Evaluation of Methods and Techniques for Language Based Sentiment Analysis for DAX 30 Stock Exchange – A First Concept of a "LUGO" Sentiment Indicator

[1]Artur Lugmayr and [2]Gerhard Gossen

[1]Tampere University of Technology, Tampere, Finland
[2]L3S Research Center, Leibniz-University Hannover
artur.lugmayr@tut.fi, gossen@l3s.de

Abstract. Social media companies are famous for creating communities, or their companies for IPOs. However, social media are as well utilized in stock exchange trading and for product promotion of securities of financial investment companies. Especially stock exchange trading is many times based on sentiment, thus fast spreading rumors and news. Within the scope of this publication, we aim at an evaluation of potential methods and techniques for language based sentiment analysis for the purpose of stock exchange trading. Within the scope of this publication we evaluate a possible technique to obtain a technical indicator based on social media, which should support investment decisions. We present a basic experimental setup and try to describe the *LUGO Sentiment Indicator* as possible tool for supporting investment decisions based on a social media sentiment analysis.

Keywords: sentiment analysis, sentiment indicators, stock exchange, securities, investment business information management.

1 Introduction

There exist many theories and models for describing the potentials of predicting the stock market behavior. Macroeconomic models in the 1950s, suggested a gradual move of the market in business cycles [12], but failed to support concrete stock exchange investment decisions. With the advent of the *Efficient Market Theory (EMT)* between the 50s and 60s, the believe that the current stock price fully includes the total market information and reacts rationally to changes was born ([10] and [9]). This theories where challenged in the 80s, when scholars contradicted these ideas. They underlined the importance to consider bubbles, anomalies, volatility, crashes, and reactions on new information, overreactions, and investor sentiment [13]. These theories led to the development of the *Behavioral Finance Theory (BFT)*, which attempts to describe changes on the stock market through emotional changes and sentiment. With the advent of today's social media, a new tool for the investigation of investor sentiment is available, and has already been experimented with. Examples are StockTwits [19], or analysis of Twitter feeds [3].

Within the scope of this research work, we especially focus on the development of a German language based sentiment indicator for the *DAX 30 Performance Index*. The basic experimental setting and the approach to develop the indicator is presented. The final result of the study shall be a *LUGO Sentiment Indicator* that shall support investors in their decision process. The indicator should fulfill the following requirements:

- integration of suitable common indicator variables;
- consideration of time-lag influences on investor sentiment;
- adequate visualization of the indicator;
- provision of a benchmarking & validation possibility;
- basic components of the indicator should be a sentiment index, trend, and trend strength;
- simple probabilistic model and
- consideration of the sentiment seesaw (see [2]);
- optionally support/resistance zones and 'hot news' shall be included.

The indicator shall give insights into the stock exchange at least three (or eventually four) times per day: opening, midday, late, and optionally early. As results, the indicator should convey: sentiment [-1, 1]; trend (bear, bull, neutral); trend strength (in %); optionally resistance/support limits (in %); and optionally 'hot news'. Based on this indicator, derivative indicators showing the changes of the indicator over time are suggested, similar to existing MACD, RSI, or CCI trading indicators. Within the scope of this publication, the basic concept of the indicator shall be outlined, and relevant research works evaluated.

2 Related Work

The goal of *Sentiment Analysis*[1] of textual data is the extraction and aggregation of opinions, sentiments and attitudes held by document authors towards persons, events, or other topics discussed in the text. Algorithms typically use dictionaries of sentiment-indicating terms such as SentiWordNet ([8] and [1]). Machine learning algorithms that detect the relevant features of the text are also frequently used ([16], [21], and [22]). For a broad overview on Sentiment Analysis see [15], while the more recent developments are summarized in [20].

The easy availability of texts from social media has inspired a lot of research in sentiment analysis, especially as this data tends to reflect current events with a very short delay. Texts from social media such as Twitter and blogs has been used to predict global social trends (see [6] and [7]), stock market indicators (see [11],[4],[23],[24], and others), product sales [14], and asset value [23].

Other relevant works include sentiment analysis of the stock exchange, as e.g. [2]. A description of a set of indicators for sentiment analysis can be found in in [25]. Sentiment proxies are described in [2]. Benchmarking of social media sentiment indicators is addressed in [18] and [5].

[1] The term *Opinion Mining* is often used interchangeably.

3 Experiment Description and Architecture

3.1. Data Sources

The main data sources of the experiment are broker house newsletters, RSS market feeds, and stock exchange data. Traditionally broker houses publish a daily or bi-daily newsletter about the possible development of the DAX 30 performance index during this particular day. It shall support private investors in their decision making, and guide him through the complex process. Mostly these newsletters contain: general trend for the day (e.g. bear, bull, neutral); intraday resistance and support levels; textual description of the past behavior; textual description about the future behavior (mid-term, and intraday description); and charts illustrating the developments. We aim at the following basic requirements for data acquisition:

- data sources should be in German language;
- weighting of data sources according the reliability of the source;
- utilization of solely public and free available data sources;
- recording of a large enough textual test data-set for a representative analysis;

3.2. General Architecture

The idea of the experimental setup is depicted in Figure 1. The idea is to mash-up stock exchange data and obtain a sentiment indicator that should support investors in their decision making process. The aim is to obtain a real-time indicator, nevertheless an indicator that is calculated at open, mid-day, and close seems to be sufficient to assist in investment decisions. The indicator should consist of:

- *sentiment index:* numerical value representing the current market sentiment based on a set of variables;
- *market trend and trend strength:* trend direction (neutral, bullish, bearish) of the market, as well as the strength of the trend;
- *resistance/support levels:* optionally, the indicator should give insights into resistance and support zones for trading, to identify buying/selling levels.

The core of the system is a textual sentiment analysis component, which is mining textual input such as broker newsletters, RSS feeds, or other relevant news on market sentiment. This component is described in further depth at a later stage of this publication, but its main functionality is *sentiment score aggregation* and *validation*. Optionally we attempt to implement a textual mining block, which shall mine the textual inputs for hot news or resistance/support zones, which is contained in the textual input materials.

The second important architectural block is the sentiment aggregator, and its underlying sentiment model. The sentiment aggregator combines sentiment relevant indicators to a sentiment index based on a probabilistic model. Sentiment relevant indicators include macroeconomic data (e.g. consumer sentiment), implicit sentiment indicators (e.g. call/put ratio), explicit sentiment indicators (e.g. sentiment questioners), and technical indicators (e.g. volatility). The ideal mix of indicators is currently under investigation. The functionality of the sentiment aggregator is the calculation of the *sentiment index; trend and trend strength;* and the provision of other insights such as *resistance/support* and *hot market news*. However, the latter two are only optional in our current considerations.

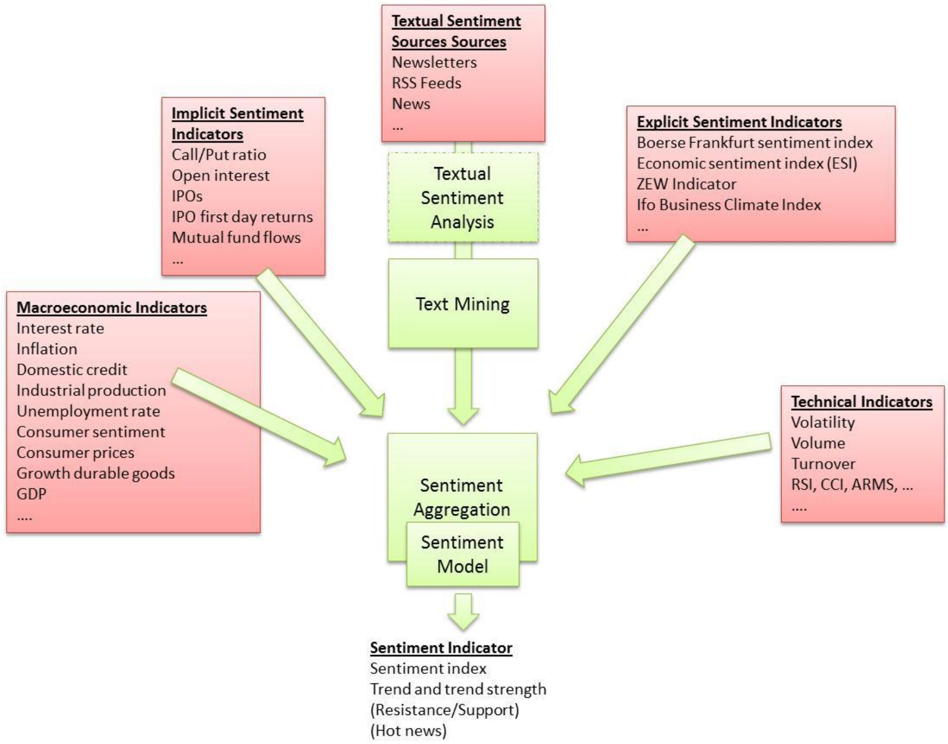

Figure 1. General description of the system architecture.

3.3. Goals of the Experiment

The experiment is divided into various different sub-goals. The goals that we are aiming at are:

1. *statistical testing of the validity of daily newsletters provided by brokers:* testing of the validity of daily trend descriptions of newsletters pushed to social media platforms from broker houses for the calendar year 2011 about this day's DAX 30 performance;
2. *selection, testing, and benchmarking of suitable common indicators:* selection of a set of indicators suitable for sentiment aggregation (yet not including the textual sentiment analysis) and creation of a sentiment model based on literature review and conducted experiments with common sentiment indicators (e.g. volatility);
3. *textual sentiment analysis implementation:* development of a textual sentiment indicator, dictionary, and test data set recording. Testing and

benchmarking the sentiment indicator on pre-recorded textual RSS news data-sets. Benchmarking of the indicator against a common sentiment indicator, and other available market data;

4. *sentiment aggregator development:* implementation of the sentiment aggregator, which combines textual sentiment analysis, common sentiment indicators, and other technical indicators;

4 Textual Sentiment Analysis of Social Media

To detect the sentiment of the market as described by market observers we use techniques from textual sentiment analysis. The input is a collection of texts such as newsletters, blog or forum posts, or news articles published before the start of trading. The result is a real value $S^t_{text} \epsilon [-1, +1]$ that captures the sentiment of the document collection. A value of -1 means that the sentiment indicates a downward turn (bear market), and a value of $+1$ an upward turn (bull market).

4.1. Goals of the Experiment

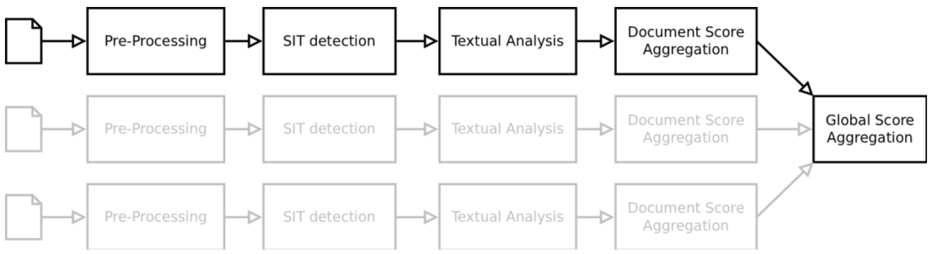

Figure 2. Architecture of the textual sentiment analysis component.

The textual sentiment analysis component consists of several steps which are described here in more detail (see also Figure 2). Many of these steps are typical for tasks in *Natural Language Processing (NLP)* and many high-quality libraries are freely available.

First the documents are pre-processed by stripping all irrelevant text, such as HTML boilerplate on web pages and standard headers and footers in newsletters. The text is then split it into sentences and tokens. As one of the languages we analyze (German) is a highly inflected language, we need to use *stemming* to remove all non-semantic prefixes and suffixes.

The sentiment analysis starts at occurrences of terms from a dictionary of words that typically express a positive or negative sentiment (*sentiment indicator terms*, *SITs*). Each such term is associated with its *sentiment polarity* $(-1, +1$ for positive/negative terms). Examples of SITs are “crash” (-1), “bear market” $(+1)$, or “rising” $(+1)$. We used a dictionary based on SentiWordNet [1], one of the largest freely available resources of sentiment indicator terms. However, because many terms typical for financial texts are missing from SentiWordNet, we augment the dictionary with terms we found in our data. We use a machine learning approach to incorporate modifications of a SIT through its local context such as negation, valence shifts [17], or conditional clauses into a *contextual sentiment score* for each sentence. Each sentence is classified using a *Support Vector Machine (SVM)*. The output is a *contextual sentiment score* for each sentence. We aggregate the sentence scores $s_{i,j}$ of a document i into a document score s_i calculating the average: $S_{text}^{t} = (\sum_{j=1}^{m} s_{i,j})/m$. Finally, we aggregate the document scores of a day to get the global textual sentiment S_{text}^{t} for day t using a weighted average $S_{text}^{t} = (\sum_{i=1}^{n} w_i s_i)/(\sum_{i=1}^{n} w_i)$ with weights w_i for individual documents to give documents created by professional market observers more weight than blog and forum posts, as they typically represent the market sentiment more correctly.

4.2. Evaluation

Typically the quality of the sentiment analysis is evaluated using a collection of documents (*gold standard dataset*) for which domain experts have provided the correct interpretation (positive/negative sentiment) by comparing the results of the algorithm with the labels provided by the experts. A different method is to measure the predictive value of the algorithm by correlating the extracted sentiment with the external indicators that we ultimately want to predict.
The advantage of the first method is that it directly measures the quality of the algorithm and allows easier iterative improvement as the algorithm. However, the creation of the gold standard dataset is very labor intensive and the labels provided are heavily biased towards the experts we chose (typically less than 70% agreement on labels provided by different domain experts). For these reasons and because no standard dataset for our domain exists, we will evaluate our algorithm using the second method and measure the quality of our algorithm by counting how often the predicted sentiment of the DAX30 corresponds to the actual development of that index.

5 Conclusions

Currently this work is still in progress, and we are preparing the data sets to perform a basic analysis. The resulting indicator can be solely one more indicator that is able to describe the happenings at the market, and is one additional parameter that influences the decision of an investor. One conclusion is already clear – the stock exchange is still <u>not</u> predicable, and social media will also <u>not</u> be able to predict the market! Social media will only be one new tool to assist investors in their decision making process.

Acknowledgements

The described work was partially funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 270239 (ARCOMEM).

References

[1] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC 2010*, volume 25, pages 2200–2204, 2010.
[2] Malcolm Baker and Jeffrey Wurgler. Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2):129–151, 2007.
[3] J. Bollen, H. Mao, and X.-J. Zeng. Twitter mood predicts the stock market. *ArXiv e-prints*, October 2010.
[4] Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, March 2011.
[5] Haiqiang Chen, Terence Tai-Leung Chong, and Xin Duan. A principal-component approach to measuring investor sentiment. *Quantitative Finance*, 10(4):339–347, 2010.
[6] Brendan O Connor, Ramnath Balasubramanyan, Bryan R Routledge, and Noah A Smith. From tweets to polls: Linking text sentiment to public opinion time series. In *International AAAI Conference on Weblogs and Social Media*, page 122â€“129, 2010.
[7] Gianluca Demartini, Stefan Siersdorfer, Sergiu Chelaru, and Wolfgang Nejdl. Analyzing political trends in the blogosphere. In *International AAAI Conference on Weblogs and Social Media*, 2011.
[8] Andrea Esuli and Fabrizio Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *LREC 2006*, volume 6, pages 417–422, 2006.
[9] Eugene F Fama. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2):383 – 417, December, 28-30 1970.
[10] Eugene F Fama and Kenneth R French. Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1):3 – 25, 1988.
[11] Eric Gilbert and Karrie Karahalios. Widespread worry and the stock market. In *International AAAI Conference on Weblogs and Social Media*, May 2010.
[12] John Maynard Keynes. *The General Theory of Employment, Interest and Money*. Atlantic Publishers & Distributors (P) Ltd., New Delhi, India, 2008.
[13] Edward R. Lawrence, George McCabe, and Arun Prakash. Answering financial anomalies: Sentiment-based stock pricing. *The Journal of Behavioural Finance*, 8(3):161–171, 2007.
[14] Yang Liu, Xiangji Huang, Aijun An, and Xiaohui Yu. ARSA: a sentiment-aware model for predicting sales performance using blogs. In *SIGIR 2007*, page 607â€“614. ACM, 2007.
[15] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135, 2008.

[16] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. In *Conference on Empirical methods in Natural Language Processing*, volume 10, pages 79–86. Association for Computational Linguistics, 2002.

[17] Livia Polanyi and Annie Zaenen. Contextual valence shifters. In James G. Shanahan, Yan Qu, and Janyce Wiebe, editors, *Computing Attitude and Affect in Text: Theory and Applications*, volume 20, pages 1–10. Springer-Verlag, Berlin/Heidelberg, 2006.

[18] Her-Jiun Sheu, Yang-Cheng Lu, and Yu-Chen Wei. Causalities between the sentiment indicators and stock market returns under different market scenarios. *International Journal of Business and Finance Research*, 4(1):159–172, 2009.

[19] StockTwits. http://wwwstocktwits.com/.

[20] Mikalai Tsytsarau and Themis Palpanas. Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, 24:478–514, 2012.

[21] Janyce Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. Learning subjective language. *Computational Linguistics*, 30(3):277–308, 2004.

[22] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity: An exploration of features for Phrase-Level sentiment analysis. *Computational Linguistics*, 35(3):399–433, 2009.

[23] Xue Zhang, Hauke Fuehres, Peter Gloor, JÃ¶rn Altmann, Ulrike BaumÃ¶l, and Bernd J. KrÃ¤mer. Predicting asset value through twitter buzz. In *Advances in Collective Intelligence 2011*, volume 113 of *Advances in Intelligent and Soft Computing*, pages 23–34. Springer Berlin / Heidelberg, 2012.

[24] Xue Zhang, Hauke Fuehres, and Peter A. Gloor. Predicting stock market indicators through twitter â€œI hope it is not as bad as I fearâ€? *Procedia – Social and Behavioral Sciences*, 26(0):55–62, 2011.

[25] M Zouaoui, G J M Nouyrigat, and F Beer. How does investor sentiment affect stock market crises? evidence from panel data. *Recherche*, 46:723–747, 2010.