Sentiment analysis of Tweets in correlation with financial investments.

Magnus L Kirø
kiro@stud.ntnu.no
Student, IDI, IME, NTNU

Supervisor:
Pinar Öztürk

Co Supervisor:
Arvid Holme

October 2, 2013

Abstract

This is very much work in progress. All input is welcome. Don't hessitate to contact me.

Background: Twitter har become mainstream and it's significance in the financial world is increasing. This makes tweets interesting as a source of stock predictions.

This article will look at the state of the art techniques that exist and compare them. Keeping in mind the trends that crystalises from the underlying sentiment analysis of stock exchange relevant tweets.

What is state of the art in sentiment analysis with twitter as the datasource?

A summary of relevant research will be provided. And thoughts of missing aspects of the current results will uttered.

Results: Rough results of my research.

Conclusion: All OK ? No?

Keywords:

Acknowledgements

Acknowledgements goes here.

Metadata

Metadata ?

Typically repositories, links to code, file downloads, websites etc. Maybe contact info.

Contents

List of Figures

List of Listings

Introduction

What and why

Findings

Outline

The outline of the document and the description of what which part is about.

State of the Art / background / related work

Twitter

Twitter is a social and information networking. It's a real-time service that connects users to the latest stories, their interests, ideas and much more. The microblogging site allows users to find and follow accounts that the user has an interest in.

At the core of Twitter you have the Tweet. The Tweet is the 140 character message. These bursts of information combined are the life blood of Twitter. Tweets lets you communicate with other users, share photos, post all kinds of information. The small size of the tweets are not a hindrance for the flow of information. [1]

The fast growing messaging service handles 1.6 billion searches every day. As of 2012 the 500 million users would generate 3.2 queries each on any given day. 340 million tweets were posted every day. [2]

Today most medium and large companies are on Twitter. Posts can contain any type of information, from promotional content to service status to financial reports.

Companies use twitter for feedback and customer relations. Questions can be asked with a hashtag of to a specific user. This makes it easy to sort filter the messages, and therefore easier to get in contact with the customer. Best Buy demonstrated the successfulness of twitter in customer relations by answering questions with a specific hashtag. In 2009 they had answered nearly 20 thousand questions using twitter. [Li and Li, 2013, p1] Market Intelligence is also a major aspect of the microbloggin sphere.

Twitter represents one of the largest and most dynamic datasets of user generated content. Along with Facebook twitter data is real time. This has

[1] About Twitter: https://twitter.com/about

[2] Wikipedia: http://en.wikipedia.org/wiki/Twitter

major implications for anyone who are interested in sentiment, public opinion or customer interaction. [Speriosu et al., 2011]

Companies who use twitter. [Jubbega, 2011]

Sentiment

Opinion mining on the web is not a new phenomenon. But in resent years it has become much more attractive to traders in the financial world. Twitter and the social media's opinion is on the rise. This means a surplus of raw data with easy access. Companies all over the world has started to use twitter and readily available tweets to their benefit. Trading with social media is part of the trend. Although there are some drawbacks and shortcomings. Noise and garbage is one of them. It's difficult to accurately sort through all the data and get only the information relevant for your use. Even if your right 80% of the time, the last 20% can prove devastating. [Stevenson, 2012]

Sentiment broadly refers to the state of mind a person has. Whereas negative of positive. Based on the current state of mind the person will do optimistic or pessimistic choices. A positive state of mind leads to optimistic judgements of future events. And a negative state of mind leads to pessimistic judgements. [Doukas et al., January 10, 2010, p4]

The intention of users are also a part of the driver in user activity on microbloggs. The users may have different roles and intentions in different communities in the microblogging sphere [Java et al., 2007]. This might also be a factor in the sentiment analysis.

Sentiment Analysis

[Li and Li, 2013] approaches the classification of sentiment in tweets in four steps. First is the topic detection. The topic is the overall theme of the message. This step extracts and identifies the topics associated with the queries of users. Following that the classification of opinion happens. This judges the polarity of the sentiment. The state of mind of the user can be recorded.

A problem that arises is the credibility of the expresser. [Li and Li, 2013] addresses this to get a better summary of the sentiment. Then aggregates, the three previously described parts of the classification, to get a truer reflection of the opinions.

[Barbosa and Feng, 2010] looks upon the problem of noise in biased and noisy data. They focus on noisy labels and add features to the tweets to increase the classification properties of the tweets. Then the tweets are classified as subjective or objective. This is to filter out the tweets that don't project a sentiment. The subjective tweets are then classified as positive or negative. Then [Barbosa and Feng, 2010] generalise the classification of tweets by using meta data about the tweets and how tweets are written. This results in a more abstract representation of tweets and the classification. [Barbosa and Feng, 2010] provides a better way to classify tweets.

Another approach to the sentiment challenges with twitter is explored by [Becker et al., 2013]. Their explorations techniques for Contextual Polarity Disambiguation and Message Polarity Classification. Constrained and supervised learning is utilised to create models of classification. They describe a system that solves these tasks with the help of polarity lexicons and dependency parsers. Expanded vocabulary is one of the main aspects of their success, as they say in their findings: "We hypothesize this performance is largely due to the expanded vocabulary obtained via unlabeled data and the richer syntactic context captured with dependency path representations." [Becker et al., 2013]

From a bit earlier [Speriosu et al., 2011] have researched the polarity classification of tweets. In contrast to [Becker et al., 2013], [Speriosu et al., 2011] has used distant supervision and labeled propagation on a graph based data structure. The data structure represents users with tweets as nodes. And tweets with bigrams, unigrams, hashtags, etc as subnodes of the tweets. A label propagation approach rivals a model supervised with in-domain annotated tweets and outperforms the noisily supervised classifier and a lexicon-based polarity ratio classifier. [Speriosu et al., 2011]

Sentiment in Finance

Use of sentiment can predict changes and momentum in the market. Bad news in an optimistic period creates cognitive dissonance in the small investors. This impacts the market by slowing down the selling rate of loosing stocks. [Doukas et al., January 10, 2010, p29]

Further we can see that optimistic sentiment has a 2% monthly average return. While the investor sentiment is pessimistic we see a drastic reduction in returns. Down to 0.34%.[Doukas et al., January 10, 2010, p5] After optimistic periods it is indicated that the monthly return is reduced to -0.49%. On the contrary there is no equivalent change after a pessimistic period. [Doukas et al., January 10, 2010, p6-7] Momentum profits are only significant when the sentiment is optimistic. [Doukas et al., January 10, 2010, p29]

Hope and fear is used by [Zhang et al., 2011] to decide the movement of the market. The sentiment is aggregated to be hopeful of fearful. This basically focuses on positivity and negativity of the sentiment of that particular day.

The daily sentiment is then compared to the market indicators of the same day to create a prediction of the market. [Zhang et al., 2011] finds that calm times give little hope or other emotions. Little turmoil results in few fluctuations in the market. And opposite, lots of emotions(hope, worry, fear), gives speed to the market.

[Brown and Cliff, 2004, p3] indicates that the sentiment does not cause subsequent market returns. For a short-term marketing timing this is bad news. However with the changes in social media over the last decade how is the situation today? With the microblogging sphere of today we can easily see the correlation of sentiment and the market indicators [todo:Citation]. But does the sentiment cause changes in the market-return? [Brown and Cliff, 2004, p3] also says that optimism is associated with overvaluation and subsequent low returns.

[Brown and Cliff, 2004, p] concludes that aggregated sentiment measures has strong co-movement with changes in the market. He also indicates that sentiment doesn't appear to be a good trading strategy. This, in the view of [Zhang et al., 2011] indicates a leap in sentiment research and what is possible with the microblogging of today.

Finance and Trading

[Brown and Cliff, 2004, p2] writes the following on over-reaction of investors: " He(Siegel (1992)) concludes that shifts in investor sentiment are correlated with market returns around the crash. Intuitively, sentiment represents the expectations of market participants relative to a norm: a bullish (bearish) investor expects returns to be above (below) average, whatever "average" may be.". In the light of resent changes in the financial world and the utilisation of sentiment from social media, the notion that opinions and sentiment of investors and market actors affect the market is not a new observation.

Trending

What is happening in the world on trending ?

Data

What data was used

Obtaining the data

Simple guide to access the twitter api: http://datascienceandprogramming.wordpress.com/2013/0
api/

Structure

Problems and shortcomings

Sentiment Analysis

What is sentiment

What can we use the sentiment for

Getting the sentiment

Techniques

- of a Tweet

Sentiment trend

Trending

The trend is your friend

What is a trend and how does it work.

Trending and it's applications

In Finance

Elsewhere

The Trend

Comparing the trend and the moving average

Results

All our results are in ones and zeroes.

Conclusion

We worked hard, and achieved very little.

Future Work

All the things I didn't have time to do my self.

Articles / SLR

Article template

file:*filename.pdf* citation:[]
* What did they use tweets for?
* How are tweets used?
* Event detection. Is the tweet about merging?
* Where can this article be useful later?
* What does this article give answers to?

Twitter Mood Predicts the Market

Twitter Mood Predicts the Stock Market

Predicting the Future with Social Media

Twitter as driver of stock price

Tweets and Trades: The Information Content of Stock Microblogs

file:*SSRN-id1702854.pdf* citation:[Sprenger and Welpe, December 2010]

* What did they use tweets for?

"We find the sentiment (i.e., bullishness) of tweets to be associated with abnormal stock returns and message volume to predict next-day trading volume." [Sprenger and Welpe, December 2010]

* How are tweets used?

* Event detection. Is the tweet about merging?

* Where can this article be useful later?

What twitter is used for, Twitter chapter.

Twitter incentives. [Sprenger and Welpe, December 2010, p4]

Description of bullishness, message volume and what it does etc.

[Sprenger and Welpe, December 2010, p52] suggest that stock microblogs can claim to capture key aspects of the market conversation.

Picking the right tweets remains just as difficult as making the right trades.

* What does this article give answers to?

Whether bullishness can predict returns. Whether message volume is related to returns, trading volume, or volatility. Whether the level of disagreement among messages correlates with trading volume or volatility. Whether and to what extent the information content of stock microblogs reflects financial market developments Whether microblogging forums provide an efficient mechanism to weigh and aggregate information

Exploiting Topic based Twitter Sentiment for Stock Prediction

file:*filename.pdf* citation:[Si et al., 2013]

* What did they use tweets for?

Predicting the stock market. Stock index time series analysis. daily one-day-ahead predictions.

* How are tweets used?

Dirichlet Process mixture model to learn the daily topic set. Vector regression. Topic-based prediction.

* Event detection. Is the tweet about merging?

* Where can this article be useful later?

Twitter's topic based sentiment can improve the prediction accuracy. [Si et al., 2013, p28]

* What does this article give answers to?

Twitter as driver of stock price

file:*Twitter as driver of stock price-Jubbega.pdf* citation:[Jubbega, 2011]

Twitter Polarity Classification with Label Propagation over Lexical Links and the Follower Graph

file:*twitter polarity classification.pdf* citation:[Speriosu et al., 2011]

* What did they use tweets for?

Polarity classification. Positive/negative.

* How are tweets used?

With label propagation. Distant supervision. Graph based data structure. user–¿tweet–¿bigram/unigram/hashtag/etc.

* Event detection. Is the tweet about merging?

* Where can this article be useful later?

Data section / sentiment /

Twitter section: What people uses twitter for.

Label propagation approach rivals a model supervised with in-domain annotated tweets and outperforms the noisily supervised classifier and a lexicon-based polarity ratio classifier. [Speriosu et al., 2011]

Twitter represents one of the largest and most dynamic datasets of user generated content.

* What does this article give answers to?

AVAYA: Sentiment Analysis on Twitter with Self-Training and Polarity Lexicon Expansion

file:*Sentiment Analysis on Twitter with Self-Training and Polarity Lexicon Expansion.pdf* citation:[Becker et al., 2013]

* What did they use tweets for?

Contextual Polarity Disambiguation and Message Polarity Classication * How are tweets used?

Constrained learning with supervised learning. Unconstrained model that used semi-supervised learning in the form of self-training and polarity lexicon expansion

* Event detection. Is the tweet about merging?

* Where can this article be useful later?

Technical approach of models and sentiment analysis. State of the art on sentiment analysis with twitter.

* What does this article give answers to?

dependency parses, polarity lexicons, and unlabeled tweets for sentiment classification on short messages

We hypothesize this performance is largely due to the expanded vocabulary obtained via unlabeled data and the richer syntactic context captured with dependency path representations. [Becker et al., 2013]

Robust Sentiment Detection on Twitter from Biased and Noisy Data

file:*Robust Sentiment Detection on Twitter from Biased and Noisy Data.pdf* citation:[Barbosa and Feng, 2010]

* What did they use tweets for?

Sentiment analysis with focus on noise reduction.

* How are tweets used?

Noisy labels. Classifies tweets as subjective or objective. Then distinguishes the subjective into positive and negative tweets. Generalization of tweet classification. Meta-information. How tweets are written. More abstract representation.

* Where can this article be useful later?

Previous work, sentiment analysis, twitter, sentiment features. * What does this article give answers to?

It provides a better way to classify tweets.

Investor sentiment and the near-term stock market

file:*Investor sentiment and the near-term stock market.pdf* citation:[Brown and Cliff, 2004]

* Where can this article be useful later?

In the finance chapter for historic value and where we have come from.

[**?**, p2] on over-reaction of investors writes: " He(Siegel (1992)) concludes that shifts in investor sentiment are correlated with market returns around the crash. Intuitively, sentiment represents the expectations of market participants relative to a norm: a bullish (bearish) investor expects returns to be above (below) average, whatever "average" may be.". In the light of resent changes in the financial world and the utilisation of sentiment from social media, the notion that opinions and sentiment of investors and market actors affect the market is not a new observation.

[Brown and Cliff, 2004, p3] indicates that the sentiment does not cause subsequent market returns. For a short-term marketing timing this is bad news. However with the changes in social media over the last decade how is the situation today? With the microblogging sphere of today we can easily see the correlation of sentiment and the market indicators [todo:Citation]. But does the sentiment cause changes in the market-return? [Brown and Cliff, 2004, p3] also says that optimism is associated with overvaluation and subsequent low returns.

* What does this article give answers to?

[Brown and Cliff, 2004, p] concludes that aggregated sentiment measures has strong co-movement with changes in the market. He also indicates that sentiment doesn't appear to be a good trading strategy. This, in the view of [Zhang et al., 2011] indicates a leap in sentiment research and what is possible with the microblogging of today.

Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear"

file:*Predicting Stock Market Indicators Through Twitter.pdf* citation:[Zhang et al., 2011]

* What did they use tweets for?

Gather hope and fear for each day using tweets. The sentiment indication of each day is compared to the marked indicators of the same day.

* How are tweets used?

Get the Positive/negative sentiment.

* Event detection. Is the tweet about merging?

* Where can this article be useful later?

Address the question of intention of users on twitter. Good summary of things done in regards to twitter. (Might be a bit outdated, from 2010).

* What does this article give answers to?

That hope, fear and worry makes the stock go down the day after. Calm times, little hope, fear or worry, makes the stock go up.

Deriving market intelligence from microblogs

file:*Deriving market intelligence from microblogs.pdf* citation:[Li and Li, 2013]

* How are tweets used?

Companies use twitter for feedback and customer relations. Questions can be asked with a hashtag of to a specific user. This makes it easy to sort filter the messages, and therefore easier to get in contact with the customer. Best Buy demonstrated the successfulness of twitter in customer relations by answering questions with a specific hashtag. In 2009 they had answered nearly 20 thousand questions using twitter. [Li and Li, 2013, p1] Market Intelligence is also a major aspect of the microbloggin sphere.

* What did they use tweets for?

[Li and Li, 2013] approaches the classification of sentiment in tweets in four steps. First is the topic detection. The topic is the overall theme of the message. This step extracts and identifies the topics associated with the queries of users. Following that the classification of opinion happens. This judges the polarity of the sentiment. The state of mind of the user can be recorded.

A problem that arises is the credibility of the expresser. This is addressed to get a better summary of the sentiment. Then [Li and Li, 2013] aggregates, the three previously described parts of the classification, to get a truer reflection of the opinions.

* Event detection. Is the tweet about merging?
* Where can this article be useful later?

stateOf-twitter / state-sentiment / data /

* What does this article give answers to?

The social media stock pickers

file:*social_media_stock_pickers.pdf* citation:[Stevenson, 2012]

Opinion mining on the web is not a new phenomenon. But in resent years it has become much more attractive to traders in the financial world. Twitter and the social media's opinion is on the rise. This means a surplus of raw data with easy access. Companies all over the world has started to use twitter and readily available tweets to their benefit. Trading with social media is part of the trend. Although there are some drawbacks and shortcomings. Noise and garbage is one of them. It's difficult to accurately sort through all the data and get only the information relevant for your use. Even if your right 80% of the time, the last 20% can prove devastating. [Stevenson, 2012]

Sentiment and Momentum

file:*SSRN-id1479197.pdf* cition:[Doukas et al., January 10, 2010]

Not Twitter. Intra-day transaction data. Sentiment affects the profitability of price momentum strategies.

Use of sentiment can predict changes and momentum in the market. Bad news in an optimistic period creates cognitive dissonance in the small investors. This impacts the market by slowing down the selling rate of loosing stocks. [Doukas et al., January 10, 2010, p29]

Sentiment broadly refers to the state of mind a person has. Whereas negative of positive. Based on the current state of mind the person will do optimistic or pessimistic choices. A positive state of mind leads to optimistic judgements of future events. And a negative state of mind leads to pessimistic judgements. [Doukas et al., January 10, 2010, p4]

Further we can see that optimistic sentiment has a 2% monthly average return. While the investor sentiment is pessimistic we see a drastic reduction in returns. Down to 0.34%.[Doukas et al., January 10, 2010, p5] After optimistic periods it is indicated that the monthly return is reduced to -0.49%. On the contrary there is no equivalent change after a pessimistic period. [Doukas et al., January 10, 2010, p6-7] Momentum profits are only significant when the sentiment is optimistic. [Doukas et al., January 10, 2010, p29]

Is Trading with Twitter only for Twits?

Document Description: Blog post that describes the findings of the atricle [todo art:ref].

The article has developed a strategy for trading stocks based on the bullishness of the tweet. [todo glossary bullishness] Bullishness as I understand it is the same as the negativity of the tweet.

The article bases it's findings on three factors. The holding time of a stock (the time from you buy it until it's sold). The history of x days (how many of the past days are used to determine the tweet signal[todo glossary tweet signal]). And the number of picks (how many stocks you hold a any given time).

It is also indicated that The main article has some good information about how tweets are built up. (Dollar-tagging for representation of a given stock, $AAPL)

Has a good figure of the system.

Indicates that the message volume and trade volume are related.

RefArticle: **??** Twitter mood Predicts the Stock Market.

Tags: buy/sell-signals, tweet signals, dollar-tagged, OpinionFinder, GPOMS,

From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series

The article uses polling data and two years of tweets as their data.

Basically a comparison of the opinion expressed on twitter and the opinion from phone enquiries.

Uses word counting to distinguish relevant tweets from the rest.

The twitter dataset is huge, typically billions of tweets.

Daily sentiment = positive tweets / negative tweets.

From binder:

2: frameworks, techniques and metrics.
3: automated new analysis.
4: same as 14.
5:
7: Techniques, noun scheme, named entities
8: Techniques, algorithms, Key Phrase Extraction Algorithm(KEA).
9: Maybe the text mining aspects of it.
11: More for the width to understand the topic.
12: Language models for financial news recommendations.
13: About the impact of Negative words.
14: Predict direction of a market stock.
15: text mining approaches.
16: text classification
18: wc, TM, expected returns, tactical asses allocation.
19: Textual analysis of stock market prediction using financial news articles.
20: same as 10.
22: Forecasting intra-day stock price trends with text mining techniques.
23: Techniques and applications for sentiment analysis.
24: Set of articles about sentiment.

25: More text mining and such.

Links

http://en.wikipedia.org/wiki/Sentiment_analysis
https://developers.google.com/prediction/docs/sentiment_analysis
http://www.slideshare.net/mcjenkins/how-sentiment-analysis-works
http://sentimentsymposium.com/
http://www.sentiment140.com/
http://www.cs.cornell.edu/people/pabo/research.html
http://www.cs.cornell.edu/home/llee/papers.html
http://www.cs.cornell.edu/home/llee/
http://cacm.acm.org/magazines/2013/4/162501-techniques-and-applications-for-sent
fulltext
http://text-processing.com/demo/sentiment/
http://www.forexsentiment.net/
http://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.
pdf
http://people.cs.pitt.edu/~wiebe/pubs/papers/emnlp05polarity.pdf

References

Luciano Barbosa and Junlan Feng. Robust sentiment detection on twitter from biased and noisy data. 2010. Coling 2010: Poster Volume, pages 36–44, Beijing, August 2010.

Lee Becker, George Erhart, David Skiba, and Valentine Matula. Avaya: Sentiment analysis on twitter with self-training and polarity lexicon expansion. 2013. Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Seventh International Workshop on Semantic Evaluation (SemEval 2013), pages 333–340, Atlanta, Georgia, June 14-15, 2013. c 2013 Association for Computational Linguistics.

Gregory W. Brown and Michael T. Cliff. Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, 11(1):1 – 27, 2004. ISSN 0927-5398. doi: http://dx.doi.org/10.1016/j.jempfin.2002.12.001. URL http://www.sciencedirect.com/science/article/pii/S0927539803000422.

John A. Doukas, Constantinos Antoniou, and Avanidhar Subrahmanyam. Sentiment and momentum. January 10, 2010. Updated May 20, 2011. Available at SSRN: http://ssrn.com/abstract=1479197 or http://dx.doi.org/10.2139/ssrn.1479197.

A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: Understanding microblogging usage and communities. 2007. 9th WebKDD and 1st SNA-KDD workshop on web mining and social network analysis, 2007.

Annika Jubbega. Twitter as driver of stock price. Master's thesis, BI Norwegian School of Management, 2011.

Yung-Ming Li and Tsung-Ying Li. Deriving market intelligence from microblogs. *Decision Support Systems*, 55(1):206 – 217, 2013. ISSN 0167-9236. doi: http://dx.doi.org/10.1016/j.dss.2013.01.023. URL http://www.sciencedirect.com/science/article/pii/S0167923613000511.

Jianfeng Si, Arjun Mukherjee, Bing Liu, Qing Li, Huayi Li, and Xiaotie Deng. Exploiting topic based twitter sentiment for stock prediction. 2013. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 24–29, Soa, Bulgaria, August 4-9 2013. c 2013 Association for Computational Linguistics.

Michael Speriosu, , Nikita Sudan, Sid Upadhyay, and Jason Baldridge. Twitter polarity classification with label propagation over lexical links and the follower graph. 2011. Proceedings of EMNLP 2011, Conference on Empirical Methods in Natural Language Processing, pages 53–63, Edinburgh, Scotland, UK, July 27–31, 2011. c 2011 Association for Computational Linguistics.

Timm O. Sprenger and Isabell M. Welpe. Tweets and trades: The information content of stock microblogs. December 2010.

Alexandra Stevenson. The social media stock pickers, Oct 23 2012. URL http://search.proquest.com/docview/1114502067?accountid=12870. Copyright - Copyright Financial Times Ltd. 2012. All rights reserved.; Last updated - 2012-10-23.

Xue Zhang, Hauke Fuehres, and Peter A. Gloor. Predicting stock market indicators through twitter “i hope it is not as bad as i fear”. *Procedia - Social and Behavioral Sciences*, 26(0):55 – 62, 2011. ISSN 1877-0428. doi: http://dx.doi.org/10.1016/j.sbspro.2011.10.562. URL http://www.sciencedirect.com/science/article/pii/S1877042811023895. ¡ce:title¿The 2nd Collaborative Innovation Networks Conference - COINs2010¡/ce:title¿.