

# Sentiment Analysis of Investor Opinions on Twitter

**Brian Dickinson, Wei Hu**

Department of Computer Science, Houghton College, Houghton, New York, USA  
Email: [wei.hu@houghton.edu](mailto:wei.hu@houghton.edu)

Received 4 May 2015; accepted 6 July 2015; published 9 July 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

The rapid growth of social networks has produced an unprecedented amount of user-generated data, which provides an excellent opportunity for text mining. Sentiment analysis, an important part of text mining, attempts to learn about the authors' opinion on a text through its content and structure. Such information is particularly valuable for determining the overall opinion of a large number of people. Examples of the usefulness of this are predicting box office sales or stock prices. One of the most accessible sources of user-generated data is Twitter, which makes the majority of its user data freely available through its data access API. In this study we seek to predict a sentiment value for stock related tweets on Twitter, and demonstrate a correlation between this sentiment and the movement of a company's stock price in a real time streaming environment. Both n-gram and "word2vec" textual representation techniques are used alongside a random forest classification algorithm to predict the sentiment of tweets. These values are then evaluated for correlation between stock prices and Twitter sentiment for that each company. There are significant correlations between price and sentiment for several individual companies. Some companies such as Microsoft and Walmart show strong positive correlation, while others such as Goldman Sachs and Cisco Systems show strong negative correlation. This suggests that consumer facing companies are affected differently than other companies. Overall this appears to be a promising field for future research.

## Keywords

Sentiment Analysis, Word2vec, Text Mining, Twitter, Stock Prediction

---

## 1. Introduction

Over the last several years there has been an explosion of growth and new activity in social networking. Various

companies such as Facebook, LinkedIn, Reddit, Pinterest, and Twitter have grown exponentially in recent years. The amount of data exchanged between users on these sites is staggering. On Facebook alone on an average day in 2014 there are 4.75 billion items being shared, 4.5 billion items “liked”, and 300 million photographs being uploaded. That translates to over 500 terabytes of data generated by Facebook users on a single day [1]. There is an incredible amount of useful information about individual opinions, feelings, and relationships contained in these transactions, but the loosely structured nature of human communication makes harnessing this data a challenge. In order to make sense of the large portion of this data which is text-based, Natural Language Processing tools can be used to rigorously categorize user generated text. One of these tools for determining useful information from massive data sources such as Twitter is a sentiment analysis.

Sentiment analysis focuses on determining the opinion of a speaker on the particular topic about which he is speaking. The most basic structure for sentiment analysis is a single word, unfortunately based on sentence structure and words with context dependent meanings, techniques that ignore sentence structure or bag of words models often fail on smaller texts. A solution to this is constructing parse trees which identify the structure of a sentence as a binary tree by separating distinct phrases. In this case using the sentiment of each word in the tree can take into account clause structures and the possibility of multiple meanings. In cases where a larger text must be analyzed, it can be treated as a collection of smaller phrases, or as a larger bag of words. Opinions are usually classified somewhere between positive or negative often with some stratification between the two. This can be done numerically or categorically. When division is categorical, it usually distinguishes between positive, negative, and sometimes neutral sentiments, otherwise the numerical classification falls somewhere on a continuum between positive and negative. These classifications can be used to determine and aggregate the sentiment of a large number of authors on a given topic.

Because of sarcasm, and even simple negations can completely reverse the predicted sentiment from the actual opinion represented, parse trees are the most accurate method of determining the sentiment of sentences. One of the most recent methods of sentiment analysis, published by Stanford [2], has advanced features which allow it to recognize some of the most difficult features of human languages. It can identify subordinating conjunctions such as “if”, “after”, “once”, “since”, and “because” and weight towards the portion of the sentence after these since it is more important to the meaning of the sentence. Similarly it can identify the purposes of certain modifiers such as “very” and “incredibly” and use them to weight the sentiment of the words they modify rather than giving them a sentiment of their own. Finally it can handle negations quite well, and is usually accurate in switching the sentiment when something is negated [3]. All of these features make this new system for sentiment analysis very promising as a foundation for applications based on sentiment.

This Stanford model makes use of deep neural networks. Deep neural networks are an expansion on early neural networks such as Perceptrons. Advances in hardware processing speeds, particularly graphics processing units, as well as an increasing interest in parallel processing have brought resurgence in the use of artificial neural networks by enabling the addition of hidden layers of neurons, and backpropagation. The additional layers allow these models to become more highly non-linear fitting closer to the data, while backpropagation enhances training efficiency on labeled data in deep neural networks. Such networks have won numerous competitions in pattern and image recognition competitions over the last five years, and appear to have great promise in improving classification accuracies in most areas of machine learning.

One topic of user sentiment that can be easily checked for correlations between public opinion and public behavior is that of stock price prediction. There are two basic methods for predicting whether the price of a given stock will rise or fall, fundamental, and technical analysis. Fundamental analysis relies on the financial data of the company to make assessments of financial stability, growth potential, and inherent value. This value can be matched against the current market price. If the estimated real value is higher than the current market price, it is believed that the company is undervalued and that the stock is more likely to rise than to fall. Similarly if the estimated value is lower than the market price, it is assumed that the stock is overvalued and that it is more likely to fall than to rise [4].

Technical analysis takes a different route. Rather than focusing on financial data, technical analysis uses historical price data to make predictions about the expected direction of price change in the future. Frequently observed patterns that appear to be occurring such as head and shoulders or double tops, as well as recent trends such as channels and uptrends are used to predict future prices. Another way technical analysis seeks to predict prices is through observing the behavior of others. One factor in technical analysis is buying and selling at the same time as company insiders, and buying and selling opposite odd-lot traders. The idea behind this is that company

insiders are best acquainted with the company's prospects, and make the most educated buying and selling decisions. Meanwhile odd-lot traders are almost always individual rather than professional traders and generally lack a strong investing background. Their trades are negatively correlated with trades by company insiders, and generally result in buying when prices are high, and selling when prices are low. By buying and selling opposite them, one can often buy when prices are low and sell when prices are high. [5] None of these methods guarantee success, but they are generally accepted as the most likely to produce results better than random guessing.

Text analysis and more specifically sentiment detection could provide an insight for investor and general public opinion on a company and its stock price on a large scale. This insight could provide more information for use in analysis techniques similar to those currently supported by technical analysis. This could be a promising method for determining the relationship between human evaluations and stock price apart from the apparent underlying values of companies uncovered by fundamental analysis.

## 2. Related Work

There has been a significant amount of research into text analysis, including sentiment analysis, as well as some interest in utilizing these tools for prediction through Twitter, however up until now these projects have primarily worked with text analysis and sentiment prediction more generally. This is one of the unique difficulties of the problem of detecting investor sentiment on Twitter. Since tweets expressing clear sentiment about a stock can look either objective or simply noisy to general models. For example one collected tweet reads “\$MSFT bullish...” which has little natural meaning, however in the context of the jargon particular to securities markets, this tweet expresses clear positive sentiment towards Microsoft Corporation. Such difficulties necessitate the construction of a sentiment classifier particular to this field of study. General models such as the Stanford NLP sentiment classifier discussed in the introduction, however, can still be immensely valuable in providing a basic framework for a context specific classifier.

### 2.1. Textual Representation

There are three primary means of representing text in statistical textual analysis. These are n-gram, vector space modeling, and character streams. The first of these techniques, n-gram representations, has been around for decades and provides the simplest most straightforward method of representing text based on simple word or character sequence counts. Vector space models are a more complex and far more recent development with the most popular implementation, “word2vec”, having been created within the last few years. Finally character stream techniques are the most recent development in the field with the first viable model having been published mere months before this writing. As such this final technique, though an extremely promising avenue for progress in the field, has been omitted from this research. Its likely impact however is significant enough to justify inclusion in any overview of textual representation techniques.

N-gram representations are based on simple character or word sequence counts. In these techniques a full corpus of related text is parsed, and every appearing character or word sequence of length  $n$  is extracted to form a dictionary of words and phrases. For example the text “the quick brown fox jumps over the lazy dog” has the following 5-gram word features: “the quick brown fox jumps”, “quick brown fox jumps over”, “brown fox jumps over the”, “fox jumps over the lazy”, and “jumps over the lazy dog”. Similarly the text “g2g ttyl” has the following 5-gram character features: “g2g t”, “2g tt”, “g tty”, and “ttyl”. Every text in the corpus can then be easily marked as a vector based on a simple count of the number of times each phrase in the dictionary occurs in the text. The main advantages of this technique are its simplicity, and flexibility to specifically match the corpus of text being studied [6].

Vector space techniques require a substantial set of text cleaned so as to include only the words of the language. The spatial relationship between the words is then analyzed as described by Mikolov *et al.* [7]. Once every word of the language has been mapped to a unique vector, collections of words can be aggregated using entry wise summation and normalization yielding a vector for any given collection of words [8]. The cosine similarity between the vectors of words in such a representation demonstrates a significant carryover in word meaning [9]. For example the vector for “queen” can be calculated from the vectors for “king”, “man”, and “woman” as follows: “king” – “man” + “woman” = “queen”. This sustained relationship between word concepts makes vector space models very attractive for textual analysis. The primary negative consequence of vector space representations is their need to form a global representation across an entire language which may cause

difficulty in interpreting the language particular to a specialized field where words might be used with slightly altered meanings and relationships. Classification in this specialized environment can require a large set of related text from which to build the model.

Character stream modeling requires only a dictionary of valid characters to learn from a text. This enables it to be language agnostic to the point that the same algorithm can work effectively on languages as diverse as English and Chinese when provided adequate training data. This is a huge advancement over previous methods which often maintain rigid requirements for the language they are designed to model. So far this technique demonstrates superior predictive power to other methods albeit at the cost of increased computational complexity [10].

## 2.2. Sentiment Analysis on Twitter

In studying sentiment analysis on noisy and biased data, it was found that a multilevel classification model can provide more robust and accurate predictions in difficult data sets [11]. In this model tweets are first classified based on objectivity versus subjectivity. This can remove a lot of the noisy advertising data that simply states facts and separately uncovers tweets that are more likely to express true sentiment. These can then be classified with greater accuracy since the vast majority of tweets lacking sentiment that would have caused difficulties have been eliminated and effectively classified as neutral. This could be particularly useful in evaluating tweets about stocks since both subjective and objective tweets can provide information about the market, but would require separate processing methods.

There have also been several deep learning approaches to sentiment classification on twitter that have been specialized to account for the relatively limited data available in a text with a maximum of one-hundred forty characters. These studies have found that a combination of specially chosen metadata and textual features, along with more traditional analyses such as n-grams can provide a more accurate classification model than simple features alone [11] [12]. In particular targeting the appearance of words in capital letters, emoticons, and elongated words such as “whaaaaaat” or “coooooool” can significantly improve classification. Any approach to sentiment analysis utilizing deep neural networks deserves consideration considering the outstanding performance such methods have demonstrated in recent years.

## 2.3. Predicting the Behavior of a Population Based on Sentiment

There is some precedent for using aggregated sentiment analysis from Twitter data to make predictions about the behavior of a population. Previous research has demonstrated a significant connection between the overall sentiment of Twitter users towards new movie releases and the box office sales figures [13] [14]. Bollen *et al.* found that the overall mood of an unfiltered Twitter stream can predict movements in the Dow Jones Industrial Average as a whole [15]. Similarly the movement of markets of highly related companies has been predicted using an aggregate of the sentiment of companies in that market [16]. It is then reasonable to believe that the sentiment related to a particular company within the Dow Jones Industrial Average might be indicative of its future short term price movements.

## 3. Data and Methods

In order to perform a comparison between stock price changes and Twitter sentiment it was necessary to collect data on both trading values and tweets related to companies with timestamps to properly match the two into a consistent stream. In order to narrow down the range of data needing collection this study focused on the thirty companies of the Dow Jones Industrial Average. Data was collected from both sources over a period of several months from November 2014 through March 2015 utilizing specialized APIs. Only the data collected between February 6th, and February 18th was fully evaluated due to computational limitations and gaps in tweet data caused by throttling and inconsistent network connectivity.

### 3.1. Twitter and Stock Price Data Collection

Twitter provides its own API through which developers may obtain limited streams of live tweets. This interface was used through the Twitter4j java library to filter all available live tweets for any containing any complete company name, or ticker symbol on the Dow Jones Industrial Average. All tweets matching the filter were saved

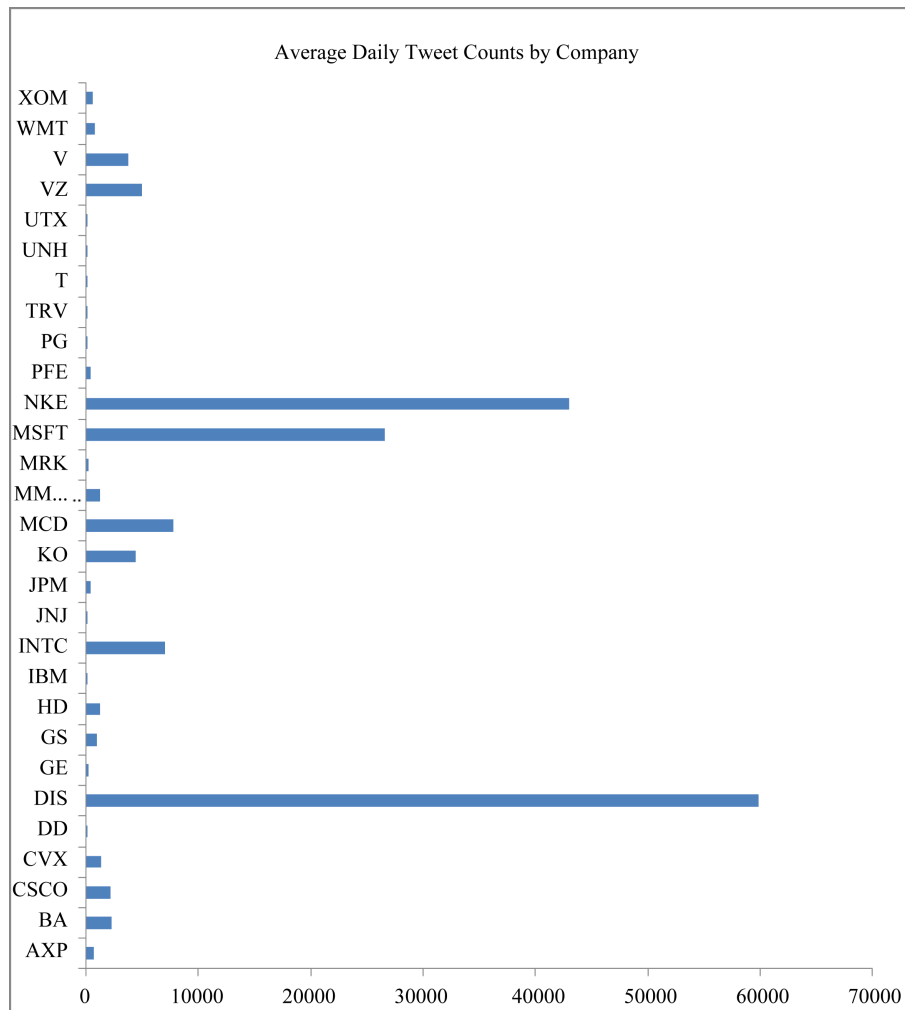
along with all available metadata including timestamp, sender, geotagging, retweet status etc. The number of tweets per day, and the distribution of tweets between companies are shown in **Figure 1** and **Figure 2**. Clearly some companies are far more commonly mentioned on Twitter. Similarly the Yahoo! Finance stock API was used to collect stock price data on the stocks of the Dow Jones Industrial Average. Every five minutes the API was queried for bid, ask, and last trade prices. This information was recorded for each stock along with the timestamp identifying precisely when the information was collected.

Because all of this information was collected in a real time streaming environment with very brief time windows, and no modifications to the data, this approach lends itself well to the type of moment by moment analysis that must be conducted for technical stock analysis.

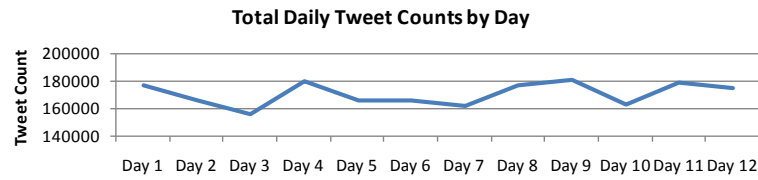
A sampling of these tweets was selected and manually labeled with sentiment values. According to this sample the distribution of sentiments can be found in **Figure 3**. It is clear from this that neutral was by far the most common sentiment expressed in my dataset, and that positive sentiment was nearly three times as common as negative sentiment. It is important to note that this is an unaltered stream of tweets obtained directly through the streaming API, and is thus a natural distribution of sentiment. This is important to contrast with many datasets which enforce an even split in their labeled and unlabeled data.

### 3.2. Sentiment Analysis of Collected Tweets

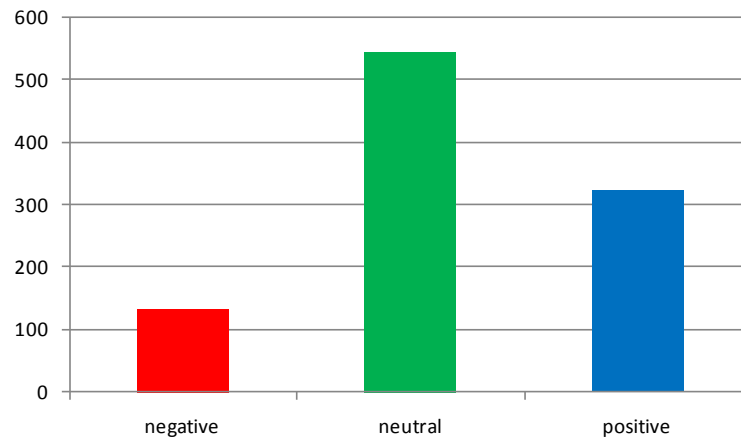
After collecting raw text data from it was also necessary to compute the sentiment value for each tweet since it



**Figure 1.** Average tweets collected for each company per day (from February 12, 2015 to February 18, 2015).



**Figure 2.** Counts of tweets received each day (from February 12, 2015 to February 18, 2015).



**Figure 3.** Tweet count distribution in the sample by sentiment.

is this sentiment which may relate closely to the stock price. Initially the Stanford NLP Sentiment Classifier was used to predict the sentiment of each tweet. In order to evaluate the accuracy of these predictions it was necessary to prepare a set of tweets labeled with true sentiment values. This was done manually to ensure accuracy, and as such the labeled set consisted of only one thousand tweets. Due to the nature of the parser and its primary training on movie reviews and newspaper articles, it was particularly inadequate for the task performing with approximately 30% accuracy.

In response to this we constructed my own classification models, one using n-gram, and the other “word2vec” textual representation techniques to preprocess raw text before using a standard random forest model for classification. Each of these models performed with accuracy between 60% and 70% on the labeled data set, an acceptably high level of accuracy for textual sentiment analysis on such short texts.

### 3.3. Correlation Analysis on Stock Price and Tweet Sentiment

Since there are only two variables involved for each company, namely sentiment and price, Pearson’s Correlation Coefficient can readily demonstrate a connection between the two. In order to calculate the correlation between these values, sentiment values over five minute increments had to be aggregated. This allows a pairing of the sentiment over a five minute period with the value of the stock after five minutes. Unfortunately these numbers are of very different kinds. The sentiment value takes into account only the previous five minutes, while the stock value at specific moment takes into account everything that has occurred before hand. In order to create a more proper comparison two techniques were used. The first changes price values to account only for the last five minutes by using the price change since the previous measurement, while the second uses a running total of sentiment to help sentiment values aggregate beyond their five minute periods. Each of these techniques still leaves values in very different ranges; to correct this both sequences were normalized to fall between zero and one. Once this normalization was complete it was possible to calculate correlations between these series for each company. Unfortunately there remained a significant amount of noise because of the limited data available for any given five minute interval, and the variability of readings. To correct this, a moving average was used with a window length of one day, and a step size of one hour. This significantly smoothed both curves and made results far more readable. This made it possible to plot these variables over the recorded time range to visually evaluate the relationship between price and sentiment for each company. A sampling of these calculations and charts may



be found in the next section.

## 4. Results and Conclusions

### 4.1. Sentiment Classification Results

This section describes the results obtained through the methodologies described in the preceding section. The foundational problem to this study was the sentiment classification which was utilized by all subsequent testing methods. The confusion matrices of the classification using n-gram and vector space representations are shown in **Table 1**, **Table 2**. Values on the top left to bottom right diagonal indicated correct predictions, while all off diagonal values indicate what errors the classifier is making in prediction. The sum of the diagonal values divided by the sum of all values in the table gives an accuracy of 68.5% for n-gram, and 63.4% for word2vec. This indicates that in this particular data set, the value of a domain specific vocabulary training outweighs the benefit of more accurate representation of the relationship between words for sentiment classification. While not outstanding for sentiment prediction in general, these figures are passable for unfiltered data in such a domain specific context.

These matrices show that using the n-gram representation predictions on positive, neutral, and negative tweets had accuracies of 55.4%, 84.6%, and 34% respectively. Similarly the same accuracies using the word2vec representation were 42.4%, 88.4%, and 12.1%. Preference for neutral sentiment is due to the overall probability of a given tweet having neutral sentiment. Since most tweets are neutral the classification model errs towards neutral predictions. This leads to a prediction biased towards positive since positive prediction is higher than negative prediction, however given sufficient training data this would not be an issue as this bias is caused by the occurrence of more positive tweets than negative ones in the sample data. The predictions remain in proportion with the true labels of the training data.

### 4.2. Price and Sentiment Correlation Analysis Results

Results of the correlation analysis were very mixed across companies, as may be seen in **Table 3**. Some companies such as Cisco Systems (CSCO) and Goldman Sachs (GS) showed a strong negative correlation of -0.90 between sentiment and price, while others such as Walmart (WMT) and Microsoft (MSFT) showed strong positive correlations of 0.85. Of these companies in particular, Microsoft and Walmart are largely consumer facing while Cisco Systems and Goldman Sachs are not. This might indicate different effects of sentiment on different sorts of companies. These results are clearly visible in the charts for the selected companies shown in **Figures 4-7**. From these it appears that in some cases there very well may be a connection between sentiment and price, but that connection may not always be the same. The first two figures show strong negative correlation, while the latter two show strong positive correlation. In particular it might be beneficial to note the Microsoft chart (**Figure 7**). In this case the sentiment remained correlated with price though the direction of both lines changed.

**Table 1.** n-gram sentiment prediction.

		Predicted Label		
		positive	neutral	negative
True Label	positive	179	126	18
	neutral	70	461	14
	negative	17	70	45

**Table 2.** word2vec sentiment prediction.

		Predicted Label		
		positive	neutral	negative
True Label	positive	137	173	13
	neutral	47	479	16
	negative	18	98	16

**Table 3.** Pearson correlation coefficients for every company for each text representation method.

Company (symbol)	word2vec representation correlation	n-gram representation correlation	Company (symbol)	word2vec representation correlation	n-gram representation correlation
WMT	0.848094	0.847447	XOM	-0.41886	-0.45982
MSFT	0.856233	0.844525	MRK	-0.4751	-0.51843
UTX	0.841241	0.839601	KO	-0.53817	-0.52242
UNH	0.796927	0.815517	MCD	-0.52038	-0.52473
GE	0.787573	0.796093	INTC	-0.51058	-0.62937
IBM	0.902387	0.745618	T	-0.69758	-0.68944
DD	0.725684	0.698454	MMM	-0.73474	-0.71081
AXP	0.58826	0.668651	NKE	-0.73293	-0.73339
PFE	0.796572	0.598452	BA	-0.73247	-0.75375
CVX	0.40015	0.424398	JPM	-0.79279	-0.75673
TRV	0.489148	0.38869	HD	-0.78832	-0.78254
VZ	0.376637	0.380748	DIS	-0.86191	-0.81751
JNJ	0.334807	0.286778	CSCO	-0.91567	-0.89665
PG	-0.28735	-0.33896	GS	-0.90957	-0.90044
V	-0.39149	-0.37326			

**Figure 4.** Cisco systems price-sentiment correlation (negative correlation).

This is important as high correlations are not necessarily significant where both lines are simply moving in the same direction throughout.

### 4.3. Conclusion

In a study of correlation between Twitter sentiment and stock price, it is expected to have three possible outcomes: positive, negative and neutral correlation. The main contribution of our work is the identification of those companies that are in each of the three categories during the time period of our investigation. The correlation has been shown to be strongly positive in several companies, particularly Walmart and Microsoft which are primarily consumer facing corporations. There is of course not a uniform connection between sentiment and price across all companies. Based on promising results in sentiment to price correlations on company groups





Figure 5. Goldman Sachs price-sentiment correlation (negative correlation).



Figure 6. Walmart price-sentiment correlation (positive correlation).

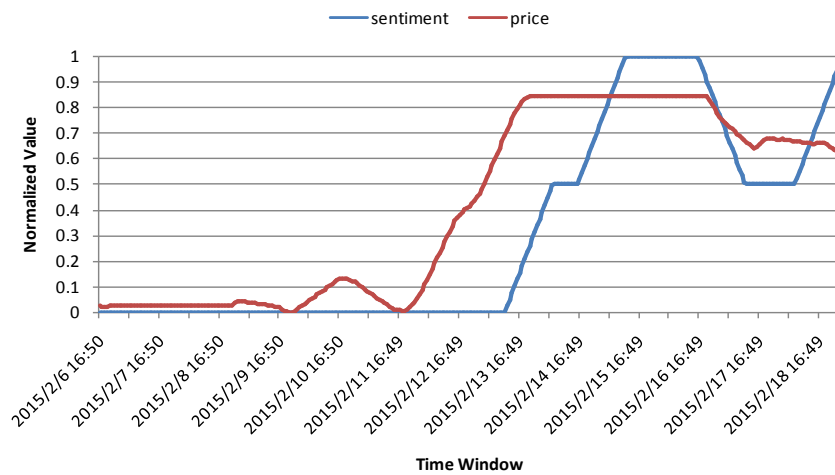


Figure 7. Microsoft price-sentiment Correlation (positive correlation).

from previous studies and on the strong correlation between sentiment and price for certain companies in this study, we believe that further research on the correlation between sentiment and stock prices is warranted.

## Acknowledgements

We would like to thank Houghton College for its financial support.

## References

- [1] Zephoria Incorporation (2015) Top 20 Facebook Statistics. Zephoria Incorporation, Sarasota.
- [2] Yang, B.S. and Cardie, C. (2014) Context-Aware Learning for Sentence-Level Sentiment Analysis with Posterior Regularization. *Proceedings of ACL*, Baltimore Maryland, June 2014, 325-335.
- [3] Richard, S., Perelygin, A., Wu, J., Chuang, J., Manning, C., Ng, A. and Potts, C. (2013) Recursive Deep Models for Semantic Compositionality over a Sentiment Treebank. *Conference on Empirical Methods on Natural Language Processing (EMNLP)*, Seattle Washington, October 2013, 1631-1642.
- [4] Lev, B. and Thiagarajan, S.R. (1993) Fundamental Information Analysis. *Journal of Accounting Research*, **31**, 190-215. <http://dx.doi.org/10.2307/2491270>
- [5] Wong, W.-K., Manzur, M. and Chew, B.-K. (2003) How Rewarding Is Technical Analysis? Evidence from Singapore Stock Market. *Applied Financial Economics*, **13**, 543-551. <http://dx.doi.org/10.1080/0960310022000020906>
- [6] William, C. and Trenkle, J. (1994) N-Gram-Based Text Categorization. *Proceedings of Annual Symposium on Document Analysis and Information Retrieval*, Las Vegas Nevada, April 1994, 161-175.
- [7] Tomas, M., Chen, K., Corrado, G. and Dean, J. (2013) Efficient Estimation of Word Representations in Vector Space. *Computation and Language*, arXiv preprint arXiv: 1301.3781.
- [8] Tomas, M., Sutskever, I., Chen, K., Corrado, G. and Dean, J. (2013) Distributed Representations of Words and Phrases and Their Compositionality. *Proceedings of Neural Information Processing Systems*, Lake Tahoe, December 2013, 3111-3119.
- [9] Tomas, M. (2013) Linguistic Regularities in Continuous Space Word Representations. *Proceedings of North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Westin Peachtree Plaza Hotel, 9-14 June 2013, 746-751.
- [10] Zhang, X. and LeCun, Y. (2015) Text Understanding from Scratch. *Computation and Language*, arXiv preprint arXiv: 1502.01710.
- [11] Luciano, B. and Feng, J.L. (2010) Robust Sentiment Detection on Twitter from Biased and Noisy Data. *Proceedings of the 23rd International Conference on Computational Linguistics*, Beijing, 23-27 August 2010, 36-44.
- [12] Tang, D.Y., Wei, F.R., Qin, B., Liu, T. and Zhou, M. (2014) Coooolll: A Deep Learning System for Twitter Sentiment Classification. *Proceedings of the 8th International Workshop on Semantic Evaluation*, Dublin, 23-24 August 2014, 208-212.
- [13] Rui, H.X., Liu, Y.Z. and Whinston, A. (2013) Whose and What Chatter Matters? The Effect of Tweets on Movie Sales. *Decision Support Systems*, **56**, 863-870. <http://dx.doi.org/10.1016/j.dss.2012.12.022>
- [14] Gregoire, M., Mikolov, T., Ranzato, M. and Bengio, Y. (2014) Ensemble of Generative and Discriminative Techniques for Sentiment Analysis of Movie Reviews. *Computation and Language*, arXiv preprint arXiv: 1412.5335.
- [15] Johan, B., Mao, H.N. and Zeng, X.J. (2011) Twitter Mood Predicts the Stock Market. *Journal of Computational Science*, **2**, 1-8. <http://dx.doi.org/10.1016/j.jocs.2010.12.007>
- [16] Si, J.F., Mukherjee, A., Liu, B., Pan, S., Li, Q. and Li, H.Y. (2014) Exploiting Social Relations and Sentiment for Stock Prediction. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Doha, 25-29 October 2014, 1139-1145.