2018 12th IEEE International Conference on Semantic Computing

Financial Sentiment Lexicon Analysis

Sahar Sohangir
Department of Computer and Electrical Engineering and Computer Sciece
Boca Raton, Florida, USA
Email: ssohangir2014@fau.edu

Nicholas Petty
Department of Computer and Electrical Engineering and Computer Sciece
Boca Raton, Florida, USA
Email: npetty2014@fau.edu

Dingding Wang
Department of Computer and Electrical Engineering and Computer Sciece
Boca Raton, Florida, USA
Email: wangd@fau.edu

***Abstract*—The modern stock market is a popular place to increase wealth and generate income, but the fundamental problem of when to buy or sell shares, or which stocks to buy has not been solved. With the availability of the Internet and its financial social networks, such as StockTwits and SeekingAlpha, investors around the world have new opportunities to gather and share their experiences. Individual experts can predict the movement of the stock market in financial social networks with reasonable accuracy, but how accurate is a large group of such experts in aggregate? One way to answer this question is by examining the sentiment of a massive group of these authors towards various stocks. By extracting the sentiment of the whole group, a collective prediction can be observed. Although sentiment extraction is a major technical challenge, the lexicon-based approach is an effective method of determining how positive or negative the content of a text document is. In this paper, we investigate if we can improve the performance of sentiment extraction from financial social media data by using lexicon-based approaches.**

***Keywords*-Sentiment analysis; opinion retrieval; natural language processing; sentiment lexicon**

I. Introduction

The Internet has become a tool of open communication for billions of people around the world, allowing interaction between individuals who may have never been able to connect previously. Crowdsourcing uses the collective wisdom of a large group of people to achieve a specific goal and has brought about a social revolution.

One website which brings these opportunities to its users is StockTwits. By leveraging Twitter's 140 character tweet system, StockTwits aggregates market analyses from the Twitter social media platform and condenses them into a focused, curated stream of data. If this stream were examined in full, it would be possible to determine the crowd's collective sentiment towards the market and make predictions from it. What makes StockTwits special is its users' ability to add a tag to their tweets to indicate whether their post is "Bullish" and they think the stock or market will improve, or "Bearish" and they think the stock or market will get worse.

In this paper, we will examine a labeled dataset from StockTwits and determine whether lexicon based sentiment analysis methods are effective for classification.

We will begin by reviewing a selection of works related to the application of machine learning and sentiment analysis on financial social media data. The next section covers our methodology. We will discuss the significance of our dataset, then compare sentiment analysis approaches through machine learning and sentiment lexicons. The following section will provide our experimental results, which show that lexicon-based approaches can offer improved performance over machine learning methods. In the last section, we summarize our conclusions and recommend the VADER system of lexicon-based sentiment analysis for classification of StockTwits tweets.

II. Related Work

Early work on Twitter and sentiment analysis comes from Bollen, et al. in [1], with their use of OpinionFinder and Google Profile of Mood States (GPOMS). These tools took tweet input and produced the author's sentiment, which was then compared against the performance of a stock market index. The authors showed that sentiment analysis of a large Twitter dataset regarding stock movement is possible. Additionally, they found that this analysis can be used for market predictions, with an accuracy of around 87%.

By expanding on the work Bollen, Mittal, and Goel in [2] looked further into sentiment analysis when applied to Twitter data. They realized that having a good sentiment analysis system was extremely important for their task, and evaluated multiple analyzers, including OpinionFinder and SentiWordNet. By stressing the importance of sentiment analysis on financial tweets, this work also leads us to examine the topic more closely. One of the most popular works in this field is by Loughran and McDonald [3]. They used the U.S. Security and Exchange Commission portal from 1994 to 2008 to make a financial lexicon and manually create six-word lists including *positive*, *negative*, *litigious*, *uncertainty*, *model strong* and *model weak*.

Supervised classification methods, such as Support Vector Machines, Naïve Bayes or ensembles [4], [5] have been deployed to perform sentiment analysis in multiple research projects. Machine learning techniques mainly use the bag-of-words [6] model. In the bag-of-words model, a text is represented as the collection of its words, disregarding the order of those words in their sentences. In addition, we do need feature engineering in machine learning methods.

Wang, et al. in [7] applied machine learning approaches, including Support Vector Machine, Naive Bayes, and Decision Tree, to classify StockTwits tweets as "bullish" or "bearish." They found that the SVM model was the most accurate at

0-7695-6360-0/18/$31.00 ©2018 IEEE
DOI 10.1109/ICSC.2018.00052

76.2%. Our research builds on this work by re-evaluating various machine learning models and then investigating lexicon-based sentiment analyzers to see if better accuracy can be attained. With an improved method of determining the overall feelings of StockTwits users, more accurate predictions can be made from their aggregate data.

III. Methodology

A. DataSet

StockTwits is a financial social network which was established in 2009. Information about the stock market, like the latest stock prices, price movement, stock exchange history, buying or selling recommendations, and so on, are available to StockTwits users. In addition, as a social network, it provides the opportunity for sharing experience among traders in the stock market. Through the StockTwits website, investors, analysts, and others interested in the market can contribute a tweet - a short message limited to 140 characters about the stock market. This message will be posted to a public stream visible to all site visitors. Moreover, messages can be labeled Bullish or Bearish by the authors to specify their sentiment towards their chosen stocks.

In our experiment, we used messages which were posted in the whole year of 2015 and the first six months of 2016. Each message includes a messageID, a userID, the author's number of followers, a timestamp, the current price of the stock, and other record-keeping attributes.

If the sentiment of top authors is known, we can predict stock prices with an accuracy of 75%. Unfortunately, only 10% of messages in StockTwits are labeled, so we can't rely on self-reported sentiment only. To increase the accuracy of stock price prediction, we need a powerful method to determine the sentiment of top authors. Therefore, sentiment lexicon is adopted to do sentiment analysis on StockTwits messages. We believe that using sentiment lexicon can vastly improve correct classification in sentiment analysis regarding various stock picks and thus exceed the current accuracy of stock price prediction.

B. Sentiment Analysis

Following the early work in sentiment analysis done in [8], [9], we examine source materials and apply natural language processing techniques to determine the attitude of the writer towards a subject. Generally speaking, the main goal in sentiment analysis is determining the attitude of a writer with respect to some topic or the overall contextual polarity to a document.

There are different methods in sentiment analysis that can help us to measure sentiments, including lexical-based approaches and supervised machine learning.

Machine learning require training data, which may also be difficult to acquire. In addition, the training process is time-consuming and computationally expensive in terms of CPU and memory requirements. Moreover, machine learning only depends on the training set to find features, and this selection may be incomplete.

With the growing popularity of social media, huge datasets of reviews, blogs, and social network feeds are being generated continuously. Concepts and methods from sentiment analysis that can help us to extract information from these areas have become increasingly important as businesses, organizations, and individuals seek to make better use of their data.

In the following section, we investigate the performance of sentiment lexicon to extract sentiment of users in financial social media.

C. Sentiment Lexicon

A sentiment lexicon is a list of lexical features which are generally labeled according to their semantic orientation as either positive or negative [10]. Due to the challenge of creating a lexicon, most research in sentiment analysis relies heavily on preexisting manually constructed lexicons. The three most common lexicons in use are LIWC [1], GI [2], and Hu-Liu04 [3]. In the following section, we briefly provide an overview of two most commonly-used sentiment lexicons - VADER and SentiWordNet. VADER uses a combination of qualitative and quantitative methods, and SentiWordNet is an extension of WordNet [11].

1) VADER: Valence Aware Dictionary for sEntiment Reasoning: VADER, as a parsimonious rule-based model for sentiment analysis, can be used in multiple domains. It is constructed from a generalized, valence-based, human-curated gold standard sentiment lexicon. In addition, the impact of grammatical and syntactical rules including punctuation, capitalization, contrastive conjunction, etc. on the sentiment of text is considered. VADER is fast enough to use online with streaming data and also it does not suffer from a speed-performance trade-off. These features make VADER one of the popular methods for sentiment analysis, especially on social media-related data.

In VADER, a group of well-established sentiment lexicons, like LIWC, ANEW, and GI, are used to construct a list. Incorporation of this list with lexical features common to sentiment expression in microblogs, including Western-style emoticons [4], sentiment related acronyms and initialisms [5], and commonly used slang [6] with sentiment value, provides over 9000 lexical feature candidates.

The wisdom-of-the-crowd is used to find an estimate for the sentiment valence of each candidate feature. Ten independent humans rate each of the features on a scale from -4 for extremely negative to 4 for extremely positive, and 0 for neutral. Only a lexical feature that has a non-zero mean rating, and whose standard deviation is less than 2.5, as determined by the aggregate of ten independent raters, is kept. These processes provide a set of 7,500 lexical features with valence

[1] www.liwc.net
[2] http://www.wjh.harvard.edu/ inquirer
[3] http://www.cs.uic.edu/ liub/FBS/sentiment-analysis.html
[4] http://en.wikipedia.org/wiki/List-of-emoticons
[5] http://en.wikipedia.org/wiki/List-of-acronyms
[6] http://www.internetslang.com/

TABLE I
PERFORMANCE OF THE MACHINE LEARNING MODELS ON SENTIMENT ANALYSIS IN THE STOCKTWITS DATASET

	Accuracy	Precision	Recall	F-measure	AUC
Logistic Regression	0.814	0.822	0.981	0.894	0.716
Naïve Bayes	0.808	0.809	0.996	0.893	0.714
Linear SVM	0.814	0.820	0.984	0.895	0.716

TABLE II
PERFORMANCE OF THE TEXTBLOB ON SENTIMENT ANALYSIS IN THE STOCKTWITS DATASET

	Accuracy	Precision	Recall	F-measure	AUC
TextBlob	0.810	0.842	0.726	0.780	0.804

TABLE III
PERFORMANCE OF THE SENTIWORDNET ON SENTIMENT ANALYSIS IN THE STOCKTWITS DATASET

	Accuracy	Precision	Recall	F-measure	AUC
SentiWordNet	0.870	0.837	0.661	0.739	0.806

TABLE IV
PERFORMANCE OF THE VADER ON SENTIMENT ANALYSIS IN THE STOCKTWITS DATASET

	Accuracy	Precision	Recall	F-measure	AUC
VADER	0.944	0.847	0.745	0.793	0.861

scores, which indicate the sentiment polarity and the sentiment intensity on a scale from -4 to +4 [12].

2) SentiWordNet: SentiWordNet is a lexical resource which uses sets of synonyms, or synsets, instead of individual terms. Their reasoning for this switch is that different senses of the same term may have different opinion-related properties. SentiWordNet assigns three numerical scores - Obj(s), Pos(s), and Neg(s) - to each synset of WordNet (version 2.0). These scores describe how Objective, Positive, and Negative the terms contained in the synset are.

SentiWordNet works based on training a set of ternary classifiers. These classifiers produce different results because they each train with a different training set and semi-supervised learning method. If all the ternary classifiers agree to assign the same label to a synset, that label will be assigned to that synset. Otherwise, each label will have a score proportional to the number of classifiers that have assigned it [13].

IV. EXPERIMENTS

In this section, we will describe how our experiment applies machine learning and lexicon based approaches to the StockTwits dataset.

Our experiment investigates if there is any relation between Bullish tweets and positive polarity, or Bearish tweets and negative polarity. In the following section, we seek to determine whether lexicon based models improve the accuracy of sentiment analysis of StockTwits data compared to machine learning approaches.

A. Machine Learning Approaches

As we mentioned before, 10% of messages in our dataset are labeled. In our experiment, we use these messages and supervised machine learning methods to classify StockTwits users' messages into either Bullish or Bearish sentiment. As Unigrams are used as features and infrequent unigrams that occur less than 300 times over all messages have been removed. In Table I, we provide the performance of Naïve Bayes, Linear Support Vector Machine (SVM), and Logistic Regression on StockTwits data based on different performance metrics. Based on Table I the performance of logistic regression, linear SVM, and Naive Bayes to classify messages to Bullish and Bearish is very close. The Accuracy of prediction is around 80%, F-measure around 90% and, Area Under the Curve is around 70%. In the following section, we try to see if we can adopt lexicons to improve the performance of the prediction.

B. Lexicon Based Approaches

1) TextBlob: The first method we used to extract the sentiment of messages in StockTwits data was TextBlob [14].

It uses a sentiment lexicon and the pattern.en sentiment analysis engine. Pattern.en leverages WordNet to score sentiment according to the English adjective used in the text. When TextBlob runs sentiment analysis on text, it returns a tuple of the form (polarity, subjectivity), where polarity is a float within the range [-1,1]. We first establish if there is any correlation between positive polarity and Bullish, and then negative polarity and Bearish. In order to compare the result of the machine learning approach to the lexicon-based approach, we apply TextBlob to the 2,522,557 messages that we used in the machine learning methods (Bearish around 500,000 and Bullish more than 2,000,000). From this set, TextBlob found 1,125,130 neutral messages. We remove all of the neutral messages and provide the result of comparing TextBlob sentiment on StockTwits data with the actual label of messages in Table II. Based on the results shown in Table II, TextBlob is not an effective method for extracting sentiment from StockTwits data. TextBlob's ineffectiveness is due to it labeling too many messages as neutral, and its performance metrics not being considerably improved in comparison to machine learning approaches.

2) SentiWordNet: Again, we consider a positive message as Bullish and a negative message as Bearish. Among all

TABLE V
NUMBER OF NEUTRAL MESSAGES

	TextBlob	SentiWordNet	VADER
neutral	1,125,130	214,972	899,503

of 2,522,557 messages, SentiWordNet found 214,972 neutral messages. All such neutral messages were removed, and then the result of SentiWordNet sentiment for each message was compared with the actual label of that message.

The result of applying SentiWordNet on StockTwits data is provided in Table III. Comparing Table III and I, it is clear that SentiWordNet can improve accuracy, precision and area under the curve values in comparison to machine learning models but still, the difference is not considerable. Although accuracy and AUC grow up around 9% f-measure reduce more than 10%.

3) VADER: Among all of 2,522,557 messages, VADER found 899,503 neutral messages and labeled them with zero. We remove all of the messages that VADER found as neutral and then compared VADER's determined sentiment with the actual label of each sentence.

Our results are shown in Table IV. We found that using VADER to predict the sentiment of the StockTwits users can improve accuracy, and area under the curve when compared to machine learning methods (Table I) TextBlob II and SentiWordNet III .

4) Combined Results: Figure 1 compares the ROC curves between machine learning methods and sentiment lexicon methods, including VADER, SentiWordNet, and TextBlob. Sentiment lexicons outperform machine learning methods based on these ROC curves. In Table V, we provide the number of messages that were labeled as neutral by TextBlob, SentiWordNet, and VADER. Fewer neutral messages indicate better performance from an analyzer, and so SentiWordNet clearly gives the best results here. However, Tables II, III, and IV reveal that, among the sentiment lexicon methods studied, VADER's higher performance metrics make it the best method for use in predicting StockTwits users' sentiment.

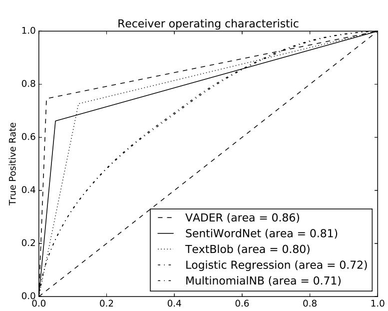

Fig. 1. Comparative Area Under the ROC curve for Lexicon versus Machine Learning based sentiment analysis

V. CONCLUSIONS

Knowing the sentiment of top authors, we can predict stock prices with accuracy of 75% but unfortunately, only 10% of messages in StockTwits are labeled. To increase the accuracy of stock price prediction, we need a powerful method for the sentiment analysis of top authors. Sentiment analysis has two main approaches - lexicon-based and machine learning. The primary drawback to machine learning is the training process, which is very time-consuming and computationally expensive. On the other hand, the lexicon-based approach does not need training data, and so it is favorable, particularly in tasks that involve high-dimensional data. There are a variety of lexicon-based methods that can be used to perform sentiment analysis. In this paper, we applied VADER, SentiWordNet, and TextBlob on StockTwits data to see if they can increase the accuracy of sentiment analysis. Logistic regression, Linear SVM, and Naive Bayes classification was used as our baseline and compared to the results of applying lexicon-based models alongside machine learning models. Based on our results, not only does VADER outperform machine learning methods in extracting sentiment from financial social media, like StockTwits, it is also faster.

REFERENCES

[1] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, pp. 1–8, March 2011.
[2] A. Mittal and A. Goel, "Stock prediction using twitter sentiment analysis," 2011.
[3] T. Loughran and B. McDonald, "When is a liability not a liability? textual analysis, dictionaries, and 10-ks," *THE JOURNAL OF FINANCE*, vol. LXVI, no. 1, 2011.
[4] E. H. Nadia Silva and E. Hruschka, "Tweet sentiment analysis with classifier ensembles," *Journal of Computational Science*, vol. 66, pp. 170–179, 2014.
[5] E. M. E. Fersini and F. Pozzi, "Automatic construction of financial semantic orientation lexicon from large scale chinese news corpus," *Decision Support Systems*, vol. 68, pp. 26–38, 2014.
[6] C. Potts and K. Pearson, "From frequency to meaning: Vector space models of semantics," *Journal of Artificial Intelligence Research*, vol. 37, pp. 141–188, 2010.
[7] G. Wang, T. Wang, B. Wang, D. Sambasivan, Z. Zhang, H. Zheng, and B. Y. Zhao, "Crowds on wall street: Extracting value from collaborative investing platforms," *The 18th ACM conference on Computer-Supported Cooperative Work and Social Computing (CSCW)*, March 2015.
[8] L. L. Bo Pang and shivakumar vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, vol. 66, pp. 79–86, 2002.
[9] M. Abadi, A. Agarwal, P. Barham, and E. Brevdo, "Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews," *Proceedings of the Association for Computational Linguistics*, vol. 66, pp. 417–424, 2002.
[10] B. Liu, "Sentiment analysis and subjectivity," in *Handbook of Natural Language Processing, Second Edition*. Chapman and Hall/CRC, 2010, pp. 627–666.
[11] C. Fellbaum, *WordNet*. Wiley Online Library, 1998.
[12] T. Mitra, C. J. Hutto, and E. Gilbert, "Comparing person-and process-centric strategies for obtaining quality data on amazon mechanical turk," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 1345–1354.
[13] A. Esuli and F. Sebastiani, "Sentiwordnet: A high-coverage lexical resource for opinion mining," *Evaluation*, pp. 1–26, 2007.
[14] S. Loria. (2017) Textblob: Simplified text processing. [Online]. Available: https://textblob.readthedocs.io/en/dev/