

Evaluating Sentiment in Financial News Articles

Robert P. Schumaker¹, Yulei Zhang², Chun-Neng Huang³ and Hsinchun Chen⁴

¹Management Information Systems

Central Connecticut State University, New Britain, Connecticut 06050, USA

rob.schumaker@gmail.com

²The W.A. Franke College of Business

Northern Arizona University, Flagstaff, Arizona 86011, USA

yulei.gavin.zhang@gmail.com

³Microsoft Corporation

Bellevue, WA 98006, USA

cnhuang@email.arizona.edu and hchen@eller.arizona.edu

⁴Artificial Intelligence Lab, Department of Management Information Systems

The University of Arizona, Tucson, Arizona 85721, USA

cnhuang@email.arizona.edu and hchen@eller.arizona.edu

Word Count: 7,057

Abstract

Can the choice of words and tone used by the authors of financial news articles correlate to measurable stock price movements? If so, can the magnitude of price movement be predicted using these same variables? We investigate these questions using the Arizona Financial Text (AZFinText) system, a financial news article prediction system, and pair it with a sentiment analysis tool. Through our analysis, we found that subjective news articles were easier to predict in price direction (59.0% versus 50.0% of chance alone) and using a simple trading engine, subjective articles garnered a 3.30% return. Looking further into the role of author tone in financial news articles, we found that articles with a negative sentiment were easiest to predict in price direction (50.9% versus 50.0% of chance alone) and a 3.04% trading return. Investigating negative sentiment further, we found that our system was able to predict price decreases in articles of a positive sentiment 53.5% of the time, and price increases in articles of a negative

sentiment 52.4% of the time. We believe that perhaps this result can be attributable to market traders behaving in a contrarian manner, e.g., see good news, sell; see bad news, buy.

Keywords: Business intelligence, text mining, financial prediction, sentiment analysis

1. Introduction

Predicting stock market behavior has always had a certain appeal to researchers. While numerous attempts have been made, the difficulty has always centered around the inability to model the behaviors of human traders. Worse yet, behavioral patterns are constantly changing, making accurate predictions quite difficult. To further add to this uncertainty, there are two entirely opposed philosophies of stock market research; fundamental and technical analysis techniques (Technical Analysis, 2005). Fundamentalists seek to leverage a security's relative data, ratios and earnings, while technicians analyze charts and modeling techniques based on historical trading volume and pricing. The basic problem becomes *does price history matter?*

Within the past several years, the role of computers in stock prediction has increased dramatically. Many of these systems have simply followed the trend of automating existing fundamental and/or technical strategies with the goal of achieving better returns than human traders by removing the elements of emotion and bias from trading (Jelveh, 2006). The downside of these types of systems is that they lack intuition and will continue to execute trades even after unfavorable news events, such as losing a costly court battle. In order to work effectively, these systems require news events to be translated into numeric data before appropriate decisions can be made. This problem introduces serious lag-time into decisions and in some cases human analysts must override trades.

Even with the addition of financial news in a quantitative trading system, other important features of the article may be missing. The addition of author sentiment in stock price prediction could improve accuracy (Das, 2010) and reduce risk (Lu and Chen, 2010). By automatically recognizing not only the terms used in the financial news article, but also the emotional cues used by the author and reflected by their choice of words, we plan to implement a machine learning approach that can learn historical price movements using these features and build a price prediction model in which future news articles could be instantly evaluated.

Our motivation is to build and test such an artifact with an emphasis on the sentiment and author tone features incorporated within a financial news article. By evaluating the intricacies of article sentiment and author tone, we seek to discover the role author sentiment can play in price prediction.

This paper is arranged as follows. Section 2 provides an overview of literature concerning Stock Market prediction, textual representations and sentiment analysis techniques. Section 3 develops our research questions. Section 4 describes our approach to the problem and introduces the Arizona Financial Text (AZFinText) system. Section 5 gives an overview of our experimental design. Section 6 provides results of our experimental findings and a discourse of their impact on stock price prediction. Section 7 presents our conclusions and a brief discussion of future research directions.

2. Literature Review

There are two theories that have had a significant impact on market prediction, Efficient Market Hypothesis (EMH) and Random Walk Theory (RWT). In EMH, the price of a security is a reflection of complete market information and when new information is introduced, the market instantly adjusts the stock price to reflect it (Fama, 1964). From EMH theory, it is believed that

markets are efficient and that price corrections occur instantly, making price prediction from market data impossible.

RWT is similar to EMH where all information is embedded in the current price and cannot be practically used for future prediction. This theory slightly differs from EMH in that under RWT, short-term price movements are viewed as indistinguishable from random activity (Malkiel, 1973). It is believed that this short-term random activity produces unpredictable near-term price movements and thus the market's non-deterministic nature makes it impossible to consistently outperform it.

Furthermore, the ability to scrutinize the decisions of traders and uncover trading behavior on the scale of a market exchange is an extremely difficult task. To lessen this problem and test the impact of both fundamental and technical trading strategies, LeBaron created an artificial stock market of simulated traders whose trading decisions could be manipulated and dissected (LeBaron, Arthur et al., 1999). LeBaron accomplished this by introducing new pieces of information into the market and adjusted the amount of decision time between when an individual trader would receive new information and then act upon it. It was discovered that traders with longer waiting times formed fundamental strategies, relying more heavily on company-specific performance data, while those with shorter waiting times developed technical strategies, such as timing trades to base their trading decisions upon. This study led to a more important contribution by finding that a lag period existed between the time information is introduced and when the market would correct itself. This apparent delay in market behavior helped to dismiss the instantaneous market correction theories and lent support to the idea that markets could be forecast for short durations of time following the introduction of new information. To follow up on establishing limits to this predictive window, Gidofalvi discovered

that there exists a twenty minute window of opportunity before and after a financial news article is released (Gidofalvi, 2001). It is within this window that weak prediction of a stock price is possible. It is further believed that this period occurs for several reasons. The predictability of price twenty minutes before an article is released is believed to account for the activity of insider traders whereas the twenty minutes following an article release is the result of reprinting and posting delays that exist across web media, where some traders may receive their information later than others (Gidofalvi and Elkan, 2003).

2.1 Financial News Articles

Markets thrive on new information and this data is constantly streaming. While many information types can move a stock price, e.g., rumors, eavesdropping and scandals; financial news articles are considered to be a more stable and trustworthy source. Many studies rooted in the Information Retrieval (IR) methodologies have examined the use of textual news articles. While most are interested in studying the relationships between articles in a pure IR point of view, few focus on the domain of financial text as it relates to quantifiable stock market data. In Cecchini's work, he examined the role that financial text may play in predicting corporate fraud or bankruptcy (Cecchini, Aytug et al., 2010). He did this by looking at the terms used in corporate 10K reporting and linked it to historical financial results. Through this pairing of text and quantitative activity, Cecchini was able to predict bankruptcy 83.9% of the time and fraud with 82.0% accuracy.

In a different type of study, Ma examined the textual role that financial news articles may play in determining levels of company cooperation and connectedness amongst one another (Ma, Sheng et al., 2009). This study constructed a social network of companies mentioned in financial

news articles and built links between co-mentioned organizations. The strength of the links can provide an idea of the amount of connectedness between companies.

The act of correlating financial news articles to a quantifiable price movement is a difficult task, even when the information within a financial news article can have a visible impact on price (Wuthrich, Cho et al., 1998; Lavrenko, Schmill et al., 2000a; Gidofalvi, 2001; Mittermayer, 2004). However, financial news articles are not the sole determiner of price movement. Sudden price movements can still occur from other sources, such as large unexpected trades (Camerer and Weigelt, 1991).

The first challenge of a textual financial prediction system is to manage the large amounts of textual information that exist for securities. This material can include required reports such as periodic SEC filings, press releases and financial news articles reporting both unexpected events and routine news alike. These textual documents can then be parsed using Natural Language Processing (NLP) techniques to identify specific article terms or phrases most likely to cause dramatic share price changes. Terms such as “factory exploded” or “workers strike” may indicate a price plunge in the near future. By automating this process, machines can map financial text terms to discrete price movements and identify arbitrage opportunities faster than a human counterpart. This leads to the next logical step where a purpose-built quantitative system could also be put in charge of executing immediate trades based on textual data.

Thus the process of obtaining timely financial documents from reputable Web sources is a critical step. There are a variety of financial news aggregation sites that provide this service. One of these sites is Comtex which offers real-time financial news in a subscription format. Another source is PRNewsWire, which offers free real-time and subscription-based services. Yahoo! Finance is a third such source and is a compilation of 45 different news sources

including the Associated Press, Financial Times and PRNewsWire among others. This source provides a variety of perspectives and timely news stories regarding financial markets.

2.2 Textual Representation

Once financial news articles have been gathered they need to be represented in machine-friendly form. There are multiple techniques available, the first of which is to tokenize the article and use each token as a feature. However, this technique is both noisy and has scalability issues from rarely used features. Furthermore, some repetitive features can detract from a prediction algorithm's accuracy. Features such as "http:" or semantically empty stopwords only add to the scalability overhead. There are a variety of ways to combat these problems which will be discussed throughout this section. The most popular way to represent text is to isolate particular features in an article, such as writing style, parts of speech or even author sentiment.

2.2.1 Writing Style

Writing style analysis is a useful representation in document comparisons where the goal is to identify an unknown author amidst a corpus of known authors. Techniques include measures of similarity between documents and focusing on word choice and syntax usage (Abbasi and Chen, 2008).

2.2.2 Parts of Speech

In a parts of speech representation, articles are represented by particular syntactic subsets, such as nouns or verbs. The most popular of which is the Bag of Words (BOW) approach which has been used extensively in textual financial research (Lavrenko, Schmill et al., 2000a; Gidofalvi, 2001). This approach is similar to tokenization, except that stopwords such as conjunctions and declaratives are removed from the representation. Better versions of BOW will also employ some form of stemming where predicates are stripped and the root term

representation remains. However, the BOW approach still has noise-related issues associated with infrequent terms as well as problems of scalability. An improved representational system is Noun Phrases which retains only the nouns and noun phrases from a document and can adequately represent important article concepts (Tolle and Chen, 2000). As a result, this technique uses fewer terms leading to better article scaling. A third representational technique is Named Entities, which is an extension of Noun Phrases. It functions by selecting the proper nouns of an article that fall within well-defined categories. This process uses a semantic lexical hierarchy (Sekine and Nobata, 2004) as well as a syntactic/semantic tagging process (McDonald, Chen et al., 2005) to assign candidate terms to pre-defined categories. Named Entities allows for better generalization of previously unseen terms and does not possess the scalability problems associated with a semantics-only approach. A fourth representational technique is Proper Nouns. This method functions as an intermediary between Noun Phrases and Named Entities. Proper Nouns is a subset of Noun Phrases by selecting specific nouns and a superset of Named Entities without the constraint of pre-defined categories. In a comparison study using these four representational techniques, it was found that the Proper Noun representation was more effective in representing financial news articles (Schumaker and Chen, 2009).

2.2.3 Author Sentiment

In general, sentiment analysis is concerned with the analysis of direction-based text and attempts to determine whether the text is objective or subjective, as well as whether the subjective parts contain either positive or negative sentiments. This classification into positive and negative sentiments is a common two-class problem (Pang, Lee et al., 2002; Turney, 2002). Additional variations include classifying sentiments as opinionated/subjective or factual/objective (Wiebe, Wilson et al., 2004). Some studies have attempted to classify

emotions, including happiness, sadness, anger, horror, etc., instead of sentiments (Subasic and Huettner, 2001; Grefenstette, Qu et al., 2004; Mishne, 2005).

Building upon these sentiment and tonal classifications, (Li and Wu, 2009) developed an SVM clustering technique which found that popular chat forum topics all shared certain sentiment values and that it was possible to predict future topical hotspots based on their similarity to the sentiment characteristics. While our research does not investigate topical similarity, the same underlying methods can be used to derive stock price movement, or online product sales, such as financial press releases and financial news (Davis, Piger et al., 2006; Devitt and Ahmad, 2007; Tetlock, 2007), or user generated content (Antweiler and Frank, 2004; Das and Chen, 2007; Ghose, Ipeirotis et al., 2007). From these investigations, it was discovered that negative sentiment may be predictive of future downward moves in firm value (Tetlock, 2007) and that positive and negative polarity in financial news is consistent with human judgment (Devitt and Ahmad, 2007) and has an impact on firm performance (Davis, Piger et al., 2006; Das and Chen, 2007; Ghose, Ipeirotis et al., 2007).

One well-known and tested tool to measure an article's sentiment is OpinionFinder. This tool aims to identify subjective sentences and classify them as either positive or negative sentiments (Wilson, Hoffmann et al., 2005a). It was developed by Wiebe's group based on a series of publications, such as the subjective sentence classifier (Riloff and Wiebe, 2003; Wiebe and Riloff, 2005), and the polarity classifier (Wilson, Wiebe et al., 2005b). It also has relatively good performance compared against the MPQA Opinion Corpus, with an accuracy of 74%, subjective precision of 78.4%, subjective recall of 73.2% and a subjective F-measure of 75.7% as compared to baseline accuracy of 55.3%.

2.3 Machine Representation

Machine learning algorithms are unable to process raw text or sentiment representations and require an additional layer of representation. One popular method is to represent article terms in binary where the term is either present or not in a given article (Joachims, 1998). This binary solution is typically implemented using a massive matrix where articles comprise one axis and the other axis is all of the terms in the corpora and other article-specific features. It leads to large but sparse matrices where the number of represented terms throughout the dataset will greatly outnumber the number of terms used in an individual article.

Once financial news articles have been represented, learning algorithms can then begin to identify patterns of behavior and build prediction models from them. One commonly accepted method, Support Vector Regression (SVR), is a regression equivalent of Support Vector Machines (SVM) but without the aspect of classification (Vapnik, 1995). Like SVM, SVR attempts to minimize its fitting error while maximizing its goal function by fitting a regression estimate through a multi-dimensional hyperplane. This method is also well-suited to handling textual input in binary and has been used in similar financial news studies (Tay and Cao, 2001; Schumaker and Chen, 2009).

3. Research Questions

From these gaps, we have formulated several research questions. The first of which is:

- Does Objectivity/Subjectivity impact news article prediction?

While we know from the literature that the tone of an article can correlate to future price direction, can the addition of objectivity and subjectivity provide improved discrete prediction?

As a follow-up to this question, we also ask:

- Does Positive/Negative Subjectivity impact news article prediction?

From the literature, we know that the subjectivity of an article will impact the share price in the same direction over the long term, but shorter periods of time have not been examined and could perhaps lead to interesting results.

4. System Design

In order to evaluate our research questions, we designed the AZFinText system. Figure 1 illustrates the basic design.

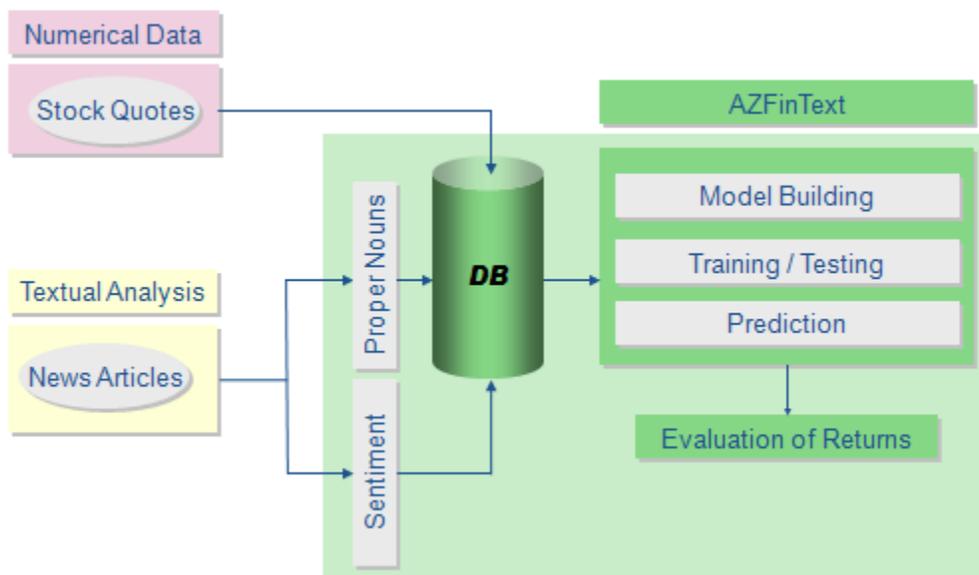


Figure 1. The AZFinText System

From the AZFinText system design in Figure 1, there are several major components to describe in detail. The first component is Numerical Data that gathers stock price data in one minute increments from a commercially available stock price database. The second component is Textual Analysis. This component gathers financial news articles from Yahoo! Finance and represents them by their proper nouns as well as by the sentiment of the article. This module further limits extracted features to three or more occurrences in any document, which cuts down the noise from rarely used terms (Joachims, 1998).

Once data is gathered, AZFinText makes price predictions, twenty minutes after the news article has been released, for each financial news article. This 20min period was selected because it follows the research of Gidofalvi who observed a twenty minute window of weak predictability (Gidofalvi, 2001). From prior empirical testing, it was found that including the proper noun representation and the stock price at the time the news article was released, provided AZFinText with superior predictive performance compared to other textual representations and different pieces of price information (Schumaker and Chen, 2009).

At the Model Building stage, we partitioned the data into three models. The first of which is the AZFinText system without sentiment information. This model only incorporates the proper nouns of the article and the price of the stock at the time the article was released. The second model of Tone is AZFinText plus three sentiment features of objective, subjective and neutral. We then used the OpinionFinder tool to make a determination of the overall tone of the article, i.e., is the article more objective or more subjective. In cases of a tie between objective and subjective, the article was marked neutral. The third model of Polarity is AZFinText plus three subjective sub-features of positive, negative and neutral. Even though an article may be marked as objective, we felt that the minority subjective aspects may make a valuable contribution to price prediction and should certainly be tested. As an example of the subjective and objective aspects of the OpinionFinder tool, we present two financial news articles, one marked subjective and the other objective.

Article marked Subjective

Baby Talk at Coach by Alyce Lomax (Dec. 5, 2005)

I have to admit: I chuckled in amazement when I saw that Coach (NYSE: COH) offers a line of baby products. I had a hard time dealing with the thought of tiny Coach purses, tiny Coach leather gloves and other miniaturized Coach wares. (And given the Coach brand name, it's guaranteed that all of this tiny stuff will bear grown-up price tags.)

Coach certainly does offer cashmere mittens, blankets, and teddy bears for infants and babies. However, the vast majority of the product line includes leather photo albums enamel frames and of course what might be seen as a must-have for the well-heeled Mom a Coach baby bag. When it comes to designer apparel and accoutrements for a baby I would imagine that lots of us would most certainly blink an eye at say the Coach pom-pom cap for baby that costs \$98 (as do the mittens mentioned above). The product line includes a \$398 diaper-and-associated-baby-necessities bag, a \$58 teddy bear, and last but not least, a \$48 key fob. (Ahem!)

Many moms certainly go for less expensive baby gear -- Target (NYSE: TGT) or Gap's (NYSE: GPS) Old Navy are big destinations for parents who juggle the needs of a newborn in a more frugal manner. I know many moms who are of the opinion that spending too much money on gear is silly when baby grows out of everything quite soon enough. (Not to mention, infants certainly don't distinguish according to brand -- if they did, parents would be in big trouble!) Regardless, though there are plenty of people who prefer luxury or aspirational brands -- Coach is a popular one as is Tiffany (NYSE: TIF).

It may have seemed silly to me at first blush, but Coach has been doing quite well as an investment over the years, having successfully attracted a cross-section of shoppers in different income brackets many of whom see the brand as an affordable luxury. Meanwhile it is the holiday season when doting grandparents and other family members are gearing up to find the perfect gift for special little someones everywhere. Coach for baby might seem like madness to some of us -- but it seems more than likely there's great method to the madness as the holidays approach. (Credit: Motley Fool)

Article marked Objective

Procter increase in prices products pays off (Nov. 1, 2005)

Procter & Gamble Co.'s quarterly profit rose 4.5 percent the company announced Tuesday after price increases and new product additions. Net income in the first quarter at Procter increased to \$2.03 billion or 77 cents a share on a sales gain of 7.6 percent.

Procter raised prices on Pampers diapers and Cascade dishwashing detergent and sales of new products such as Tide with Febreze scent increased. The company was trying to recover cost increases of raw materials including oil and resin. Procter was forecast to earn 76 cents a share according to the average estimate of 12 analysts surveyed by Thomson Financial.

Procter is adding health and beauty items to grab market share from Kimberly-Clark Corp. The company had announced earlier this year increases on Ivory soap Cascade dishwashing detergent and Gain laundry detergent to try to recover some of the higher costs of raw materials including resin. So far the majority of the company's price increases have been accepted by retail and passed on to the end consumer Christopher Ferrara an analyst at Merrill Lynch & Co. in New York wrote in a report. Procter also expanded sales in China and other

developing markets which account for about 23 percent of Procter's annual revenue.

Cincinnati-based Procter and Gamble (NYSE: PG) has a product line-up that includes Crest toothpaste and Pampers diapers. (Credit: Cincinnati Business Courier)

For the machine learning algorithm we chose to implement the SVR Sequential Minimal Optimization (Platt, 1999) function through Weka (Witten and Eibe, 2005). This function allows discrete numeric prediction instead of classification. We selected a linear kernel and ten-fold cross-validation. A similar prediction method was employed in the forecasting of futures contracts (Tay and Cao, 2001).

AZFinText is then trained on the data and performs price predictions for each financial news article encountered. Evaluations are then made regarding the effect of predicted stock price in terms of the models generated.

5. Experimental Design

For the experiment, we selected a consecutive five week period of time to serve as our experimental baseline. This period of research was from Oct. 26, 2005 to Nov. 28, 2005 and incorporates twenty-three trading days. The five-week period of study was selected because it gathered a comparable number of articles to prior studies: 6,602 for Mittermayer (Mittermayer, 2004) and 5,500 for Gidofalvi (Gidofalvi, 2001). We also observe that the five-week period chosen did not have unusual market conditions and was a good testbed for our evaluation. In order to identify the companies with the most likelihood of having quality financial news, we limited our scope of activity to only those companies listed in the S&P 500 as of Oct. 3, 2005. Articles gathered during this period were restricted to occur between the hours of 10:30am and 3:40pm. Even though trading starts at 9:30am, we felt it important to reduce the impact of overnight news on stock prices and selected a period of one-hour to allow these prices to adjust.

The 3:40pm cut-off was selected to disallow any +20 minute stock predictions to occur after market hours. A further constraint to reduce the effects of confounding variables was introduced where two articles on the same company cannot exist within twenty minutes of each other or both will be discarded. The above processes filtered the 9,211 candidate news articles gathered during this period to 2,802, where the majority of discarded articles occurred outside of market hours. Similarly, 10,259,042 per-minute stock quotations were gathered during this period. This large testbed of time-tagged articles and fine-grain stock quotations allow us to perform a systematic evaluation.

AZFinText's predictions were then analyzed against a three metric evaluation of Closeness, Directional Accuracy and a simple Trading Engine. Closeness, or how close AZFinText's predicted +20min value was to the actual +20min price, is measured in terms of Mean Squared Error (MSE) where $MSE = (1/n)\Sigma(\text{Predicted} - \text{Actual})^2$ (Cho, Wuthrich et al., 1999). Directional Accuracy is simply how often AZFinText was correct in predicting the price direction of the +20min stock (Gidofalvi, 2001). For a Trading Engine, AZFinText utilized a modified version of Lavrenko's Trading Engine (Lavrenko, Schmill et al., 2000a) that examines the percentage return of the stock. When a stock demonstrates an expected movement exceeding 1%, then \$1,000 worth of that stock is either bought or shorted and then disposed of after twenty minutes. This modified version differs slightly from Lavrenko's original design, where Lavrenko traded in blocks of \$10,000 instead of \$1,000. We further assume zero transaction costs, consistent with Lavrenko.

An example of AZFinText in operation is shown in Figure 2.

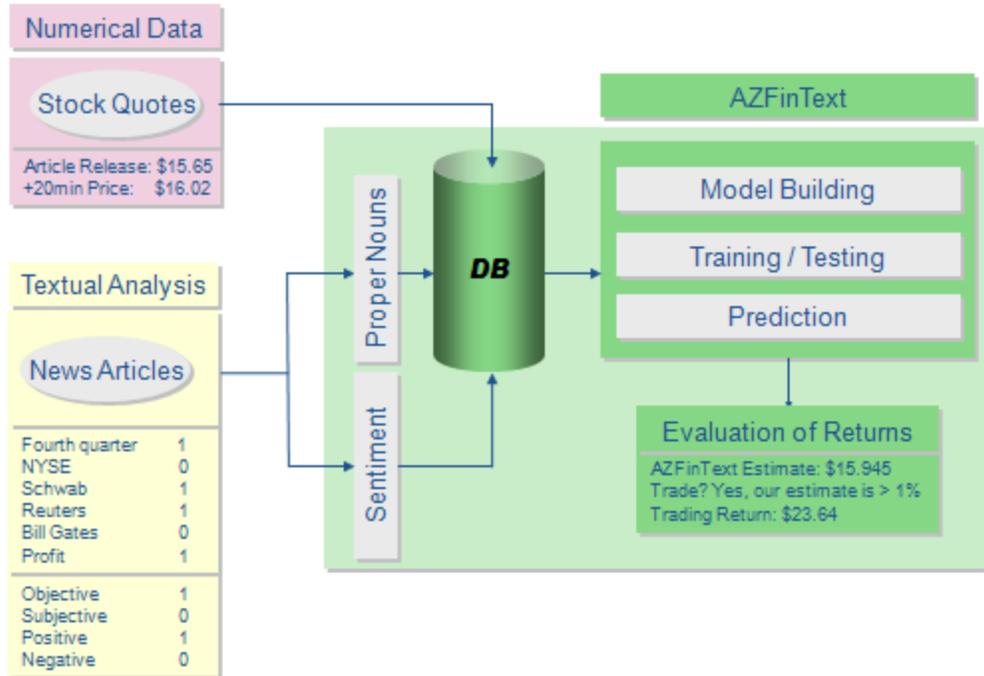


Figure 2. AZFinText Textual Example

The first task is to extract financial news articles. The entire corpus of financial news articles are represented by their Proper Nouns in binary. If a particular Proper Noun feature is present in the article, that feature is given a 1, else a 0 and then stored in the database. Similarly, each financial news article is evaluated by OpinionFinder to identify its overall tone and polarity. In tandem, stock quotations gathered on a per minute basis and stored. To build a model, we first pair together the representational Proper Nouns and stock quotation at the time the article was released, for each financial news article. Then, depending upon the particular model that is tested, data is aggregated and passed to our machine learning component for training and testing. Stock price predictions are then made for each financial news article and passed along to the evaluation instruments.

From the example above, AZFinText derived a prediction price of \$15.945 which is greater than 1% of the stock price at the time the article was released, \$15.65. Our trading engine makes a trade and sells twenty minutes later, for a trade return of \$23.64 or 2.36%.

6. Experimental Results

To answer our first research question of *does Objectivity/Subjectivity impact news article prediction*, we tested AZFinText (Baseline), against Tone (AZFinText plus binary representations of objective, subjective and neutral). We further broke apart Tone into individual Objective, Subjective and Neutral components to further investigate Tone’s impact, as shown in Table 1.

	Baseline	Tone	Objective	Subjective	Neutral
# Articles	2,802	2,802	2,662	61	79
Closeness	0.0516	0.0565	0.0544	0.103	0.0930
Direction	50.4%	49.8%	49.5%	59.0%	53.2%
Trading	2.41%	2.00%	2.03%	3.30%	0.42%

Table 1. Article Tone results (all p-values versus Baseline < 0.05)

Tone performed poorly against Baseline in all three metrics. Although it may seem that the addition of sentiment variables harmed AZFinText’s predictive capability, when Tone was subdivided into its three constituent parts of Objective, Subjective and Neutral, it became apparent that several of Tone’s components were depressing its overall score. Objective articles were performing poorly in Directional Accuracy versus Baseline (49.5% to 50.4% respectively) and Neutral articles had poorer Trading Returns versus Baseline (0.42% to 2.41% respectively). By contrast, Subjective articles performed better with 59.0% Directional Accuracy and a 3.30% Trading Return. From these results, Baseline, which did not include any sentiment analysis in its model, had the best Closeness score of 0.0516 while subjective articles had the best Directional Accuracy (59.0%) and Trading Return (3.30%). This would imply that the author’s use of subjectivity in financial news articles demonstrated marked price movement immediately following article release. These results were all significant as all values versus Baseline’s values had p-values < 0.05. Digging further into the subjective trading returns, of the 61 opportunities

to trade, there were 10 instances where AZFinText executed a trade for an average return of \$3.30 (standard deviation of \$7.92). The maximum return was \$390 and the minimum was -\$60.

To answer our second research question of *does Positive/Negative subjectivity impact news article prediction*, we tested AZFinText (Baseline) against Polarity, which was AZFinText plus the binary representations of positive, negative and neutral. We further broke apart Polarity into its constituent components of Positive, Negative and Neutral to further investigate Polarity’s impact, as shown in Table 2.

	Baseline	Polarity	Positive	Negative	Neutral
# Articles	2,802	2,802	619	1,077	1,106
Closeness	0.0516	0.0556	0.0521	0.0576	0.0557
Direction	50.4%	49.4%	50.1%	50.9%	47.6%
Trading	2.41%	2.29%	1.73%	3.04%	1.98%

Table 2. Article Polarity results (all p-values versus Baseline < 0.05 except Positive Closeness)

Polarity performed poorly versus Baseline in all three metrics (0.0556 to 0.0516 in Closeness, 49.4% to 50.4% in Directional Accuracy and 2.29% to 2.41% in Trading Returns). We suspect that the addition of the polarity variables was detrimental to AZFinText’s predictions. However, by breaking apart Polarity into its component pieces, interesting results occurred. As shown in Table 2, Baseline again performed best in measures of Closeness (0.0516). Negative subjective articles performed best in both Directional Accuracy (50.9%) and Trading Returns (3.04%). We believe that this may be a psychological reflection of market dynamics because negative emotions can have a larger and more lasting impact than positive or neutral ones. These results were all significant as all values versus Baseline’s values had p-values < 0.05 except for Positive Closeness, which was statistically equivalent to Baseline. Looking closer at the negative polarity results, of the 1,077 opportunities to execute a trade,

AZFinText traded in 82 of them for an average return of \$3.04 (standard deviation of \$22.30). The maximum return was \$610 and the minimum was -\$250.

To pursue why negative articles were easier to predict, we looked at the three component pieces of Polarity (Positive, Negative and Neutral) in terms of AZFinText’s ability to correctly predict price direction, as shown in Table 3.

Correct Predictions	Positive	Negative	Neutral
Upswings	46.0%	52.4%	49.5%
Downswings	53.5%	49.1%	46.0%

Table 3. Article Polarity versus Directional Accuracy (all p-values < 0.05)

From this table, AZFinText worked best at predicting downswings of price in Positive Polarity articles (53.5%) and price upswings on both Negative and Neutral Polarity articles (52.4% and 49.5% respectively). All p-values < 0.05. AZFinText exhibited seemingly counter-intuitive results where positive articles were easier to predict price decreases whereas negative and neutral articles were easier to predict price increases. We believe that perhaps this result can be attributable to market traders behaving in a contrarian manner, e.g., see good news, sell... see bad news, buy.

An example of this contrarian behavior is shown in the following PRNewsWire article, *Goodyear, Tire Industry Association Join Forces on Certification*, in which Goodyear describes their training certification program.

The industry-first collaboration between a tire company and the international industry trade association includes two one-day seminars in 48 cities next year. Tire technicians from the Goodyear Dunlop and Kelly dealer network as well as Wingfoot Commercial Tire Centers are eligible for the Goodyear/TIA joint certification. Steve McClellan Goodyear's vice president of commercial tire systems said the customized training combines TIA's commercial tire service program and Goodyear's hands-on experience. TIA's certification process offers technicians valuable information that will help them perform their jobs better and with more care particularly as we embark on an aggressive campaign to expand our service business through a strong network of tire servicing outlets. McClellan said Goodyear seeks to improve its service business revenue which is more stable than cyclical commercial tire sales. Knowledgeable tire technicians across more than 2,000 commercial tire centers will deliver unrivaled service to our customers in their quest to lower costs. The service equation becomes a win-win for Goodyear and its dealer

network. Al Cohn manager of strategic initiatives for Goodyear's commercial tire systems said dealer training is a major initiative. Our vision is to do more than just manufacture quality tires. We want to create business solutions that are measurable, repeatable and sustainable. That means helping dealers to work with fleets to manage their tire costs from original equipment to replacement and retreads and delivering service and value along the way. In addition the course exceeds OSHA training requirements for improved safety awareness. And that's why Goodyear/TIA technician certification makes sense. Through this collaboration the training provides a competitive advantage for independent dealers to grow their tire and service business, he said. Goodyear/TIA certification also may reduce dealer workers compensation costs, Cohn added. Goodyear commercial tire systems offer complete products and services to the trucking industry including a full range of original equipment and replacement tires. In addition the company's cradle-to-grave tire and service network includes retreading tire management tools and business solutions for tomorrow's trucking fleets. For more information on Goodyear's line of commercial tires, go to <http://www.goodyear.com/truck>

This article is a rather normal public relations piece that was marked with positive polarity and had a stock price of \$15.39 when released. AZFinText predicted the +20min stock price to decline to \$15.136 and Goodyear's actual +20min stock price was \$15.22, a clear decline. Since the price was stable prior to article release, it would appear contrarian trading was taking place. While it could be possible that some external event occurred simultaneously that forced the price downward, AZFinText found a statistically sufficient number of these declines occurring that if this trend were noticed by the company, they may have done far better managing their share price to not release any news articles during this time, positive or otherwise.

7. Conclusions

From our investigation we found several interesting results. The first of which was that AZFinText was best able to predict Subjective articles in Directional Accuracy (59.0% to 50.4%) and Trading Returns (3.30% to 2.41%), but not Closeness (0.103 versus 0.0516). We felt that the subjectivity of the articles may have had an impact on trading behavior. The second notable result was that AZFinText was best able to predict Negative Subjective articles in Directional Accuracy (50.9% to 50.4%) and Trading Returns (3.04% to 2.41%), but not Closeness (0.0576 versus 0.0516). We believe that these results are attributable to investors reacting more strongly

to negative articles, which further adds to the ideas of Devitt and Ahmad (2007). The third notable result was that AZFinText found evidence of Contrarian trading activity. AZFinText was better able to predict downswings in Positive articles (53.5%) and upswings in Negative and Neutral articles (52.4% and 49.5% respectively). This stands in contrast to the work of Tetlock (2007) whom observed that negative sentiments should be indicative of downward price movement. It is possible that this work and that of Tetlock observed differing snapshots of market trading activity that led to differing results. Clearly more research is needed to determine the full extent of market trading behavior on mechanized prediction.

We would also suggest several future directions for this area of research. The first of which is to investigate the role of verbs and adverbs as a textual representation method. Perhaps this representational scheme will lead to better predictivity than Proper Nouns alone. It would be beneficial in future studies to draw comparisons between various textual representations. The second future direction would be to investigate other machine learning techniques. While SVR has proven itself in the textual financial domain, perhaps other techniques could identify different types of hidden market patterns. The third future direction would be to explore the role of negation in the OpinionFinder tool and its impact on price direction. While OpinionFinder has satisfactory accuracy at 74%, an unaccounted for negation could introduce a small amount of noise into the results. It would be important to determine its overall impact before constructing a real-time trading system.

References

- Abbasi, A. and H. Chen (2008). Writeprints: A Stylometric Approach to Identify-level Identification and Similarity Detection in Cyberspace. *ACM Transactions on Information Systems* 26(2).
- Antweiler, W. and M. Frank (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance* 59(3): 1259-1294.
- Camerer, C. and K. Weigelt (1991). Information Mirages in Experimental Asset Markets. *Journal of Business* 64(4): 463-493.
- Cecchini, M., H. Aytug, et al. (2010). Making Words Work: Using Financial Text as a Predictor of Financial Events. *Decision Support Systems* doi: 10.1016/j.dss.2010.07.012.
- Cho, V., B. Wuthrich, et al. (1999). Text Processing for Classification. *Journal of Computational Intelligence in Finance* 7(2).
- Das, S. (2010). The Finance Web: Internet Information and Markets. *IEEE Intelligent Systems* 25(2).
- Das, S. and M. Chen (2007). Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. *Management Science* 53(9): 1375-1388.
- Davis, A., J. Piger, et al. (2006). Beyond the numbers: an analysis of optimistic and pessimistic language in earnings press releases. Technical Report, Federal Reserve Bank of St. Louis.
- Devitt, A. and K. Ahmad (2007). Sentiment Polarity Identification in Financial News: A Cohesion-Based Approach. *Association of Computational Linguistics*, Prague, Czech Republic.
- Fama, E. (1964). The Behavior of Stock Market Prices. *The Journal of Business* 38(1): 34-106.
- Ghose, A., P. Ipeirotis, et al. (2007). Opinion Mining Using Econometrics: A Case Study on Reputation Systems. *Association of Computational Linguistics*, Prague, Czech Republic.
- Gidofalvi, G. (2001). Using News Articles to Predict Stock Price Movements. Department of Computer Science and Engineering, University of California, San Diego.
- Gidofalvi, G. and C. Elkan (2003). Using News Articles to Predict Stock Price Movements. Technical Report. Department of Computer Science and Engineering, University of California, San Diego.
- Grefenstette, G., Y. Qu, et al. (2004). Coupling Niche Browsers and Affect Analysis for an Opinion Mining Application. *7th International Conference on "Recherche d'Information Assistee par Ordinateur"*, Avignon, France.
- Jelveh, Z. (2006). How a Computer Knows What Many Managers Don't. The New York Times.

- Joachims, T. (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features. *European Conference on Machine Learning*, Chemnitz, Germany.
- Lavrenko, V., M. Schmill, et al. (2000a). Language Models for Financial News Recommendation. *International Conference on Information and Knowledge Management*, Washington, DC.
- LeBaron, B., W. B. Arthur, et al. (1999). Time Series Properties of an Artificial Stock Market. *Journal of Economic Dynamics and Control* 23(9-10): 1487-1516.
- Li, N. and D. D. Wu (2009). Using Text Mining and Sentiment Analysis for Online Forums Hotspot Detection. *Decision Support Systems* 48: 354-368.
- Lu, H.-M. and H. Chen (2010). Financial Text Mining: Supporting Decision Making Using Web 2.0 Content. *IEEE Intelligent Systems* 25(2).
- Ma, Z., O. Sheng, et al. (2009). Discovering Company Revenue Relations from News: A Network Approach. *Decision Support Systems* 47: 408-414.
- Malkiel, B. G. (1973). A Random Walk Down Wall Street. New York, W.W. Norton & Company Ltd.
- McDonald, D. M., H. Chen, et al. (2005). Transforming Open-Source Documents to Terror Networks: The Arizona TerrorNet. *American Association for Artificial Intelligence Conference Spring Symposia*, Stanford, CA.
- Mishne, G. (2005). Experiments with Mood Classification in Blog Posts. *SIGIR*, Salvador, Brazil.
- Mittermayer, M. (2004). Forecasting Intraday Stock Price Trends with Text Mining Techniques. *Hawaii International Conference on System Sciences*, Kailua-Kona, HI.
- Pang, B., L. Lee, et al. (2002). Thumbs up?: Sentiment classification using machine learning techniques. *Association for Computational Linguistics*, Philadelphia, PA.
- Platt, J. C. (1999). Fast Training of Support Vector Machines using Sequential Minimal Optimization. Advances in Kernel Methods: Support Vector Learning. B. Scholkopf, C. Burges and A. Smola. Cambridge, MA, MIT Press: 185-208.
- Riloff, E. and J. Wiebe (2003). Learning Extraction Patterns for Subjective Expressions. *Conference on Empirical Methods in Natural Language Processing*, Sapporo, Japan.
- Schumaker, R. P. and H. Chen (2009). Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFinText System. *ACM Transactions on Information Systems* 27(2).
- Sekine, S. and C. Nobata (2004). Definition, Dictionaries and Tagger for Extended Named Entity Hierarchy. *Language Resources and Evaluation Conference*, Lisbon, Portugal.

- Subasic, P. and A. Huettner (2001). Affect Analysis of Text Using Fuzzy Semantic Typing. *IEEE Transactions on Fuzzy Systems* 9(4): 483-496.
- Tay, F. and L. Cao (2001). Application of Support Vector Machines in Financial Time Series Forecasting. *Omega* 29: 309-317.
- Technical Analysis. (2005). The Trader's Glossary of Technical Terms and Topics. Retrieved Mar. 15, 2005, 2005, from <http://www.traders.com>.
- Tetlock, P. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance* 62(3): 1139-1168.
- Tolle, K. M. and H. Chen (2000). Comparing Noun Phrasing Techniques for Use with Medical Digital Library Tools. *JASIS* 51(4): 352-370.
- Turney, P. D. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Association for Computational Linguistics (ACL'02)*, Philadelphia, PA.
- Vapnik, V. (1995). The Nature of Statistical Learning Theory. New York, Springer.
- Wiebe, J. and E. Riloff (2005). Creating Subjective and Objective Sentence Classifiers from Unannotated Texts. *Sixth International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City, Mexico.
- Wiebe, J., T. Wilson, et al. (2004). Learning Subjective Language. *Computational Linguistics* 30(3): 277-308.
- Wilson, T., P. Hoffmann, et al. (2005a). OpinionFinder: a system for subjectivity analysis. *Human Language Technology Conference*, Vancouver, Canada.
- Wilson, T., J. Wiebe, et al. (2005b). Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. *Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, Vancouver, Canada.
- Witten, I. H. and F. Eibe (2005). Data Mining: Practical Machine Learning Tools and Techniques. San Francisco, Morgan Kaufmann.
- Wuthrich, B., V. Cho, et al. (1998). Daily Stock Market Forecast from Textual Web Data. *IEEE International Conference on Systems, Man, and Cybernetics*, San Diego, CA.