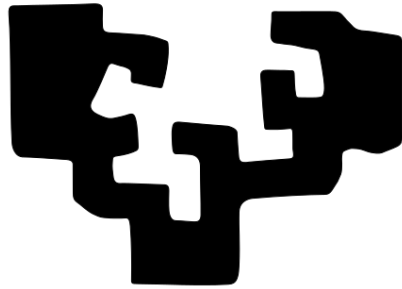


eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

Advanced Techniques in AI

Sentiment Analysis On The Stock Market

Mikel Iparragirre and Aimar Ugarte

Introduction	2
Stock market prediction using machine learning techniques	3
Examples	3
Sentiment analysis	4
Sentiment Analysis Algorithms	4
Rule based	4
Automatic systems	5
Hybrid	5
Tools	5
Sentiment analysis on the stock market	6
Twitter and the stock market, the power of twitter	7
Sentiment analysis on Twitter	8
Conclusion	9
References	10

Introduction

Since before the first formal stock exchange house, a roofless courtyard, was opened in Amsterdam in 1611, the stock market has been an important pillar of the economy. Throughout the years it has expanded into a global state, it has influenced history and in recent years, with the birth and development of Artificial Intelligence, has experienced drastic changes.

In the 1980s the trading floors started to implement electronic trading systems, instead of using the old Open Outcry where professionals used shouts and hand signals to transfer information. What were once floors filled with buzzing activity and constant movement turned into a colder, more organized affair. This opened the door for AI into a very profitable environment. The first prediction methods used focused on quantitative facts but as crashes like the 1987 black Monday proved there was more than numbers to this problem.

Encompassing what is basically the whole world's economy, the stock market is a complicated affair. The constant rise and fall of prices and values without apparent reason makes many question its rationale. In the end the human factor is an important one in economy with the mood of the investors, buyers and other interested people affecting its performance. That is why Sentiment Analysis is used to predict market share changes and the social Twitter, with its digital, contained and mood signaling information of significant people, is the perfect target.

Stock market prediction using machine learning techniques

Predicting the stock market has always been a classic problem for investors as the efficient-market hypothesis (EMH) states that it is not possible to predict or outperform the overall market through stock selection or market timing. But technical analysts believe that most information about the stocks are reflected in recent prices and so if trends in the movements are observed then prices can be easily predicted. In addition, the stock market is strongly affected by politics, general economy, bank rate, exchange rate, expeculations, psychology of investors, etc.

Artificial Neural Networks (ANN) and Support Vector Regression (SVR) are two machine learning algorithms which have been most widely used for predicting stock price and stock market index values.

Examples

In the report “*Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm*” by Wei Shen, Xiaopen Guo, Chao Wu and Desheng Wu they introduced The Artificial Fish Swarm Algorithm (AFSA) to train radial basis function neural network (RBFNN). Their experimented on the Shanghai Stock Exchange showed that their new hybrid algorithm, RBF optimized by AFSA, had a relatively small average forecasting error.

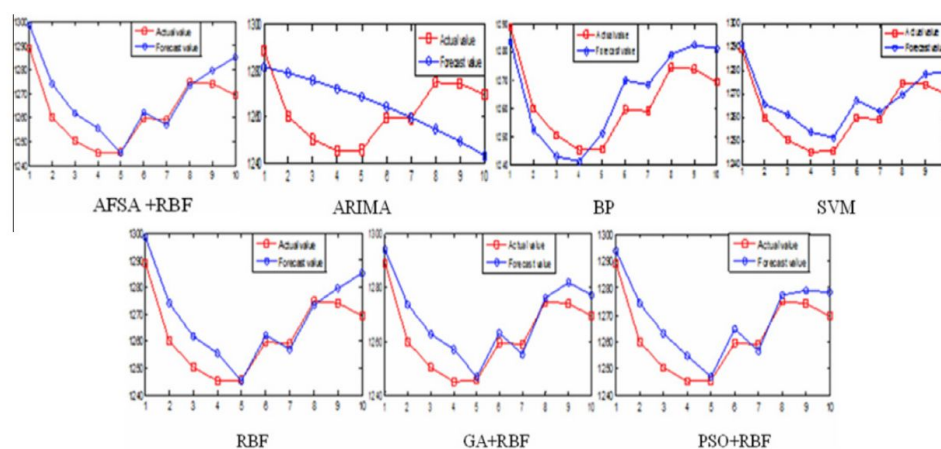


Fig. 5. Comparison of the forecasting results of seven discussed models.

This work mainly employs quantitative indicators, however, stock movements are affected not only by quantitative factors, but also by non-quantitative factors, such as breaking news, macroeconomic policies and regulations, psychological factors, etc. These non-quantitative factors can be integrated into mathematical algorithms using text-mining techniques to effectively increase forecast accuracy. One of these text-minig techniques is Sentiment analysis, of which we are going to talk next.

Sentiment analysis

Sentiment analysis also known as Opinion Mining is a field within Natural Language Processing (NLP) is the automated system that aims to determine the attitude of a speaker or writer with respect to some topic or emotional reaction to a document, interaction, or event.

Usually, besides identifying the opinion, these systems extract attributes of the expression:

- Polarity: if the speaker express a positive or negative opinion,
- Subject: the thing that is being talked about,
- Opinion holder: the person, or entity that expresses the opinion.

The method and algorithms used in sentiment analysis can be classified in categories:

- Rule-based systems that perform sentiment analysis based on a set of manually crafted rules.
- Automatic systems that rely on machine learning techniques to learn from data.
- Hybrid systems that combine both rule based and automatic approaches.

Sentiment Analysis Algorithms

Rule based

Rule-based approaches define a set of rules in some kind of scripting language that identify subjectivity, polarity, or the subject of an opinion.

A basic example would be the following:

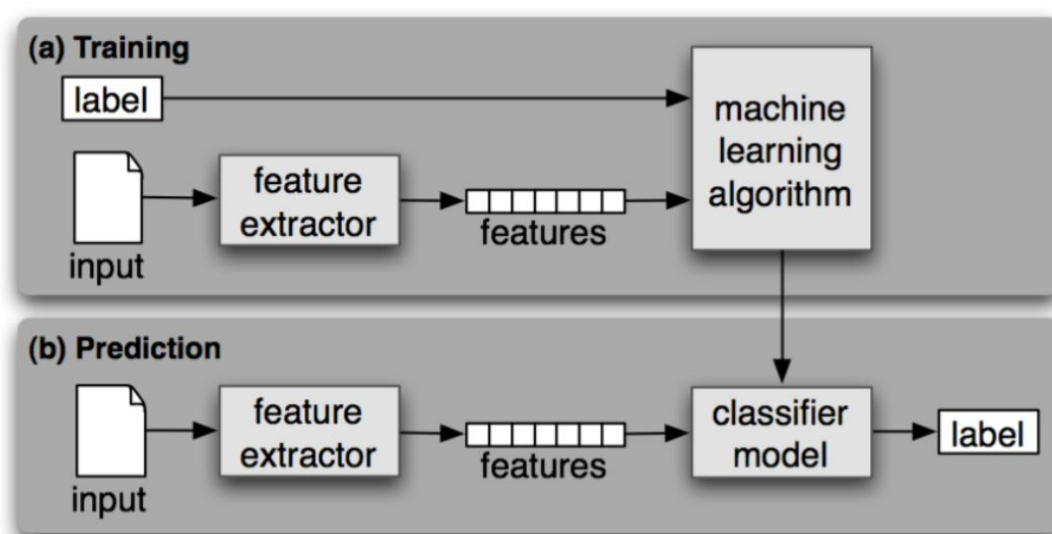
1. Define two lists of polarized words, a list of negative words such as bad, worst, ugly, etc and a list of positive words such as good, best, beautiful, etc.
2. Given a text:
 - a. Count the number of positive words that appear in the text.
 - b. Count the number of negative words that appear in the text.
3. If the number of positive word appearances is greater than the number of negative word appearances return a positive sentiment. If the number of negative word appearances is greater than the number of positive word appearances return a negative sentiment. Otherwise, return neutral.

This system is naïve since it does not take into account how words are combined in a sequence. A more advanced processing can be made, but these systems get very complex quickly.

Automatic systems

The automatic systems do not rely on manual crafted rules (such as the positive and negative words list on rule based systems) but on machine learning techniques. The task is modeled as a classification problem where the classified is fed with test and returns the corresponding category; positive, negative or neutral.

For this systems two type of input is used, training input and prediction input. Training input is used to create a database for the algorithm were the model associates certain type of characteristics to the corresponding category. This information is used to predict the prediction input, categorizing the characteristics of each input text, like shown in the next picture.



Hybrid

The concept of hybrid methods is very intuitive: just combine the best of both worlds, the rule-based and the automatic ones. Usually, by combining both approaches, the methods can improve accuracy and precision.

Tools

There are some open source sentiment analysis tools available online:

1. [GATE](#) plugins
2. Stanford Sentiment Analysis Module (Deeply Moving: Deep Learning for Sentiment Analysis)
3. LingPipe (Sentiment Analysis Tutorial)

And some open Source Dictionary or resources:

1. SentiWordNet
2. Bing Liu Datasets (Opinion Mining, Sentiment Analysis, Opinion Extraction)
3. General Inquirer Dataset (General Inquirer Categories)

Sentiment analysis on the stock market

One of the fields that sentiment analysis can be applied is the stock market. Using sentiment analysis algorithms it can somewhat predict or anticipate the market movements.

For these algorithms to work, the right input is needed; news, articles, trends, investors blogs, tweets, etc. The algorithm will be fed automatically with this input, classifying it as positive, negative or neutral. If the general feeling for one specific stock is positive it would be a good buying opportunity, otherwise it would be the right moment to sell.

All this system can be automated, from the input recollection to the stocks buy and sell actions.

The next graph shows how accurate the sentiment analysis prediction can be, the blue line been the stock price of Microsoft and the green line the sentiment moving average calculated by Sentdex.



Twitter and the stock market, the power of twitter

The microblogging service of Twitter is one of the biggest social networks today. Its power to distribute information is unparalleled by any other social network. News show the latest tweet of Angela Merkel or the king of Spain and people are sent to prison for making a joke about a dead fascist.

The best example of the power of twitter and its power over the stock market is President of The United States of America Donald J. Trump. Being one of (if not the most) important person on the planet with the power to enact acts of war without congressional approval has a lot of sway over the masses and the stock alike. Even before being officially invested his tweet about the infamous F-35 sixth generation fighter project made Lockheed, the company developing it, lose 2% of its share in the following hours while Boeing's, its competitor, rose 0.7%. In another case Toyota suffered a loss of 1.2 billion dollars in five minutes after Trump tweeted negatively about the plan to build a new Corolla factory in Mexico. There are now apps that notify you if Trump is tweeting about a company you are investing in and there are studies about the overall effects of his tweets on the American economy.

However, Twitter's power does not come only from presidents. Recently Tesla Motors has suffered after its CEO Elon Musk was accused by the Securities and Exchange Commission of fraud for misleading investors over a post in twitter where he claimed to have "funding secured" for a buyout of the company at 420 dollars a share. Elon Musk's mental state at the time of the tweet appears to not have been lucid. The company's stock was hardly hit by this debacle and on one day dropped up to 14 percent. In the end, the lawyers convinced Elon Musk to accept a deal with the SEC and avoid courts. Musk paid a \$20 million fine and he stepped down from CEO for three years but the SEC allowed him to remain as chief executive among other changes.

With its influence clear the only thing left is to apply Sentiment analysis on Twitter.

Sentiment analysis on Twitter

In sentiment analysis predictions of the stock market using Twitter the key to good results comes not only from the tools used, which are many, but also the data, the measures of mood selected and the preprocessing done to the data before analysing it.

In the highly cited article of 2010 "Twitter mood predicts the stock market" by Johan Bollen, Huina Mao and Xiaojun Zeng, they achieved an accuracy of 86.7% in predicting the changes in the closing values of the Dow Jones Industrial Average (DJIA) and a reduction of the Mean Average Percentage Error (MAPE) by more than 6%. They used two mood tracking tools, OpinionFinder to measure positive vs negative mood and Google-Profile of Mood States (GPOMS) to measure mood in 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy). A Granger causality analysis and a Self-Organizing Fuzzy Neural Network (SOFNN) were used to investigate that mood states as measured were predictive of changes in DJIA closing values. Their Granger causality analysis made them conclude that the Calm mood had the most correlation with the DJIA results .

In another article, inspired by the 2010 investigation, Anshul Mittal and Arpit Goel set to again test the hypothesis that emotions and moods of individuals affect their decision making process and thus leads to a direct correlation between "public sentiment" and "market sentiment". They too used DJIA data (6 months) and twitter tweets (476 million) of the same timeframe but the moods measurement was different, being only Calm, Happy, Alert and Kind. In their prediction they saw too that SOFNN algorithms were the most accurate, but reached only 75.56%. In their A Granger causality analysis they concluded that unlike in the Bollen article, both Happy and Calm moods had the greatest correlation with the DJIA values. The authors noted that taking into accounts more factors, like better mapping of the public it would be possible to obtain a better correlation between mood and DJIA results.

In the following years more papers have been published about predicting stock market movements using twitter sentiment analysis. They have used different sample sizes and types, used different algorithms and mood prediction systems but in the end most of them yielded the same results of high prediction rates on short term changes .

Conclusion

Twitter mood and market share changes haven been linked with each other repeatedly by several papers and the idea that Stock Market movement can not be predicted, as the Efficient Market Hypothesis says it is long gone. What is certain is that it is possible to analyse and predict general changes in the market with the proper methods and tools by looking at the general mood of society on twitter.

References

[Predicting stock market index using fusion of machine learning techniques](#)

[Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm](#)

[Efficient-market hypothesis \(Wikipedia\)](#)

[Stock market prediction system with modular neural networks](#)

[Sentiment Analysis nearly everything you need to know](#)

[Sentiment analysis \(Wikipedia\)](#)

[Opinion Mining and Sentiment Analysis](#)

[Investor Sentiment in the Stock Market](#)

[How Accurate Can Sentiment Analysis for Stocks Be?](#)

[Elon Musk Steps Down as Chairman in Deal With S.E.C. Over Tweet About Tesla](#)

[Lockheed F-35 fighter project in doubt after Trump tweet encourages rival Boeing](#)

[Toyota Stock Drops Immediately After Trump Tweet About Mexican Factory](#)

[How Does President Trump's Twitter Use Impact Forex, Markets And Stocks?](#)

[Twitter mood predicts the stock market](#)

[Predictive Sentiment Analysis of Tweets: A Stock Market Application](#)

[Analyzing Stock Market Movements Using Twitter Sentiment Analysis](#)

[Stock Prediction Using Twitter Sentiment Analysis](#)

[Twitter as a tool for forecasting stock market movements: A short-window event study](#)

[Using Twitter to Predict the Stock Market](#)