Stock Prediction Using Twitter Sentiment Analysis

Anshul Mittal
Stanford University
anmittal@stanford.edu

Arpit Goel
Stanford University
argoel@stanford.edu

ABSTRACT
In this paper, we apply sentiment analysis and machine learning principles to find the correlation between "public sentiment" and "market sentiment". We use twitter data to predict public mood and use the predicted mood and previous days' DJIA values to predict the stock market movements. In order to test our results, we propose a new cross validation method for financial data and obtain 75.56% accuracy using Self Organizing Fuzzy Neural Networks (SOFNN) on the Twitter feeds and DJIA values from the period June 2009 to December 2009. We also implement a naive protfolio management strategy based on our predicted values. Our work is based on Bollen et al's famous paper which predicted the same with 87% accuracy.

1. INTRODUCTION
Stock market prediction has been an active area of research for a long time. The Efficient Market Hypothesis (EMH) states that stock market prices are largely driven by new information and follow a random walk pattern. Though this hypothesis is widely accepted by the research community as a central paradigm governing the markets in general, several people have attempted to extract patterns in the way stock markets behave and respond to external stimuli.

In this paper, we test a hypothesis based on the premise of behavioral economics, that the emotions and moods of individuals affect their decison making process, thus, leading to a direct correlation between "public sentiment" and "market sentiment". We perform sentiment analysis on publicly available Twitter data to find the public mood and the degree of membership into 4 classes - Calm, Happy, Alert and Kind (somewhat like fuzzy membership). We use these moods and previous days' Dow Jones Industrial Average (DJIA) values to predict future stock movements and then use the predicted values in our portfolio management strategy.

Related work Our work is based on Bollen et al's strategy [1] which received widespread media coverage recently. They also attempted to predict the behavior of the stock market by measuring the mood of people on Twitter. The authors considered the tweet data of all twitter users in 2008 and used the OpinionFinder and Google Profile of Mood States (GPOMS) algorithm to classify public sentiment into 6 categories, namely, Calm, Alert, Sure, Vital, Kind and Happy. They cross validated the resulting mood time series by comparing its ability to detect the public's response to the presidential elections and Thanksgiving day in 2008. They also used causality analysis to investigate the hypothesis that public mood states, as measured by the OpinionFinder and GPOMS mood time series, are predictive of changes in DJIA closing values. The authors used Self Organizing Fuzzy Neural Networks to predict DJIA values using previous values. Their results show a remarkable accuracy of nearly 87% in predicting the up and down changes in the closing values of Dow Jones Industrial Index (DJIA).

The rest of the paper is organized as follows. The second section briefly discusses our general approach towards solving the problem and the following sections discuss the individual components in greater detail. In Section 3, we briefly discuss the dataset that we have used for this paper and data preprocessing measures adopted. Section 4 discusses the sentiment analysis technique developed by us for the purpose of this paper. Section 5 includes in detail, the different machine learning techniques to predict DJIA values using our sentiment analysis results and presents our findings. In Section 6, we use the predicted values and devise a naive strategy to maintain a profitable portfolio.

2. ALGORITHM
The technique used in this paper builds directly on the one used by Bollen et al. [1]. The raw DJIA values are first fed into the preprocessor to obtain the processed values. At the same time, the tweets are fed to the sentiment analysis algorithm which outputs mood values for the four mood classes for each day. These moods and the processed DJIA values are then fed to our model learning framework which uses SOFNN to learn a model to predict future DJIA values using them. The learnt model as well as the previous DJIA and mood values are used by the portfolio management system which runs the model to predict the future value and uses the predicted values to make appropriate buy/sell decisions. Figure 1 shows a brief flow diagram of our technique.

The following sections discuss each component of our technique in greater detai

3. DATASET
In this project, we used two main datasets-

- Dow Jones Industrial Average (DJIA) values from June 2009 to December 2009. The data was obtained using Yahoo! Finance and includes the open, close, high and low values for a given day.
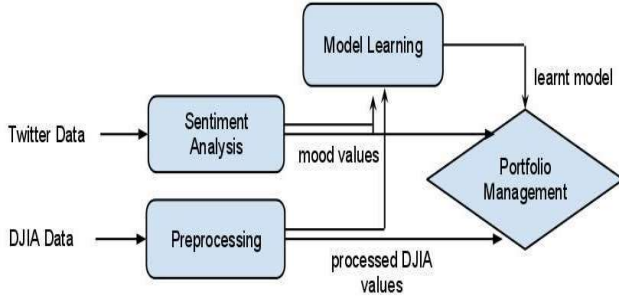
Figure 1: Our technique

- Publicly available Twitter data containing more than 476 million tweets corresponding to more than 17 million users from June 2009 to December 2009. The data includes the timestamp, username and tweet text for every tweet during that period. Since we perform our prediction and analysis on a daily basis, we split the tweets by days using the timestamp information.

Data Preprocessing

The data obtained from the above mentioned sources had to be pre-processed to make it suitable for reliable analysis. We pre-processed the DJIA data in the following manner-

1. While the Twitter data was available for all days lying in the giving period, the DJIA values obtained using Yahoo! Finance was (understandably) absent for weekends and other holidays when the market is closed. In order to complete this data, we approximated the missing values using a concave function. So, if the DJIA value on a given day is x and the next available data point is y with n days missing in between, we approximate the missing data by estimating the first day after x to be $(y+x)/2$ and then following the same method recursively till all gaps are filled. This approximation is justified as the stock data usually follows a concave function, unless ofcourse at anomaly points of sudden rise and fall.

2. If we observe the general movement of stock markets, it is associated with a few sudden jumps/falls and a brief period of small fluctuations around the new value. However, such jumps/falls are due to some major aberrations and cannot be predicted. Moreover, as we know the public memory is very short and even though the market may be trading at a much higher level than the previous year, that does not mean that calmness will be much higher than previous year; public mood is a very local metric. Therefore, we adjusted our stock values by shifting up/down for steep falls/jumps, respectively; making sure that we do not disturb the daily directional trend (up/down movement of stock prices).

3. Even after shifting the values in step 2, the values contained significant periods of volatile activity which are

very difficult to predict. We pruned our dataset by removing these periods for final training and testing.

Finally, in order to ensure that values were small and comparable, we computed the z-score of each point in the data series $((x-\mu)/\sigma)$ and used that in our analysis (The original values were of the order of 10^4, so MATLAB was giving a precision error when computing functions like $exp(-x^2)$)

4. SENTIMENT ANALYSIS

Sentiment analysis was an important part of our solution since the output of this module was used for learning our predictive model. While there has been a lot of research going on in classifying a piece of text as either positive or negative, there has been little work on multi-class classification. In this project, we use four mood classes, namely, Calm, Happy, Alert, and Kind. We tried several standard tools like OpinionFinder, SentiWordnet [5] etc. for our problem but found them inadequate and/or inefficient and therefore decided to develop our own analysis code. The methodology we adopted in finding the public sentiment is as follows-

1. **Word List Generation**
   We develop our own word list based on the well known Profile of Mood States (POMS) questionnaire. POMS is an established psychometric questionnaire which asks a person to rate his/her current mood by answering 65 different questions on a scale of 1 to 5 (For example, rate on a scale of 1 to 5 how tensed you feel today?). These 65 words are then mapped on to 6 standard POMS moods- Tension, Depression, Anger, Vigour, Fatigue and Confusion. In order to do automate this analysis for tweets, the word list needs to be appropriately extended. Bollen et al. [1] used the Google n-grams data for the same. We followed a much simpler approach of extending the list by considering all commonly occuring synonyms of the base 65 words using SentiWordNet and a standard Thesaurus.

2. **Tweet Filtering**
   As mentioned earlier, the tweet data is enormous and will take several hours to be processed if used as it is (which makes the task of daily predictions difficult). Therefore, we filtered and considered only those tweets which are more likely to express a feeling, i.e. we consider only those tweets which contain the words "feel", "makes me", "I'm" or "I am" in them.

3. **Daily Score Computation**
   We used a simple word counting algorithm to find the score for every POMS word for a given day-

   $$\text{score of a word} = \frac{\#\text{of times the word matches tweets in a day}}{\#\text{of total matches of all words}}$$

   The denominator accounts for the fact that the number of tweets could vary from one day to another. This works well for our problem because of the nature of tweets which contain simple sentence structures and only a maximum of 140 characters (in most cases much less). We tried using the Stanford coreNLP software for word tagging and then using a word's position in

the sentence to find its importance. However, similar to our experience working with OpinionFinder, we observed that this process, besides being extremely slow was not too beneficial.

4. **Score Mapping**

We map the score of each word to the six standard POMS states using the mapping techniques specified in the POMS questionnaire. We then map the POMS states to our four mood states using static correlation rules (for example, happy is taken as sum of vigour and negation of depression).

It is important to note that, given our formulation, it does not make much sense to compare the value of one mood against another; they should only be used to compare mood trends across days. We cross validated the results of our sentiment analysis technique by comparing the values returned by our algorithm around significant events like Thanksgiving day and Michael Jackson's death. As shown in Figure 2, the moods show a sharp rise of various mood states on Thanksgiving whereas one day after MJ's death, there is a sharp decline in happiness.
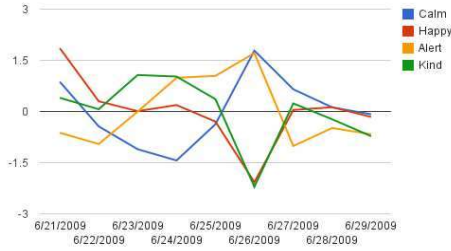
(a) Various moods after Michael Jackson's death on 25 June 2009
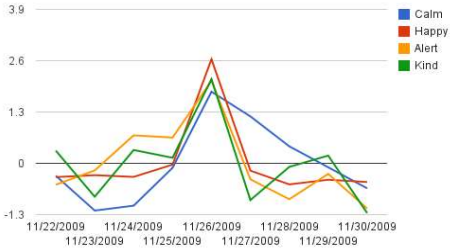
(b) Various moods on Thanksgiving day on 26 November 2009

Figure 2: Cross validation of our sentiment analysis by analyzing moods on some important events

Granger Causality In order to ascertain whether the mood values returned by our algorithm can be used to predict the future stock movements, we computed the p-values using Granger Causality analysis. Granger Causality analysis finds how much predictive information one signal has

about another over a given lag period. The p-value measures the statistical significance of our result i.e. how likely we could obtain the causality value by random chance; therefore, lower the p-value, higher the predictive ability. It is clear from Table 1 that calmness and happiness are most helpful in predicting the DJIA values as per the Granger causality. Moreover, the causality is best when we use the past 3 or 4 days' data. Figure 3 shows the DJIA values superimposed with sum of calmness and happiness over a brief period, indicating the clear predictive trend.

Table 1: p-values obtained using Granger causality analysis with different lags (in days)

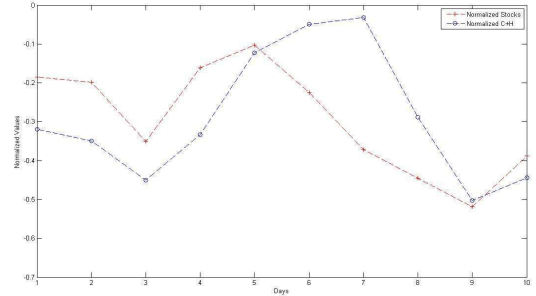| Lag	Calm	Happy	Alert	Kind
1	0.0207	0.4501	0.0345	0.0775
2	0.0336	0.1849	0.1063	0.1038
3	0.0106	0.0658	0.1679	0.1123
4	0.0069	0.0682	0.3257	0.1810
5	0.0100	0.0798	0.1151	0.1157

Figure 3: The DJIA, and Calmness + Happiness curves superimposed to show correlation

In the next section, we use the results of our sentiment analysis algorithm to learn a model that can predict the stock index and its movement.

5. MODEL LEARNING AND PREDICTION

Granger causality is based on linear regression, but the correlation between stocks and moods is certainly non linear. Therefore, after finding a causality relation between the past 3 days moods and current day stock prices, we tried 4 different learning algorithms (Linear Regression, Logistic Regression, SVMs, Self Organizing Fuzzy Neural Networks) to learn and study the actual correlation. For SVM we used the LIBSVM [2] library, but we implemented the other three in MATLAB ourselves as we could not find good working libraries for them. The Self Organizing Fuzzy Neural Network (SOFNN) is a five layer fuzzy neural network which uses ellipsoidal basis function (EBF) neurons consisting of a center vector and a width vector. We implemented the online alorithm for creating SOFNNs as introduced in [3] in which neurons are added or pruned from the existing network as new samples arrive. Neural networks have been considered to be a very effective learning algorithm for decoding nonlinear time series data [4], and financial markets often follow nonlinear trends. The authors in [1] showed the

3

Table 2: DJIA 5-SCV Accuracy Using 4 Different Algorithms

Algorithm	Evaluation	I_D	I_{CD}	I_{CHD}	I_{CAD}	I_{CKD}	I_{CHAD}	I_{CHKD}
Linear Regression	MAPE	7.28%	7.26%	7.66%	7.05%	7.43%	7.57%	7.78%
	Direction	64.44%	64.44%	71.11%	64.44%	64.44%	68.89%	71.11%
Logistic Regression	Direction	60%	60%	60%	60%	60%	60%	60%
SVM	Direction	59.75%	59.75%	59.75%	59.75%	59.75%	59.75%	59.75%
SOFNN	MAPE	9.71%	9.66%	11.03%	9.22%	11%	10.52%	11.78%
	Direction	64.44%	71.11%	75.56%	68.89%	73.33%	73.33%	73.33%

87% correlation using SOFNNs only, and our results also indicate that SOFFNs do the best among all other algorithms, giving nearly 75.56% accuracy.

In order to measure accuracy, we developed a novel validation technique called the *k-fold sequential cross validation (k-SCV)*. In this method, we train on all days upto a specific day and test for the next k days. The direct k-fold cross validation method is not applicable in this context as the stock data is actually a time series unlike other scenarios where the data is available as a set. Therefore, it is meaningless to analyze past stock data after training on future values. For the purpose of our analysis, we use $k = 5$.

Our Granger Causality analysis indicates that Calm and Happy are causative of the DJIA values. But to confirm the inverse dependence of other mood dimensions on DJIA we investigated a total of 7 different possibilities. In Table 2, I_D denotes the 5-SCV accuracy when only the past 3 days DJIA values are given as features. Similarly I_{CD}, I_{CHD}, I_{CAD}, I_{CKD}, I_{CHKD}, I_{CHAD} denote the accuracy when features are the past 3 days DJIA values (represented by D) along with the past 3 days mood values (C=Calm, H=Happy, A=Alert, K=Kind) in different combinations. MAPE indicates the Mean Absolute Percentage Error between our predicted values and the actual normalized stock values. The Direction accuracy indicates the percentage matchings in the trends (up/down) predicted by our training vs the actual daily trends in the stocks. When using classification algorithms, we used the up/down trends as class inputs and used the algorithm to directly predict trends whereas when using regression algorithms, we fed the normalized stock values as input and used the predicted stock values to obtain direction (up/down) trends.

We find that I_{CHD} gives the best results in all the algorithms considered, indicating that Calmness and Happiness are more predictive of the stock values, confirming the Granger causality analysis and unlike the [1] result. They showed 87% correlation when features were Calm and DJIA values of past 3 days, and reported the accuracy only on a specific test set, without reporting any cross validation error average. Our results are also in conjunction with the philosophy that happiness should in general be causative of the stock values. Figure 4 shows a graph of the normalized stock values as predicted by our SOFNN algorithm vs the actual normalized stock values when trained on the Happy+Calm+DJIA feature set. We find a very close correlation and hence the small value of Mean Absolute Percentage Error (MAPE) and good Directional Accuracy.

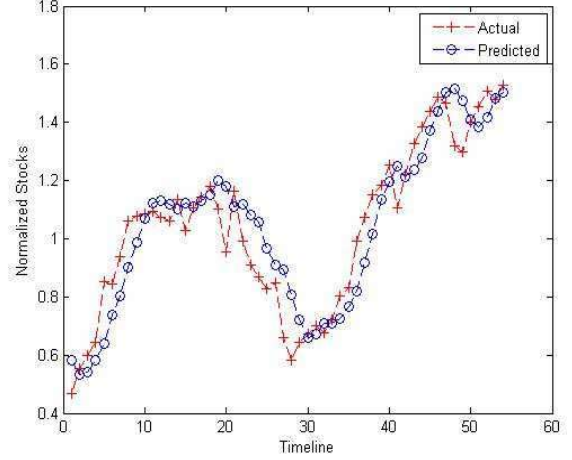We can draw several important conclusions from the Table

Figure 4: Predicted vs Actual Stock Values using SOFNN on Calm+Happy+DJIA for 40 Consecutive Days

2. Firstly SVMs and Logistic Regression perform badly on this dataset, giving the same percentage values for Direction Accuracy for all mood combinations. This shows that classification (directly predicting trends) is not the ideal methodology for this problem. Linear Regression performs pretty good, which is in conjunction with the Granger Causality results, whereas SOFNN performs the best. If we look at the Direction Accuracy for SOFNN we observe that the best value is for I_{CHD} (75.56%). Also on adding any other mood dimension the Direction Accuracy worsens, ie. the Direction Accuracy is clearly lower for I_{CHAD}, I_{CHKD}. This shows that by adding more features we would essentially be overfitting the data. If we try to remove some feature and find the Direction Accuracy, we observe that the result still worsens. Hence Calm and Happiness are indeed more indicative of the stock values than any other moods. If we consider the MAPE values, we find that Calm and Alert mood dimensions do marginally better than the others, but they are poor in predicting the Direction or the trends of stock movement.

To compare the significance of our results on the Directional Accuracy with those of [1], we see that 5 fold sequential cross validation is clearly a better indicator. [1] have used a specific test period which gives a 87.6% Directional Accuracy, and they have proved that statistically the probability of that event happening with random success is pretty low.

But we have on the other hand cross validated over the entire period, using our 5-fold SCV technique. As mentioned earlier, this technique makes more sense for stock data because the usual cross validation would essentially be using future stock values to predict the past ones, which is an incorrect technique when financial data is concerned.

6. PORTFOLIO MANAGEMENT

Having predicted the DJIA closing values one day in advance, we can use these predicted values to make intelligent sell/buy decisions. We develop a naive greedy strategy based on a simple assumption that we can hold at most one stock at any given time (or s stocks if all stocks are always bought and sold together) Following are the steps/features of our strategy-

- **Pre-computation**
  We maintain a running average and standard deviation of actual adjusted stock values of previous k days

- **Buy Decision**
  If the predicted stock value for the next day is n standard deviations less than the mean, we buy the stock else we wait.

- **Sell Decision**
  If the predicted stock value is m standard deviations more than the actual adjusted value at buy time, we sell the stock else we hold.

Note that the above strategy has three parameters- k, n and m. Our experiments show that the optimal parametrization is $n = m = 1$ and $k = 7$ or 15 (Note that in order to trade effectively, we needed a large enough test set, containing at least 30-40 entries, thereby limiting the scope of our experiments because of limited test data.) The profit obtained using our strategy is as follows-

$k = 7$ - $Profit = 527.2$ Dow Points
$k = 15$ - $Profit = 543.65$ Dow Points

The total range of stock movement in the same period is 920.72 Dow Points.

Please note that while the above analysis is entirely in terms of Dow points, it is easy to correlate a profit in Dow points with a monetary value. For example, the Dow Diamonds is an exchange-traded fund that holds the 30 stocks that comprise the DJIA and researchers have shown a 98.5% correlation between the movement of DJIA and the Dow Diamonds. Similarly, we can find many other funds which can translate this profit in Down point to a corresponding profit in Dollars.

7. CONCLUSIONS AND FUTURE WORK

We have investigated the causative relation between public mood as measured from a large scale collection of tweets from twitter.com and the DJIA values. Our results show that firstly public mood can indeed be captured from the large-scale Twitter feeds by means of simple natural language processing techniques, as indicated by the responses towards a variety of socio-cultural events during the year 2009. Secondly, among the observed dimensions of moods, only calmness and happiness are Granger causative of the DJIA by 3-4 days. Thirdly, a Self Organizing Fuzzy Neural Network performs very good in predicting the actual DJIA values when trained on the feature set consisting of the DJIA values, Calm mood values and Happiness dimension over the past 3 days. The performance measure we have used is k-fold sequential cross validation, which is more indicative of the market movements for financial data. Finally a naive implementation of portfolio management using our strategy indicates a decent profit over a range of 40 days.

Our results are in some conjunction with [1], but there are some major differences as well. Firstly our results show a better correlation between the calm and happy mood dimensions with the DJIA values, unlike their result, which showed high correlation with only calm mood dimension. Secondly, we haven't been able to obtain high percentage result of 87%, but our 75.56% result using k-fold sequential cross validation gives stronger evidence that the correlation is over the entire range of data. The profits in out naive implementation of portfolio management shows that our MAPE estimates are pretty much accurate.

Finally, its worth mentioning that our analysis doesn't take into account many factors. Firstly, our dataset doesn't really map the real public sentiment, it only considers the twitter using, english speaking people. It's possible to obtain a higher correlation if the actual mood is studied. It may be hypothesized that people's mood indeed affect their investment decisions, hence the correlation. But in that case, there's no direct correlation between the people who invest in stocks and who use twitter more frequently, though there certainly is an indirect correlation - investment decisions of people may be affected by the moods of people around them, ie. the general public sentiment. All these remain as areas of future research.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] J. Bollen and H. Mao. Twitter mood as a stock market predictor. *IEEE Computer*, 44(10):91–94.
[2] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.
[3] G. P. Gang Leng and T. M. McGinnity. An on-line algorithm for creating self-organizing fuzzy neural networks. *Neural Networks*, 17(10):1477–1493.
[4] A. Lapedes and R. Farber. Nonlinear signal processing using neural network: Prediction and system modeling. In *Los Alamos National Lab Technical Report*.
[5] A. E. Stefano Baccianella and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*. LREC.