7-1-2013

Predictive Analytics On Public Data - The Case Of Stock Markets

Stefan Nann
University of Cologne, Cologne, Germany, stefan.nann@gmail.com

Jonas Krauss
University of Cologne, Cologne, Germany, krauss@wim.uni-koeln.de

Detlef Schoder
University of Cologne, Cologne, Germany, schoder@wim.uni-koeln.de

PREDICTIVE ANALYTICS ON PUBLIC DATA – THE CASE OF STOCK MARKETS

Nann, Stefan, University of Cologne, Pohligstr. 1, 50969 Köln, Germany, nann@wim.uni-koeln.de

Krauss, Jonas, University of Cologne, Pohligstr. 1, 50969 Köln, Germany, krauss@wim.uni-koeln.de

Schoder, Detlef, University of Cologne, Pohligstr. 1, 50969 Köln, Germany, schoder@wim.uni-koeln.de

Abstract

This work examines the predictive power of public data by aggregating information from multiple online sources. Our sources include microblogging sites like Twitter, online message boards like Yahoo! Finance, and traditional news articles. The subject of prediction are daily stock price movements from Standard & Poor's 500 index (S&P 500) during a period from June 2011 to November 2011. To forecast price movements we filter messages by stocks, apply state-of-the-art sentiment analysis to message texts, and aggregate message sentiments to generate trading signals for daily buy and sell decisions. We evaluate prediction quality through a simple trading model considering real-world limitations like transaction costs or broker commission fees. Considering 833 virtual trades, our model outperformed the S&P 500 and achieved a positive return on investment of up to ~0.49% per trade or ~0.24% when adjusted by market, depending on supposed trading costs.

Keywords: Predictive Analytics, Data Mining, Sentiment Analysis, Financial Markets, Twitter, Social Media.

1 Introduction

Recent academic discourse on information systems speaks of a paradigm shift toward data-intensive computing and "big data"-based studies (Hey, Tansley and Tolle, 2009). A new level of connectedness among peers creates a huge database by providing new ways for the dissemination and consumption of data and ever-easier means of collecting vast amounts of public data from various on- and offline resources, including posts, tweets, Web documents, and news feeds. Evolving data mining technologies and the increasing processing power of today's computers support the desire to appropriately analyze at least parts of today's growing (public) data deluge in real time, and thus tackle the core question of what meaningful information can be derived through algorithmic analyses and what predictive value can be inferred in automated fashion from public data.

Social media data in particular have been the subject of academic research in the recent past (Schoder et al., 2013). Asur and Huberman (2010) investigated the predictive power of tweets for box office returns, analyzing some 2.89 million tweets. Bollen, Mao and Zeng (2010) collected and classified tweets to forecast daily closing values of the Dow Jones Industrial Average. Among many others (Koch and Schneider, 2002; Forster, 2002; Antweiler and Frank, 2004; Wang, Jank and Shmueli, 2008; Xu et al., 2012), these studies represent the research field of predictive power in publicly available data. In their research essay, Shmueli and Koppius (2011) suggest a framework they call "predictive analytics," which is concerned with the assessment of predictive power in empirical research and statistical inference, and propose six roles for its application. This paper adopts the role of "the assessment of the predictability of empirical phenomena" from their framework.

The work presented herein extends previous attempts to assess the predictive power of social media talk by aggregating data from multiple resources (Twitter, eleven online message boards, and traditional news) and considering an extended period of six-month of data. The subjects of prediction are daily stock price movements from the Standard & Poor's 500 index (S&P 500). Most research concerned with stock market predictions based on online data is mainly theoretical in nature and does not take into account real-world limitations such as broker fees, bid/ask differences, and liquidity. To demonstrate the potential practical application of our findings, we describe a simple trading model based on the predictor, considering commission fees, transaction costs, and stock liquidity.

2 Literature Review

This section provides a brief summary of past literature concerned with the predictive power of online data for financial markets and shows how results improved with the progress of time. While older research found that discussion followed market movements, more recent results clearly detect predictive value in online data for stock price changes.

There are two main research streams which are relevant for the scope of this work. The literature can be categorized in works related to sentiment analysis and works related to predictive power of user generated content (UGC). Sentiment analysis is a broad research field and is applied on many different domains (Berger, Della Pietra and Della Pietra, 1996; Pang, Lee and Vaithyanathan, 2002; Whitelaw, Garg and Argamon, 2005; Abbasi, Chen and Salem, 2008; Boiy and Moens, 2009; Choi, Kim and Myaeng, 2009; Lin and He, 2009; Narayanan, Liu and Choudhary, 2009; Mizumoto, Yanagimoto and Yoshioka, 2012; Fang, Datta and Dutta, 2012). The second major research stream relevant for this study relates to works of predictive power of UGC. Although this topic is much broader and applies to many different domains, the following articles focus on making predictions for developments in financial markets based on UGC.

In many studies both streams are tied together since the value of user generated content can be captured better when it is analyzed with automated methods (Antweiler and Frank (2004), Das and Chen (2007), Bollen, Mao and Zeng (2010), Zhang and Swanson (2010), Sprenger and Welpe (2010)).

These authors apply sentiment analysis for extraction of predictive value from UGC and to study its impact on the stock market. Antweiler and Frank (2004) studied the predictive power of online message boards for the stock market by analyzing 1.5 million messages from Yahoo! Finance (http://finance.yahoo.com) and Raging Bull. Applying sentiment analysis, they found that the number of messages is a predictor for stock turnover. However, their model's performance would not deliver a significant return on investment, as plausible transaction costs would be too high. Das, Martinez-Jerez and Tufano (2005) found that sentiment follows stock price returns.

All more recent research applies sentiment analysis to a changing number of messages from a variety of online resources. While Oh and Sheng (2011) looked at a comparably small subset of messages from Stocktwits, Bollen, Mao and Zeng (2010) collected a large amount of ~9.85 million microblog postings from Twitter (http://www.twitter.com). Schumaker et al. (2012) looked at a small sample (9211 news articles) of traditional news articles; the amount of data for other studies falls between these parameters. However, the latest work focuses on a single source of data (mainly Twitter), leaving out other, well-researched sources such as Yahoo! Finance, Raging Bull, or traditional finance news – which may or may not improve the results.

Traditional news evaluated by natural language processing can carry alpha information as well (Cohen and Frazzini, 2008; Schumaker and Chen, 2009; Dion et al., 2011). Alpha information refers to information that is not yet reflected in stock price levels, thus leading to future stock price movement according to the Efficient Market Hypothesis (EMH) (Fama, 1970).

There is a number of studies over the last decade which found predictive evidence of UGC on stock return (Bagnoli, Beneish and Watts, 1999, Tumarkin and Whitelaw, 2001, Jones, 2006, Gu et al., 2006, Das and Chen, 2007, and Sabherwal, Sarkar and Zhang, 2008).

Bollen, Mao and Zeng (2010), Zhang and Swanson (2010), Sprenger and Welpe (2010), Oh and Sheng (2011), and Xu et al. (2012) are examples of recent studies that have found clear evidence for the predictive power of online communication for stock price movements. Oh and Sheng (2011) examined ~72,000 microblog postings from Stocktwits.com, extending over a three-month period, to predict stock price movements. Applying sentiment analysis, they found microblog messages predict future stock price movement. They also briefly evaluated potential return on investments, finding that simple (not adjusted returns) deliver better results than market-adjusted returns.

Our work and that of other researchers hypothesizes that online talk in social media and microblogs has predictive power over future stock price movements. Microblog posts in particular are characterized by strong focus on their subject because they are succinct, happen in nearly real-time, and have high posting frequencies (Xu et al., 2012; Oh and Sheng, 2011; Bollen, Pepe and Mao, 2010; Java et al., 2007).

Andrew Lo (2004) provides a theoretical foundation for the predictive power of public data over financial markets with the Adaptive Market Hypothesis (AMH), as suggested by Brown (2012). Taking behavioral economics and finance research (De Long et al., 1990; Hirshleifer, 2001; Camerer and Loewenstein, 2004; Tetlock, 2007; Xu and Zhang, 2009; Zhang and Swanson, 2010) into consideration, the AMH describes an evolutionary model of individuals adapting to a changing environment via simple heuristics. It provides an explanation for the existence of alpha information and how learning and competition gradually restore market efficiency (Neely, Weller and Ulrich, 2009). Thus, social media, microblog posts, and news could be considered contributors to the competition and learning process that drives prices (Brown, 2012).

This paper contributes by providing a case study of a virtual trading model based on the predictive power of online communication. Our first goal is to demonstrate that it is indeed possible to trade based on an online message board predictor and achieve a positive return on an investment (adjusted by the market). We gathered communication from Twitter, eleven online message boards, and Yahoo! Finance's news stream, thus extending the scope of data utilized in earlier research.

3 Methodology and Sentiment Analysis

For this paper, we collected 2,971,381 messages concerned with stocks of the S&P 500 index during a six-month period from June 1 to November 30, 2011. Table 1 shows the different online sources which we have used to collect the messages.

The first step is to assign these messages to stocks. For that we filtered the dataset by either looking for messages in sub-forums concerned with particular stocks on online message boards or by using Twitter's "cash tag". The cash tag is a stock's ticker symbol with a preceding dollar sign ($). Not considering spam at this point (more details on how we filter spam in the following paragraph) we can rely on that a tweet which contains a cash tag refers to the stock price or anything related to the financial value of the underlying company (e.g. $MSFT for company Microsoft).

In sources other than Twitter stock specific communication can be accessed in sub-forums which exist for each component of the S&P 500 index. In these sub-forums users exclusively discuss topics related to a particular company. For example Yahoo! Finance provides direct access to a stock's sub-forum by appending the ticker symbol to the following hyper-reference: http://finance.yahoo.com/mb/ (e. g. http://finance.yahoo.com/mb/MSFT for company Microsoft). With a similar approach specific sub-forums can be accessed for all other sources listed in table 1. With this method, each tweet / forum post can clearly be assigned to a single company which is important to ensure only relevant communication is considered when calculating sentiment.

Thus we get a relatively precise assignment of messages to a specific company / stock which helped us to avoid some common name entity conflicts as mentioned in Yerva & Miklós & Aberer (2010). Although we cannot rely with full certainty, we can assume that people talk about the company or company-related issues when posting in the financial discussion board about Apple and not about the fruit apple. Nevertheless, after collecting all messages we applied a spam filter which cleaned our data set. Our self-developed spam filter searches for example for posts which intend to insult other users without contributing relevant information. Most of these posts can be identified by the usage of scurrile and nasty language. Everything related was removed from the data set.

Following the state-of-the-art research as established in the previous section, in the next step we applied sentiment analysis to microblog messages, forum posts, and traditional news using a Naïve Bayes classifier with an adapted bag-of-words in combination with part-of-speech tagging to find negations and spam filtering based on keywords. The basis of the applied sentiment methodology can be found in Krauss and Nann and Schoder (2012). One key finding in this study indicates that the quality of sentiment recognition depends on how specific the sentiment analysis algorithms are adjusted to the analyzed context. The more context-specific the algorithms are designed the higher the quality of sentiment recognition will be. This is determined e.g. by the choice of bag-of-words and the adjustment of part-of-speech tagging. For example, authors use different language and words to write a positive review about a digital camera and say something positive about their favorite stock. In the current study we adjusted the sentiment analysis very specifically to the stock market domain.

For this reason we initially read a few hundred tweets and posts from the available dataset and manually annotated it with positive or negative sentiment. This sample data was used to train our self-developed sentiment algorithm. In the next step we applied the trained algorithm to a newly and not annotated data sample from the available data set to determine the precision of the sentiment algorithm. Also in a manual process we defined lists of positive (e.g. buy, long, call, etc.) and negative (e.g. sell, short, put, etc.) words which resulted in our bag-of-words. During text analysis the algorithm scans the content for these words. We also manually defined specific words for part-of-speech tagging. The word "don't" for example will be used during part-of-speech-tagging, if a user posts "I don't sell my shares", this will be recognized and labeled with positive sentiment since the key word "don't" will give the negative key word "sell" the opposite meaning.

After all, the algorithm calculates a ratio (decimal number) based on the occurrences of positive and negative labels in a tweet or post. The sum of all ratios for all messages of a specific stock represents the aggregated sentiment value which was used to predict the daily stock price.

Messages of all sources were considered with equal weight in our model. It is obvious that different data sources contain different value contributions. Tweets will probably have a lower half-life than traditional news which usually references a longer time period. It is subject of further research to design a process which evaluates these aspects of every single source separately. Please also refer to section 7 Discussion and Research Suggestions.

Source	URL	Number of Messages
Clearstation	http://www.clearstation.com	12,743
Free Realtime	http://quotes.freerealtime.com	610
Hotstockmarket	http://www.hotstockmarket.com	1,818
Investor's Hub	http://investorshub.advfn.com	20,359
Investor Village	http://investorvillage.com	37,180
The Motley Fool	http://www.fool.com	10,587
Raging Bull	http://ragingbull.quote.com	15,331
Silicon Investor	http://www.siliconinvestor.com	21,442
Stockhouse	http://www.stockhouse.com	37,119
The Lion	http://www.thelion.com	5,766
Twitter	http://www.twitter.com	1,801,345
Yahoo! Finance Boards	http://messages.finance.yahoo.com	802,476
Yahoo! Finance News	http://finance.yahoo.com	204,605

Table 1. Data Sources

As not all stocks from the S&P 500 index receive equal attention in social media, there are substantial differences in the average number of messages written each day for different stocks. For instance, Apple Inc. is one of the most discussed equities on the Web and many more messages are posted for Apple Inc. than for other index components. Thus we require an adjustment of sentiment values based on the average number of messages. For this work we chose the simple moving average (SMA) to achieve comparability for equities of differing attention levels. Sentiment values were used as a stock price movement predictor, with positive values indicating an upward movement and negative values indicating a downward movement.

We began calculating our predictor one month after commencing data collection since we used a 30-day simple moving average (SMA30) to calculate sentiment values. Thus, predictions of stock price movements were made each trading day from June 1 to November 30, 2011. For each trading day t, the predictor considered sums of positive and negative messages for each stock on the S&P 500 index and weighted them based on the SMA30. This value was used to predict stock price change on day t + 1, predicting an increase in the case of positive values and a decrease in the case of negative values. Sentiment values can assume any value larger or smaller than zero.

$$Sentiment\ Predictor = \frac{\text{No. of positive messages}}{\text{SMA30 of positive messages}} - \frac{\text{No. of negative messages}}{\text{SMA30 of negative messages}}$$

Through weighting current messages in relation to the SMA30, it is possible to compare stocks that have significantly different attention levels and thus strongly differing message averages. We chose a 30-day average because it takes into account enough days to even out positive and negative peaks in communication without being too static in comparison to longer periods. Longer periods would carry the danger of ignoring short-term anomalies in communication – e.g. in the case of earning releases or bankruptcies – which often lead to a strong increase in message numbers.

Each trading day, we determined the level of sentiment (threshold) for each stock where the historical ratio of correct to total predictions is maximized. Sentiment values count as prediction signals only if the absolute sentiment value lies above the threshold. For instance, for a threshold of two, a sentiment value of one would not be considered a signal to trade. We systematically determined historic prediction ratios by conducting a sensitivity analysis of sentiment thresholds each trading day for each stock. This was done by looking at all past trading days, comparing the sentiment on day t with the stock price change on day t + 1, summing dates on which a positive/negative sentiment on day t corresponded with a positive/negative stock price change on day t +1, and finally building the ratio for each threshold. Results were statistically significant on a level of 0.05 (right-sided significance test). On average, we obtained 10 stocks per trading day that had statistically significant prediction ratios greater than 0.5. We further found that a look ahead of one day delivered the best results, which means that sentiment values of day t predicted stock price changes of day t + 1 with the highest accuracy compared to predictions for day t + 2 or day t + 3.

$$\frac{sum\ correct\ predictions\ (sentiment\ threshold)}{total\ predictions\ (sentiment\ threshold)} \rightarrow max$$

4 Results: Prediction Ratio of Sentiment Signals

Our results show that publicly available data in microblogs, forums, and news on day t have predictive power for stock price changes on day t + 1. We confirm the findings of Bollen, Mao and Zeng (2010), Sprenger and Welpe (2010), and Oh and Sheng (2011) for single source-based predictions and extend their validity to the case of using multiple sources in aggregation. Table 2 displays the overall prediction performance and the performance for each month.

Predictions	All	Ratio Buy Predictions	Ratio Sell Predictions
Entire period	60.38%	60.69%	60.03%
November 2011	57.54%	60.71%	54.74%
October 2011	63.76%	63.16%	64.81%
September 2011	52.63%	48.98%	55.86%
August 2011	67.34%	67.52%	67.18%
July 2011	60.99%	63.38%	56.79%
June 2011	59.73%	59.01%	60.61%

Table 2. Prediction Ratios

The percentage values display the ratios of correct predictions for all analyzed stocks over the entire period from June 1 to November 30, 2011. For example 60.38% means that ~60 percent of all stocks for which sentiment had significant prediction ratios in the past delivered correct predictions in the period considered. In sum, the algorithm made 1,300 predictions over the entire period (126 trading days). Table 2 lists only predictions for rising stock prices through positive sentiment and predictions of falling stock prices through negative sentiment. Both cases do not differ significantly.

5 Trading Model and Model Parameters

To extend the academic body of literature, and especially to go beyond more recent research as illustrated by, for example, Bollen, Mao and Zeng (2010), Sprenger and Welpe (2010), and Oh and Sheng (2011), we demonstrate the potential practical application of our findings. Here we describe a simple trading model based on the predictor, considering commission fees, transaction costs, and stock liquidity. Most research concerned with stock market predictions based on online data is mainly theoretical in nature and does not take into account real-world limitations when considering a return on investments. Although we are taking some of these factors into account we do not propose a complete trading strategy which could be executed on the stock market as is. We did not execute trades under real market conditions.

Our model is based on assumptions and simplifications that are as practical as possible. However, on real-world trading floors and in real trading environments, there are many factors that can influence a trading model that works perfectly in theory. In our opinion, it is most critical to control for large spreads (differences between the bid and ask prices of stocks), the traded volume of a stock (the more a stock is traded, the higher is its liquidity and the higher the chance to buy or sell the stock for the desired price), and broker commissions, which become particularly relevant if a strategy is based on several trades per day (as is ours).

Table 3 shows our three most relevant and important assumptions to simulate a practical trading strategy. For our trading model, we considered only stocks from the S&P 500 index (as described in the section above). The S&P 500 is one of the most important U.S. stock market indices (and probably also in the world), and stocks on the index are traded mostly in the United States. Therefore, all stocks from this index have comparably high trading volumes, which are important to ensure liquidity. To enforce this criterion, we considered only stocks with a trading volume larger than 3 million shares traded on average per day. For these stocks, it is almost guaranteed that shares can be bought in the morning at the opening of the stock market and be sold at the closing bell. Further, we only considered stocks that cost more than $10 on the day the trade is to be executed. This is important to guarantee a relatively small spread.

Criteria	Description
Tradable Stocks	Only trading stocks from S&P 500 index
Stock Volume	> 3,000,000 traded shares/per day on average
Stock Price	> $10 (on the trading day selected)

Table 3. Assumptions for our trading model

The difference between bid and ask prices (spread or transaction costs) is the highest cost factor for a trading model which is based on multiple trades on one day. Transaction costs are commonly expressed as basis points in a finance context. A basis point (bp) is a unit of measure used in finance to describe the percentage change in the value or rate of a financial instrument. One basis point is equivalent to 0.01% (1/100th of a percent) or 0.0001 in decimal form[1]. The difference between the ask/bid price of a stock, which is the price that has to be paid when buying/selling the stock, and the actual stock price typically lies between 10 bp and 20 bp (0.1% to 0.2%) per transaction. However, these values are based on our assumptions that a stock has a trading volume of more than 3 million shares on average per day and is worth at least $10 on the day it is traded. For example, if a stock trades at $10.02 the broker might charge $10.03, which implies trading costs of 10 bp or 0.1%.

For our trading model, we followed a simple trading rule:

- If the sentiment predictor is positive, the strategy is to buy the stock on market open (open long position) and sell the stock (close long position) on market close (on the same day).
- If sentiment predictor is negative, the strategy is to sell the stock (open short position) on market open and buy the stock (close short position) on market close (on the same day).

We assumed that we could buy the stocks for their opening prices and sell them for their closing prices every day[2]. Considering criteria from Table 3 and its intersection with our algorithm's sentiment signals, we obtained about 7 tradable stocks on average each day (833 trades which meet the criteria on 126 trading days). We obtained a daily overall return on investment (ROI) by summing individual

[1] http://www.investopedia.com/ask/answers/05/basispoint.asp#axzz1tcGlLhM7

[2] Buying a stock for the opening price and selling for the closing price is a simplification and is not necessary applicable in practice. There are many reasons why it may not be possible to buy the stock for the opening price and sell it for the closing price (e.g., transaction execution time).

ROIs for these stocks. This was done for the entire period from June 1 to November 30, 2011. To simulate the model in a more realistic way, we also adjusted our ROI with market movement in the considered time period. For this purpose, we used the change of the SPY certificate, which is one of the most liquid and heavily traded titles on the financial market. The trading rule for adjusting ROIs with the market was as follows:

- If the signal is positive, we buy (long) the stock and sell (short) the market.
- If the signal is negative, we sell (short) the stock and buy (long) the market.

The rule of adjusting results with changes of the market is a form of hedging. Concretely, this means that we limited losses if a positive/negative sentiment predictor turned out to be wrong and the stock price actually rose/fell. In many cases, stocks correlate with general market movement, which means that a stock often rises/falls when the market rises/falls. This is the reason that we traded the opposite of the current sentiment predictor on the market. This means that if a stock fell after a positive sentiment, we limited our losses by being short on a falling market. Table 4 (adjusted by market) and Table 5 (unadjusted) summarize the performance of our model with various levels of transaction costs.

6 Results from Trading Model Considerations

Transaction Costs	0 bp	5 bp	15 bp	20 bp	25 bp
Total ROI	196.84%	155.19%	71.89%	30.24%	-11.41%
Avg. ROI (per trade)	0.2363%	0.1863%	0.0863%	0.0363%	-0.0137%
No. of trades	833	833	833	833	833

Table 4. Performance with adjustment by the market (SPY certificate)

The numbers in table 4 show that hedging with the market results in a lower total ROI (~197%, when not considering transaction costs). In total our model required 833 trades over the entire period of time resulting in average ROIs of 0.24% (adjusted) and 0.49% (unadjusted) without considering transaction costs (0 bp). Average ROIs are calculated by dividing the total ROI by the number of trades. This results in an average ROI per trade. In Tables 4 and 5 we also display the resulting performances of our trading model for different trading costs expressed as basis points. Broker commission fees need to be subtracted, after all. This is generally a constant amount that highly depends on the broker. Usually, one can calculate it as $10 per transaction, independent of the traded volume. However, considering the return of our model it would be possible to subtract a constant broker fee and still receive a positive result.

Transaction Costs	0 bp	5 bp	15 bp	20 bp	25 bp
Total ROI	409.42%	367.77%	284.47%	242.82%	201.17%
Avg. ROI (per trade)	0.4915%	0.4415%	0.3415%	0.2915%	0.2415%
No. of trades	833	833	833	833	833

Table 5. Performance without adjustment by the market

Figure 1 shows the comparison of cumulated daily ROIs of our (unadjusted) trading model (without transaction costs) and cumulated daily price changes of the SPY certificate. We used the SPY certificate because it has very high trading liquidity and very accurately represents the current market value of the S&P 500 index. Thus, the SPY certificate is a suitable benchmark for our specific trading model. To compare its performance with our model the rule is as follows: buying the SPY certificate on the start day (price: 131.87) of the trading model and selling it on the last day (price: 124.99). The certificate lost 5.22% in this time; therefore, we outperformed it in almost every scenario from Tables 4 and 5 only when assuming transaction costs of 25bp and applying market adjustment that the certificate could beat our model.

It is obvious that a significant part of the performance of our trading model has been achieved during the period from the end of July until the beginning of September. This might be a result from a certain market phase well suited for our model or statistical effects such as a temporary increase in correct signals for equities making large price changes during that time; more on that in the discussion section.

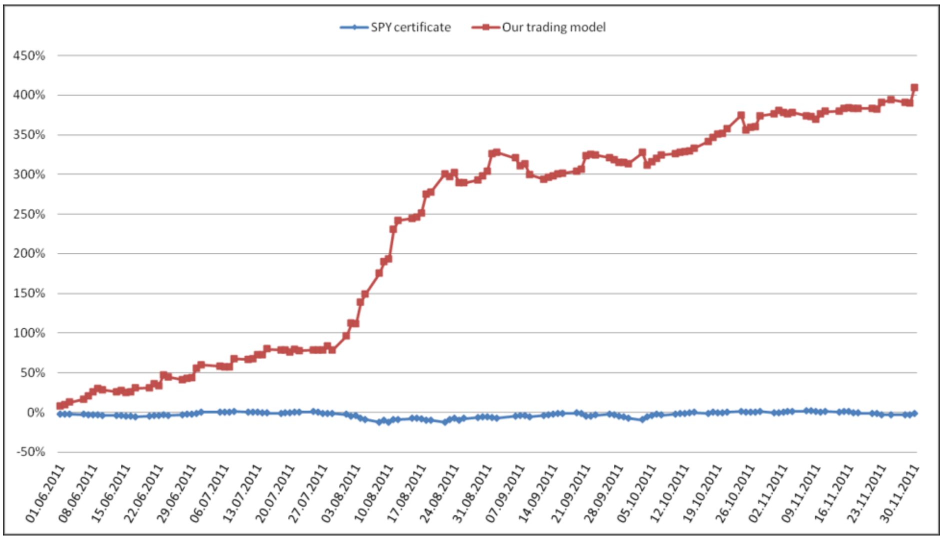

Figure 1. Comparison of cumulative returns on investments for our trading model and the SPY certificate from June 1 to Nov 30.

7 Discussion and Research Suggestions

Our findings confirm previous studies that investigated the relationship between online talk and financial markets and extend their validity to the case of multiple sources. It is obvious that predictive power rests in publicly available data. However, until now it has been uncertain whether this power has substantial value. Antweiler and Frank (2004) wrote: "This effect is statistically significant but economically quite small in comparison to plausible transaction cost." At least for the period between June 1 and November 30, 2011, we show that a positive ROI was achieved by trading based on public message board data – contrary to their results.

However, causality is uncertain and should be subject to further research. We would emphasize in particular that the period of time we used for our analysis is rather short. Thus, we cannot rule out that certain market conditions that were present during that period were responsible for the positive ROI of our model (e.g. the period from July to September when signals performed significantly better than on average). Long-term studies are required extending over different market and economic phases to address this limitation. Additionally, we used open and close prices, which is a simplification: real-world trading results will differ, either for better or worse.

Another shortcoming at this stage of our research is the comparison of models based on different data sources. We assume that Twitter, online message boards, and traditional news differ with respect to their predictive power for stock price movements. Thus, for further research we would like to explore differences in prediction ratios between the various sources we analyzed in this study. Additionally, we point to the method of sentiment analysis applied herein. It is clear that more sophisticated methods would provide better quality of correctly detected sentiment values in texts. Thus, the trading model's performance could potentially be improved by reaching higher quality in sentiment recognition.

Further it is necessary to analyze effects of applying different sentiment analysis methodologies. The quality a particular methodology delivers is probably related to the performance of the predictor. We hypothesize that a higher quality in sentiment analysis should lead to an increased number of correct predictions. Thus, as an extension of this work, we would like to compare the performance of the predictor when based on our own sentiment algorithms with the performance when other sentiment analysis tools are used for text classification. Improvement of performance could potentially also be achieved by applying separate sentiment analysis methods for different sources. E. g. Twitter communication is different in nature from forum communication, thus an adapted classification method might improve sentiment quality (Sriram et al. (2010)).

References

Abbasi, A., Chen, H. and Salem, A. (2008). Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums. ACM Trans. Inf. Syst. (26:3) 1-34.

Anonymized (2012). Towards Universal Sentiment Analysis through Web Mining.

Antweiler, W. and Frank, M. (2004). Is all that talk just noise? The information content of Internet stock message boards. Journal of Finance. 59(3):1259-1295.

Asur, S. and Huberman, B.A. (2010). Predicting the Future with Social Media. arXiv:1003.5699v1.

Bagnoli, M. E., Beneish, M. D. and Watts, S. G. (1999). Whisper Forecasts of Quarterly Earnings per Share. http://ssrn.com/abstract=74369, retrieved 04/22/2012.

Berger, A.L., Della Pietra, S.A., and Della Pietra, V.J. (1996). A Maximum Entropy Approach to Natural Language Processing. Computational Linguistics (22:1) 39-71.

Boiy, E. and Moens, M. (2009). A Machine Learning Approach to Sentiment Analysis. In Multilingual Web Texts, Information Retrieval (12:5) 526.

Bollen, J., Mao, H. and Zeng, X. (2010). Twitter mood predicts the stock market. arXiv Server.

Bollen, J., Pepe, A. and Mao, H. (2010). Modeling public mood and emotion: Twitter sentiment and socioeconomic phenomena. Proceedings from 19th International World Wide Web Conference Raleigh, North Carolina.

Brown, G.W. (1999). Volatility, sentiment, and noise traders. Financial Analyst Journal, Vol. 55 No. 2, pp. 82-90.

Brown, E.D. (2012). Will Twitter Make You a Better Investor? A Look at Sentiment, User Reputation and Their Effect on the Stock Market. SAIS 2012 Proceedings. Paper 7.

Das, S. and Chen, M. (2007). Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web. Management Science 53(9):1375-1388.

Das, S., Martinez-Jerez, A. and Tufano, P. (2005). eInformation: A Clinical Study of Investor Discussion and Sentiment. Financial Management 34(3):103-137.

Camerer, C. F. and Loewenstein, G. (2004). Behavioral Economics: Past, Present and Future. In Advances in Behavioral Economics, C.F. Camerer, G. Loewenstein and M. Rabin, eds. Princeton, NJ: Princeton University Press.

Choi, Y., Kim, Y. and Myaeng, S.H. (2009). Domain-Specific Sentiment Analysis Using Contextual Feature Generation. In Proceeding of the 1st international CIKM Workshop on Topic-sentiment Analysis for Mass Opinion, ACM, Hong Kong, China, pp. 37-44.

Cohen, L. and Frazzini, A. (2008). Economic Links and Predictable Returns. Journal of Finance 63:1977-2011.

De Long, J., Shleifer, A., Summers, L. and Waldmann, R. (1990). Noise trader risk in financial markets. Journal of Political Economy 98:703-738.

Dion, M., Shaikh, V., Pessaris, A., Malin, S., Smith, R., Lakos-Bujas, D., Okamoto, S. and Hlavaty, B. (2011). News Analytics - Can They Add Value to Your Quant Process? – Using a Language Recognition Algorithm to Analyse News Flow. J.P. Morgan Europe Equity Research, June 3rd, 2011.

Fama, E. (1970). Efficient capital markets: a review of theory and empirical work. Journal of Finance. 25:383-417.

Fang F., Datta, A., and Dutta, K. (2012). A Hybrid Method for Cross-domain Sentiment Classification Using Multiple Sources. ICIS 2012, Orlando, Florida, USA.

Forster, M. R. (2002). Predictive Accuracy as an Achievable Goal of Science. Philosophy of Science (69:3), pp. S124-S134.

Gu, B., Konana, P., Rajagopalan, B. and Chen, M. (2007). Competition Among Virtual Communities and User Valuation: The Case of Investing-Related Communities. Information Systems Research 18(1):68-85.

Hey, T., Tansley, S. and Tolle, K. Eds. (2009). The Fourth Paradigm – Data-Intensive Scientific Discovery. Microsoft Research, Redmond, Washington.

Hirshleifer, D. (2001). Investor Psychology and Asset Pricing. Journal of Finance 56(4):1533-1597.

Java, A., Song, X., Finin, T. and Tseng, B. (2007). Why We Twitter: Understanding Microblogging Usage and Communities. Joint 9th WEBKDD and 1st SNA-KDD Workshop. San Jose, CA.

Jones, C., (2006). A Nonlinear Factor Analysis of S&P 500 Index Option Returns. The Journal of Finance, Vol. LXI, No. 5, Oct. 2006, pp. 2325-2363.

Koch, S. and Schneider, G. (2002). Effort, co-operation and co-ordination in an open source software project: GNOME. Information Systems Journal 12(1): 27-42.

Krauss, J., Nann, S., and Schoder, D. (2012). Towards Universal Sentiment Analysis through Web Mining. Poster Session, European Conference on Information Systems, Barcelona, Spain.

Lin, C. and He, Y. (2009). Joint Sentiment/Topic Model for Sentiment Analysis. In Proceeding of the 18th ACM conference on Information and Knowledge Management, ACM, Hong Kong, China, pp. 375-384.

Lo, A.W. (2004). The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective, Journal of Portfolio Management: 30th Anniversary Issue), 15-29.

Mizumoto, K., Yanagimoto, H., and Yoshioka, M. (2012). Sentiment Analysis of Stock Market News with Semi-supervised Learning. Proceedings of the 2012 IEEE/ACIS 11th International Conference on Computer and Information Science, 325-328.

Narayanan, R., Liu, B. and Choudhary, A. (2009). Sentiment analysis of conditional sentences. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1, Association for Computational Linguistics, Singapore, pp. 180-189.

Neely, C. J., Weller, P. A. and Ulrich, J. M. (2009). The Adaptive Markets Hypothesis: Evidence from the Foreign Exchange Market. Journal of Financial and Quantitative Analysis (44:2)467-488.

Oh, C. and Sheng, O. (2011). Investigating Predictive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movement. ICIS 2011, Shanghai, China.

Pang, B., Lee, L. and Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing – Volume 10, Association for Computational Linguistics, pp. 79-86.

Sabherwal, S., Sarkar, S. and Zhang, Y. (2008). Online talk: does it matter?. Managerial Finance 34(6):423-436.

Schoder, D., Gloor, P., and Metaxas, P. T. (2013). Social Media and Collective Intelligence – Ongoing and Future Research Streams. In Künstliche Intelligenz, Special Issue on Social Media. Forthcoming 2013.

Schumaker, R. and Chen, H. (2009). Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFinText System. ACM Transactions on Information Systems 27(2).

Schumaker, R., Zhang, Y., Huang, C.-N., and Chen, H. (2012). Evaluating sentiment in financial news articles. Decision Support Systems, Volume 53, Issue 3, June 2012, Pages 458–464.

Shmueli, G. and Koppius Otto R. (2011). Predictive Analytics. In Information Systems Research MIS Quarterly, Vol. 35, No. 3, pp. 553-572.

Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., and Demirbas, M. (2010). Short text classification in twitter to improve information filtering. In F. Crestani, S. Marchand-Maillet, H.-H. Chen, E. N. Efthimiadis, & J. Savoy (Eds.), Proceeding of the 33rd international ACM SIGIR conference on research and development in information retrieval (SIGIR 2010), Geneva, Switzerland, July 19–23 (pp. 841–842). ACM.

Sprenger, T.O. and Welpe, I.M. (2010). Tweets and Trades: The Information Content of Stock Microblogs. http://ssrn.com/abstract=1702854, retrieved 04/22/2012.

Tetlock, P. (2007). Giving content to investor sentiment: The role of media in the stock market. Journal of Finance 62:1139-1168.

Tumarkin, R. and Whitelaw, R. (2001). News or noise? Internet postings and stock prices. Journal of Financial Analysts 57(3) 41-51.

Wang, S., Jank, W. and Shmueli, G. (2008). Explaining and Forecasting Online Auction Prices and Their Dynamics Using Functional Data Analysis. Journal of Business and Economic Statistics. (26:3), pp. 144-160.

Whitelaw, C., Garg, N. and Argamon, S. (2005). Using Appraisal Groups for Sentiment Analysis. In Proceedings of the 14th ACM International Conference on Information and Knowledge Management, ACM, Bremen, Germany, pp. 625-631.

Xu, W., Li, T., Jiang, B., and Cheng, C. (2012). Web Mining For Financial Market Prediction Based On Online Sentiments. PACIS 2012 Proceedings. Paper 43.

Xu, S. and Zhang, X. (2009). How do social media shape the information environment in the financial market?. ICIS 2009.

Yerva, Surender Reddy, Miklós, Zoltán, and Aberer, Karl (2010). It was easy, when apples and blackberries were only fruits. EPFL working paper, http://infoscience.epfl.ch/record/151616/files/LSIR_WePS3_Paper.pdf, retrieved 03/30/2013.

Zhang, Y. and Swanson, P. (2010). Are day traders bias free? – Evidence from Internet stock message boards. Journal of Economics Finance 34:96-112.