 1$ if $D$ is a positive news and $Y = -1$ otherwise. Define $P(Y = 1) = \pi$ and $P(Y = -1) = 1 - \pi$, where $\pi$ is a prior distribution of $Y$. Here, the classical Bayesian classifier assumes that $(w_1, w_2, \ldots)$ is jointly selected according to $Y = 1$ or $Y = -1$ with probability

$$P(D = (n(w_1), n(w_2), \ldots)|Y = 1) \propto \prod_i p_i^{n(w_i)};$$

$$P(D = (n(w_1), n(w_2), \dots)|Y = -1) \propto \prod_i q_i^{n(w_i)},$$

where $p_i = P(w_i \in D|Y = 1)$ and $q_i = P(w_i \in D|Y = -1)$. Since $\{Y = 1\}$ and $\{Y = -1\}$ form a partition to the set of news, by Bayes theorem,

$$P\big(Y = 1\big|D = (n(w_1), n(w_2), \dots)\big) \propto \prod_i p_i^{n(w_i)}\,\pi;$$

$$P\big(Y = -1\big|D = (n(w_1), n(w_2), \dots)\big) \propto \prod_i q_i^{n(w_i)}\,(1-\pi).$$

Thus, the odds is defined by the ratio of the two probabilities, that is

$$\frac{P\big(Y = 1\big|D = (n(w_1), n(w_2), \dots)\big)}{P\big(Y = -1\big|D = (n(w_1), n(w_2), \dots)\big)} = \left(\frac{\pi}{1-\pi}\right)\prod_i \left(\frac{p_i}{q_i}\right)^{n(w_i)}.$$

A convenient classification rule is to classify $Y = 1$ if the odds is larger than one and $Y = -1$ otherwise. In other words, we can classify $Y = 1$ if $\beta_0 + \sum_{i\geq 1} n(w_i)\,\beta_i > 0$ and $Y = -1$ otherwise, where $\beta_0 = \log\left(\frac{\pi}{1-\pi}\right)$ and $\beta_i = \log\left(\frac{p_i}{q_i}\right)$.

Ignorance of the overlapping counts oversimplifies the selection process. We propose an adjustment to the above mentioned classical Bayesian classifier by the idea of Rosenblatt's perceptron ([11]). Let $M$ be the set of misclassification news. When $\beta_0 + \sum_{i\geq 1} n(w_i)\,\beta_i > 0$ , $y$ should be one. Otherwise, $y$ should be negative one. Hence, $y(\,\beta_0 + \sum_{i\geq 1} n(w_i)\,\beta_i)$ should always be positive. However, for a news in $M$, $y(\,\beta_0 + \sum_{i\geq 1} n(w_i)\,\beta_i) < 0$. The idea of Rosenblatt is to find $\beta_0$ and $\beta_i, i \geq 1$ such that the sum $L$ is minimized, where

$$L = -\sum_{k\in M} y_k \left(\beta_0 + \sum_{i\geq 1} n(w_i)_k\,\beta_i\right).$$

By a strict forward differentiation,

$$\frac{\partial L}{\partial \beta} = -\sum_{k\in M} y_k\,(n(w_1), n(w_2), \dots)_k;$$

$$\frac{\partial L}{\partial \beta_0} = -\sum_{k\in M} y_k.$$

A gradient descent method implies the following update iteration step:

$$\begin{bmatrix}\beta_0\\ \beta\end{bmatrix} = \begin{bmatrix}\beta_0\\ \beta\end{bmatrix} + \rho\begin{bmatrix} y_k \\ y_k(n(w_1), n(w_2), \dots)_k\end{bmatrix}, \qquad (*)$$

where $\rho$ is a learning rate and the beginning values of $\beta_0$ and $\beta$ are $\log\left(\frac{\pi}{1-\pi}\right)$ and $\left(\log\left(\frac{p_i}{q_i}\right), i \geq 1\right)$ respectively.

## 7 The implementation of the adjusted Bayesian classifier

The procedure begins with a segmentation of the set of collected news into a training set and a testing set. 70% of the news are used for training. We use the training set to train a classifier.

The prior probability of $Y$ is estimated by $\hat{\pi} = \frac{\#\{Y=1\}}{n}$, where $n$ is the total number of training news. In words, $\hat{\pi}$ is the proportion of positive news in the training set. By dividing the training set into the set of positive news and the set of not positive news. We can estimate

$$\hat{p}_i = \frac{the\ i^{th}\ column\ sum\ of\ the\ training\ positive\ news - terms\ matrix}{the\ number\ of\ training\ positive\ news};$$

$$\hat{q}_i = \frac{the\ i^{th}\ column\ sum\ of\ the\ training\ not\ positive\ news - terms\ matrix}{the\ number\ of\ training\ not\ positive\ news}.$$

Having produced the probability estimates, estimates of $\beta_0$ and $\beta$ are $\hat{\beta}_0 = \log(\hat{\pi}/(1-\hat{\pi}))$ and $\hat{\beta}_i = \log(\hat{p}_i/\hat{q}_i)$ respectively. We then classify the news in the training set as positive when $\hat{\beta}_0 + \sum_{i\geq 1} n(w_i)\,\hat{\beta}_i > 0$ and not positive otherwise. For each of the misclassification news, we adjust the $\hat{\beta}_0$ and $\hat{\beta}_i$ using the Rosenblatt's updated iteration (*). After adjustment of the $\beta_0$ and $\beta$, we reclassify the news in the training set. This procedure continues until there is no misclassification news or there cannot be any further improvement on the correct classification rate.

The resulting estimates of the $\beta_0$ and $\beta$ are employed to classify news in the testing set. The results of the classification are then compared with their true classes. It is found that the correct classification rate of positive news is 83.5%. Similarly, the same procedure is applied to classify negative news. We find that the correct classification rate of negative news on a testing set is 84.15%.

## 8 Conclusion

In this paper, we have proposed to select news related to a single ticker and a method with an adjusted Bayesian classifier to predict the direction of related share price movement in short-term. The performance of the stocks are determined by the volume weighted prices of the following 10 days after the news is released. Our experiment has shown that the method is effective in predicting the movement directions on a testing set of news.

## References

1. Boyi Xie, Rebecca Passonneau, Leon Wu, and Germán G. Creamer. 2013. Semantic frames to predict stock price movement. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL), pages 873--883. Association for Computational Linguistics.Author, F., Author, S.: Title of a proceedings paper. In: Editor, F., Editor, S. (eds.) CONFERENCE 2016, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2016).
2. Johan Bollen, Huina Mao, and Xiao-Jun Zeng. 2010. Twitter mood predicts the stock market. J. Comput. Sci. 2 (1) (2011) 1-8.Author, F.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010).
3. Roy Bar-Haim, Elad Dinur, Ronen Feldman, Moshe Fresko, and Guy Goldstein. 2011. Identifying and Following Expert Investors in Stock Microblogs. In Proceedings of EMNLP-2011.
4. David Leinweber and Jacob Sisk. 2011. Event-driven trading and the new news. The Journal of Portfolio Management, 38(1):110--124.
5. Shimon Kogan, Dimitry Levin, Bryan R. Routledge, Jacob S. Sagi, and Noah A. Smith. 2009. Predicting risk from financial reports with regression. In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, NAACL '09, pages 272--280, Stroudsburg, PA, USA. Association for Computational Linguistics.
6. Joseph Engelberg. 2008. Costly information processing: Evidence from earnings announcements. San Francisco Meetings Paper.
7. W. Y. Wang, Z. Hua, A semiparametric gaussian copula regression model for predicting nancial risks from earnings calls, in: The 51st Annual Meeting of the Association for Computational Linguistics (ACL-14), 2014, pp. 1155 - 1165.
8. X. Ding, Y. Zhang, T. Liu, J. Duan, Using structured events to predict stock price movement: An empirical investigation, in: The Conference on Empirical Methods on Natural Language Processing (EMNLP-14), 2014, pp. 1415-1425.
9. X. Ding, Y. Zhang, T. Liu, J. Duan, Deep learning for event-driven stock prediction, in: Proceedings of the 24th International Joint Conference on Artificial Intelligence (ICJAI-15), 2015, pp. 2327-2333.
10. X. Ding, Y. Zhang, T. Liu, J. Duan, Knowledge-driven event embedding for stock prediction, in: Proceedings of the 26th International Conference on Computational Linguistics (COLING-16), 2016, pp. 2133-2142.
11. Rosenblatt, Frank. "The perceptron: a probabilistic model for information storage and organization in the brain." Psychological review 65.6 (1958): 386.
12. URL: www.cnbc.com