Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media

Trading Strategies to Exploit Blog and News Sentiment

Wenbin Zhang and Steven Skiena
{wbzhang@cs.sunysb.edu} and {skiena@cs.sunysb.edu}
Department of Computer Science, Stony Brook University
Stony Brook, NY 11794-4400 USA

Abstract

We use quantitative media (blogs, and news as a comparison) data generated by a large-scale natural language processing (NLP) text analysis system to perform a comprehensive and comparative study on how company related news variables anticipates or reflects the company's stock trading volumes and financial returns. Building on our findings, we give a sentiment-based market-neutral trading strategy which gives consistently favorable returns with low volatility over a long period. Our results are significant in confirming the performance of general blog and news sentiment analysis methods over broad domains and sources. Moreover, several remarkable differences between news and blogs are also identified.

Introduction

The efficient market hypothesis asserts that financial markets are "informationally efficient", which means current stock prices already reflect all known information and all occurred facts. Therefore, investors cannot make excess profits from the market if their trading strategies are based on known information, because market prices are efficiently collecting and aggregating various information and keep changing without delay. Nevertheless, some encouraging results prove the conditional usage of efficient market hypothesis. For example, (Chan 2003) shows that stock prices appear to under-reaction to information, which suggests news may provide a feasible way to analyze financial markets.

Our primary goal is to study the relationship between stock market data and linguistic media data, and to illustrate the extent to which they can contribute to the design of investment strategies. Our main contributions are:

- *Comparative Study of Blogs and News* – We conduct a thoughtful comparative study of four different linguistic sources including one news and three blog-type sources.
- *Large-Scale Analysis* – We give comprehensive results of analyzing stock market using roughly one terabyte of media data and thousands of companies. This scale of analysis has never been previously attempted in literatures.
- *Sentiment-oriented Equity Trading* – Based completely on blog/news sentiment data, we propose a market-neutral stock trading strategy and yield intriguing returns with low variance (ignoring both transaction costs and the timing issue described in Section *Media Timing Issues*).
- *Validation of Sentiment Analysis Methods* – Perhaps another important contribution of our paper is the strongest validation to date of the accuracy of our media sentiment analysis methodology of *Lydia*.

This paper is organized as follows. First we review related work. We then describe the media and financial data we work with. After that, we give a complete analysis of the correlation between major stock variables and major media variables. Finally, we propose and evaluate a market-neutral trading strategy based on media data and give conclusions.

Related Work

Previous work is divided between the finance and computer science academic communities.

We first survey research from the financial realm. Tetlock (Tetlock, Saar-Tsechansky, and Macskassy 2007) investigates the negative words in firm-specific news articles and claims that firms' stock prices under-react to the underlying negative information of news articles. Chan (Chan 2003) examines monthly returns to a subset of stocks after public news about them is released and finds that investors react slowly to information, especially after bad news.

From the Computer Science side, intense researches are delivered by text mining or machine learning communities. Their basic idea is to quantify linguistic information with text mining techniques, get the predefined set of features of training data, and then build various models with statistical approaches or statistical learning algorithms. They are shown in survey paper (Mittermayer and Knolmayer 2006). In addition, there has also been substantial interest in the opinion mining and NLP community on using financial text streams to verify sentiment analysis methods. Pang and Lee (Pang and Lee 2008) gave a detail review in this domain.

Stock and Media Data

Stock Data

Our stock price and volume data is obtained from Thomson Datastream Services (Datastream). Here we only consider the stocks listed in the New York Stock Exchange because those stocks have more intensive media coverage than others. We downloaded the data of all 3238 stocks within the period from 2005 to 2009, for their daily open, close, high, low prices, turnover volumes, and monthly market caps.

Media Data

Company-related blog and news data was generated using the *Lydia* ((Lloyd, Kechagias, and Skiena 2005), *http://www.textmap.com*), a high-speed text processing system, which reduces large text streams to time series data on the frequency of sentiment of underlying media entities. In this paper, we compare four different sources:

- *Dailies*, which includes the coverage of over 500 nationwide and local newspapers, from 2005 to 2009.
- *Twitter*, which is a free social networking and microblogging service. We only have data for 2009.
- *Spinn3r RSS Feeds*, which is a collection of blogs worldwide. We have data for 2008 and 2009.
- *LiveJournal*, which includes all the blogs provided by LiveJournal. We have data from 2006 to 2008.

How *Lydia* sentiment analysis works can be found from (Bautin, Vijayarenu, and Skiena 2008; Godbole, Srinivasaiah, and Skiena 2007). The *Lydia* sentiment data consists of time series of favorable (positive) and unfavorable (negative) words co-referenced with occurrences of each named entity (here denoting companies). Let p and n denote the number of raw positive and negative references, which occurs a total of N times in the corpus (including neutral references). Then we can give below derived measures:

- polarity $= (p - n)/(p + n)$
- subjectivity $= (n + p)/N$

Polarity indicates percentage of positive sentiment references among total sentiment references, while subjectivity indicates proportion of sentiment to frequency of occurrence. These derived measures could provide additional information that raw references cannot.

Media Timing Issues

Proper interpretation of our results requires careful attention to the timing of our news spidering (text retrieval) agents. For the Dailies news corpus we employ, the spidering program begins to download news at 11pm EST every day, a process which can take up to 12 hours. Thus while the bulk of our news was certainly retrieved before the 9:30AM opening of the NYSE each day, we cannot guarantee that it is unpolluted by news reporting events after the market opening. However, for the three blog-type medias (Twitter, Spinn3r, and LiveJournal), accurate time stamp will be provided. Nevertheless, our experimental results show the general consistency across all four corpores.

Correlation of Media / Stock Data

Media Frequency vs. Trading Volume

To compensate for technical variations in spidering efficacy, we use normalized article counts instead of raw article counts to correct for fluctuations in the total volume of news spidered each day. Some significant observations are:

- *Strength of correlation* – For all four depositories, the correlation coefficients between logged normalized article counts and logged stock trading volume are more than 0.4. Therefore, more media coverage lead to more trades.
- *Influence of Market Sectors* – We find that for some sectors, like “Aerospace & Defense”, the correlation is quite strong (corr > 0.7). By contrast, some other sectors, like “Software & Computer Services”, are less sensitive to media exposure (corr < 0.2).
- *Breakdown by Market Capitalization* – Figure 1 shows the breaking down analysis of market capitalizations, which indicates that the correlation coefficients become stronger and stronger with the increasing of market capitalization.

Another similar problem is the relationship between firms’ media references and their corresponding market capitalizations. Our analysis shows that their correlation coefficient is larger than 0.42, and it is statistically significant, which indicates bigger firms receive more media coverage.

Media Polarity vs. Stock Returns

A more interesting question is the return of stocks. We believe the return of stocks is relevant to the public opinion of corresponding firms, say, how good or how bad people think about these firms. Previous section tells us “polarity” is a quantitative term to describe how good a firm is.

Variable Selections We consider three different performance measures for a given stock s: change of stock prices, stock returns ($R(s)$), and abnormal return $R(s) - R(NYSE)$, in which $R(NYSE)$ in the index return of NYSE. In our correlation analysis we correlate each news variable from [polarity, change of polarity, percentage change of polarity], to each stock variable from [change of stock prices, stock return, stock abnormal return]. This gives six combination pairs for testing. Our experiments show (polarity, stock return) pair has the most significant correlations among all the combinations, so we only give the analysis results for polarity versus stock returns later on.

Correlation Analysis with Shifting of Time Figure 4 examines how much today’s polarity is correlated with stock returns on proximate days. For Dailies, we see that (1) the correlation coefficient of today’s polarity versus previous return decrease gradually, and (2) for days 1 and later, all the correlation coefficients are almost zero, and all those correlations are not statistically significant. This proves that today’s news almost have no predictive power for the return of tomorrow or later days. We also notice that the return of day 0 has the best correlation with polarity.

The efficient market hypothesis states that the market reflects public information in the stock price within a very short time. Therefore, Dailies’ polarity shown in Figure 4 illustrated this theory perfectly, i.e., the correlation between news polarity and stock returns disappear after 1 day.

Blogs vs. News Figure 4 also shows Spinn3r data are very similar to Dailies data. However, Twitter is somewhat different in that its polarity also has some correlationship with tomorrow’s or the day after tomorrow’s return. That is, stock market incorporates Twitter sentiment slower than

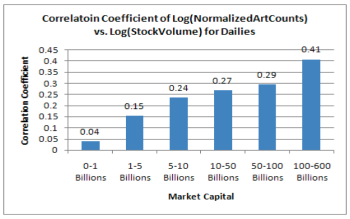

Figure 1: Media Article Counts vs. Stock Trading Volume analysis for Dailies news, broken down by market capitalization.

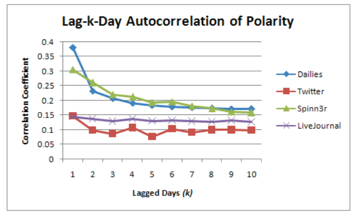

Figure 2: Polarity's Lag-k-Day autocorrelation for Dailies, Twitter, Spinn3r, and LiveJournal respectively.

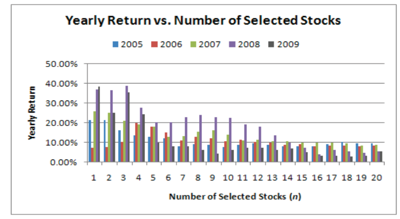

Figure 3: Yearly return vs. number of selected top and bottom stocks for Dailies depository. We tune n from 1 to 20.

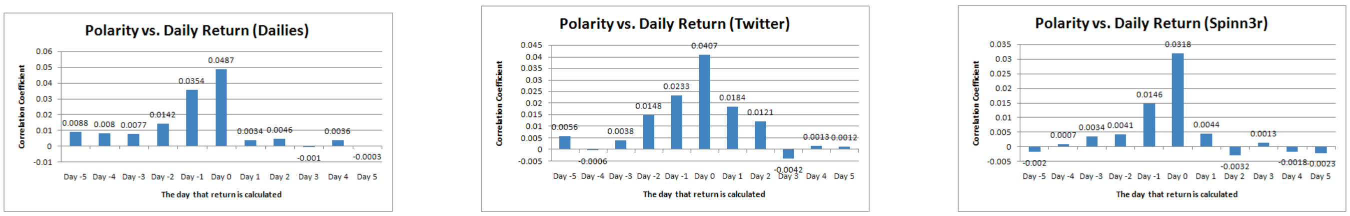

Figure 4: Correlation analysis of Polarity vs. Daily Return for Dailies, Twitter, and Spinn3r respectively. The correlation coefficients are calculated with time lags from -5 to 5 days. Please note: we do not show LiveJournal result here because its data volume is too small.

news, may need 2 or 3 days. In our opinion, there are two possible reasons (1) Our Twitter database only contains data for 0.5 year, and datasize is small and thus the result is less accurate, and (2) The sentiment of Twitter is more persistent between neighbor days.

To dig this problem further, we show polarity's Lag-k-Day autocorrelation for the four media sources in Figure 2. This figure indicates Dailies and Spinn3r have the similar autocorrelation levels, while Twitter and LiveJournal have the lower levels in terms of the strength of autocorrelations. One surprising fact is that all those three blog-type sources have more moderate slopes than Dailies, especially the curves of Twitter and LiveJournal are near flat. This is a very important difference between news and blogs, i.e., the sentiment conveyed by blogs can last longer than news. The underlying reason is that news has more significant recency effect.

Strengthening the Correlation The sentiment-return correlation can be greatly improved by (1) removing companies with the most neutral sentiment; (2) selecting market sectors those are strongly affected by news sentiment; (3) selecting big firms only. Actually, these methods could make the correlation coefficient as strong as 0.2∼0.6. Indeed, our analysis shows that sectors “Life Insurance” and “Financial Services” have the strongest correlation with news sentiment, while sectors “Fixed Line Telecommunications” and “Beverages” have near zero correlation with news sentiment.

Subjectivity vs. Trading Volume

Our analysis over all blog and news sources shows subjectivity is positively and significantly correlated with trading volume. This conclusion coincides with the result from Antweiler and Frank (Antweiler and Frank 2004) that controversial opinions are associated with more trades.

A Sentiment-Based Trading Agent

The Market-neutral Strategy

Now we design a market-neutral trading agent to demonstrate the predictive power of news data. Our market-neutral strategy first ranks companies by their reported sentiment each day, then goes long (short) on equal amounts of positive (negative) sentiment stocks. In this strategy, we identify four tunable parameters that impact our returns substantially:

- n: The number of stocks selected from the top/bottom.
- s: Sentiment analysis window size.
- h: Portfolio holding days.
- C_l and C_u: The lower and upper bounds of market caps.

Performance Evaluation

In this section, we backtest our market-neutral strategy with real blog/news and stock data over a period from 2005 to 2009, and then evaluate returns. Here we performed experiments to isolate one parameter while fixing the other three:

- *Diversification* – Figure 3 shows the impact of the number of selected stocks for Dailies. With the increasing of the number of selected stocks, the return decreases.

  Blogs vs. News – Comparing with news result in Figure 3, Figure 5 shows the performance of blog-type sources, running the same experiments. Spinn3r has similar performance with Dailies, while Twitter and LiveJournal have much lower performance, basically because Twitter and LiveJournal do not have plenty of data. However, all four sources show the same performance trend.
- *Sentiment Analysis Period* – For four of the five years studied (except 2008), yearly returns decrease with the increasing of sentiment analysis period s. This is consistent with the efficient market hypothesis, since longer periods dilute the freshness of the news.

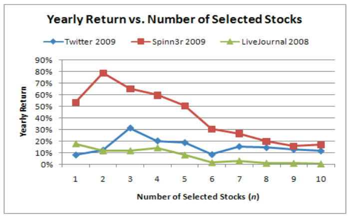

Figure 5: Yearly return vs. number of selected stocks for blog-type depositories.

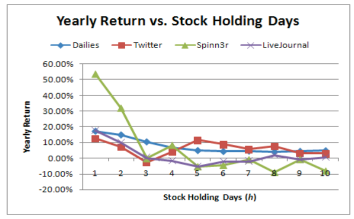

Figure 6: Yearly return vs. stock holding days. Here we use Dailies data for 2006.

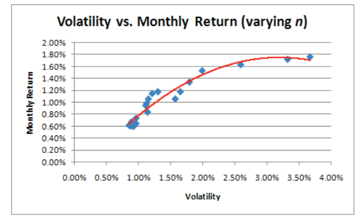

Figure 7: Portfolio Landscape of Monthly Return vs. Volatility analysis for Figure 3.

- *Holding Period* – For all the five years, longer holding time leads to lower returns. Again, the market will quickly reflect all the news information, and thus we will not benefit from extra holding days. Moreover, quickly redeeming the investment frees up capital to invest in more recently reported-on stocks.

  Blogs vs. News – Figure 6 give the comparison of Dailies news and blogs. We notice that with news data, the performance monotonously and gradually decrease with the increase of holding days, but with blogs data, there are much more fluctuations in the performance curves.
- *Market Capitalization* – Our experiments showed an interesting influence of market capitalizations. Both large and small firms showed greater returns than medium-size firms. The return for small firms is enhanced because their price is more affected by news events/sentiment. For large firms, we more accurately measure sentiment due to the greater volume of media coverage.

From these experiments, we conclude that our agent should hold small numbers of selected stocks, use short sentiment-calculation and stock holding periods, and avoid holding medium-sized firms.

Strategy Comparison Above strategy is also called the *best-sentiment strategy*. To further validate the correctness of our sentiment analysis, we design *Worst-sentiment Strategy* and *Random-selection Strategy*, the former does the opposite of *Best-sentiment Strategy*, i.e., long bottom stocks and short top stocks, while the latter just randomly picks stocks to be long and short. Our simulation results show that *worst-sentiment strategy* always produces negative returns and *random-selection strategy* oscillates about zero return.

Long vs. Short Our result shows most profits come from going short in 2008 and from going long in 2009, which due to the collapse of the broad market in 2008 and recovery in 2009, and thus this result perfectly validates the market-neutrality of our strategy.

Returns vs. Volatility Returns only capture part of investment performance. The degree of risk (volatility) taken on to achieve these returns determines to amount of leverage which can safely be employed, and the overall desirability of a given portfolio. In our analysis, we use monthly returns and the standard deviation of monthly returns to measure the risk-return horizon. Figure 7 demonstrates the tradeoff between risk and return, with a scatter plot of performance vs. volatility for strategies differing only in the number of selected stocks. Increased diversification reduces risk. The result is consistent with modern portfolio theory regarding risk and return, therefore investors should choose a diversification level to balance return and the risk he can afford.

Conclusions

We have shown that raw or derived blog/news variables are significantly correlated with some indicators in stock markets. Based on blog/news sentiment data, we design a market-neural strategy, which is able to generate consistent returns for investors. In addition, our analysis also reveals many similar and distinct properties between news and blog sources, such as opinions in blogs are more persistent and decay more gradually over time than news.

References

Antweiler, W., and Frank, M. Z. 2004. Is all that talk just noise? the information content of internet stock message boards. *Journal of Finance* 3:1259–1294.

Bautin, M.; Vijayarenu, L.; and Skiena, S. 2008. International sentiment analysis for news and blogs. In *Second Int. Conf. on Weblogs and Social Media (ICWSM 2008)*.

Chan, W. S. 2003. Stock price reaction to news and no-news: Drift and reversal after headlines. *Journal of Financial Economics* 70:223–260.

Datastream, T. http://www.datastream.com/.

Godbole, N.; Srinivasaiah, M.; and Skiena, S. 2007. Large-scale sentiment analysis for news and blogs. In *Proceedings of the First International Conference on Weblogs and Social Media*, 219–222.

Lloyd, L.; Kechagias, D.; and Skiena, S. 2005. Lydia: A system for large-scale news analysis. In *Proc. 12th String Processing and Information Retrieval (SPIRE 2005)*, volume LNCS 3772, 161–166.

Mittermayer, M., and Knolmayer, G. F. 2006. Text mining system for market response to news: A survey. *Working Paper No 184*.

Pang, B., and Lee, L. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* Vol. 2, No 1-2:1–135.

Tetlock, P. C.; Saar-Tsechansky, M.; and Macskassy, S. 2007. More than words: Quantifying language to measure firms' fundamentals. In *Proceedings of 9th Annual Texas Finance Festival*.