Analysing Market Sentiments:
Utilising Deep Learning to Exploit Relationships within the Economy

Tobias Daudert
Insight Centre for Data Analytics, National University of Ireland Galway
tobias.daudert@insight-centre.org

Abstract

In today’s world, globalisation is not only affecting inter-culturalism but also linking markets across the globe. Given that all markets are affecting each other and are not only driven by fundamental data but also by sentiments, sentiment analysis regarding the markets becomes a tool to predict, anticipate, and milden future economic crises such as the one we faced in 2008. In this paper, an approach to improve sentiment analysis by exploiting relationships among different kinds of sentiment, together with supplementary information, from and across various data sources is proposed.

1 Introduction

Nowadays, modern societies and their welfare depend on market economies with the financial markets at the heart. This means that millions of people around the globe are affected by changes in the markets (Khadjeh Nassirtoussi et al., 2014). The financial markets are not only driven by fundamental factors (e.g. trends in GDP, inflation, employment, monetary and fiscal policy) but also by psychology-related factors such as public mood. This behaviour can be observed in economic bubbles which indicate irrational and emotional actions of the market participants (Khadjeh Nassirtoussi et al., 2014). In today’s increasingly complex global economy, more sophisticated approaches are needed to provide better insights and prevent future economic crises (Khadjeh Nassirtoussi et al., 2014). Therefore, it is important to understand what the markets, particularly the financial markets as a proxy for the market economies, are influenced by, how they react, and how strong the influence is.

Given the existence of different sentiments and alternative data all affecting each other, an experimental approach to exploit their relations is proposed. The aim of this approach is to derive additional information from different data types in order to improve sentiment analysis. In this paper, we outline the proposal and detail its components, focusing on its fine-grained construction to exploit contextual sentiments in microblogs data and news data individually, as well as their mutual relationships.

2 Background

Within the last 17 years, sentiment analysis aiming at measuring public mood in respect of the financial domain grew to an important field of research. Liu (2015) defines sentiment analysis, synonymously called opinion mining, as “the field of study that analyses people’s opinions, sentiments, appraisals, attitudes, and emotions toward entities and their attributes expressed in written text”. Market sentiment is defined as “the feeling or tone of a market, or its crowd psychology”.

Prior research has shown how sentiments and opinions can affect market dynamics, thus, making the financial domain a high-impact case study for sentiment analysis in text (Goonatilake and Herath, 2007; Van De Kauter et al., 2015). Sentiments are extracted from various sources of data, such as news. Within this source, one can find discussions regarding macroeconomic factors, company-specific reports, or political information, which

http://www.investopedia.com/terms/mmarketsentiment.asp
can be relevant to the market (Sinha, 2014). Good news tends to lift markets and increase optimism, bad news tend to lower markets (Schuster, 2003; Van De Kauter et al., 2015). As an example, Bollen et al. (2010) showed that changes in public mood reflect value shifts in the Dow Jones Industrial Index three to four days later. Additionally, evidence has been found that both quantitative measures (e.g. the quantity of news, market fluctuation) and qualitative indicators, (e.g. linguistic style and tone) affect investors’ behaviour (Loughran and Mcdonald, 2011; Takala et al., 2010; Tetlock et al., 2008). Given the link between sentiment and market dynamics, the analysis of public sentiment becomes a powerful method to predict the market reaction. Despite its active improvement over the past 17 years, the field of sentiment analysis with regard to the markets is not thoroughly explored. The relationship between textual and investor sentiment is complex and it is unclear to what extent they affect each other and how much each of them affects the markets (Kearney and Liu, 2014). The quality of interpreting such sentiment can determine the predictability of the financial markets. Thus, researchers have started targeting this problem lately. However, no well-rounded theoretical and technical framework exists (Khadjeh Nassirtoussi et al., 2014). Dedicated Natural Language Processing (NLP) approaches for analysing investor’s behaviour are non-existent at all (Kumar and Ravi, 2016). This is partially due to the interdisciplinary nature of the field since it combines economics, computer science and natural language processing (Khadjeh Nassirtoussi et al., 2014). NLP and computer science researchers generally lack the economics background and are too technically-focused whereas researchers with an economics background are lacking the knowledge to develop profound language processing tools. The current state of the art is based on lacking research in multiple sub-fields which are related but not exclusive to: 1) linguistic issues; 2) data type and data sourcing problems; 3) narrow research focus; 4) different sentiment types; and 5) generally accepted and available benchmark data, and evaluation methods. Below, each of the issues identified are detailed.

2.1 Linguistic Issues
Research effort is required to overcome and address complex linguistic issues such as sarcasm, irony, and poorly-structured and/or colloquial language. More syntax-based techniques such as parse-trees for pattern recognition require more attention (Khadjeh Nassirtoussi et al., 2014). Additionally, the process of content analysis is still in need of more authoritative and field-specific dictionaries (Kearney and Liu, 2014). On the side of semantics, aspect-level approaches can be used to provide a richer sentiment analysis (Schouten and Frasincar, 2016). Furthermore, the construction of more and improved customised ontologies are required to tackle the problem of text classification with a higher accuracy in each domain. In the case of the features used, researchers currently focus on word occurrence methods (Khadjeh Nassirtoussi et al., 2014; Kumar and Ravi, 2016). However, additional value could be added using abstraction for feature reduction or weighting schemes for semantic compression. All this leads to machine learning-based sentiment analysis approaches with low accuracy results (Takala et al., 2010).

2.2 Data Type and Data Sourcing Issues
Most of the present research is only based on specific data sources such as financial or corporate news. To overcome this threshold, it could be fruitful to combine “qualitative information from textual sentiment into equity asset pricing models [...] as publicly available documents or media articles may contain additional hard-to-quantify information” (Kearney and Liu, 2014; Kumar and Ravi, 2016). As textual sentiment in articles, blogs, or posts mostly reflects opinion, it would not only be of additional value to use these types of textual data together with fundamental figures but also to integrate technical signals such as the moving average convergence divergence (MACD) or the relative strength index (RSI) as additional features into machine learning approaches dealing with public sentiment. Only a few researchers (Butler and Kėšelj, 2009; Hagenau et al., 2013; Rachlin et al., 2007; Schumaker and Chen, 2009;
Schumaker et al., 2012; Zhai et al., 2007) have experimented with such a hybrid approach. Their findings are promising; most of the researchers could improve the accuracy of their classifiers by using additional, non-textual data. As an example, Zhai et al. (2007) used news, divided into company specific news and general market news, together with stock price information and stock price based calculated trading indicators. The results show the highest accuracy for the approach combining all available information as presented in Figure 1. Only Rachlin et al. (2007) could not achieve the highest accuracy using a joint approach; they achieved the best results only based on stock prices and from stock price generated indicators. In contrast to the proposed research in this paper, all the works mentioned are purely focusing on stock price prediction. Different data is used to achieve a higher accuracy on stock price forecasts instead of using the multitude of information to additionally improve the textual sentiment classifier itself. The textual sentiment, or media sentiment as described in section 2.4, used as input to their proposed models is not the same sentiment their predictions, as output, are referring to. For example, Schumaker et al. (2012) is claiming that a correlation between textual sentiment, news, and stock prices exists but a possible influence of stock prices on textual sentiment remains unexplored.

| Data sets      | Accuracy (%) |
|----------------|--------------|
| Price          | 58.8         |
| Direct news    | 62.5         |
| Indirect news  | 50.0         |
| Combined news  | 64.7         |
| Price & News   | 70.1         |

Figure 1: Prediction Accuracy as achieved by Zhai et al. (2007)

However, online postings bear challenges given their informal writing nature which significantly differs from professionally written media articles or corporate documents (Kearney and Liu, 2014). Text short in length (such as microblog messages) can be quite opinionated, dense in information, dependent on the modeling of economic context, and challenging to parse due to the different vocabularies used (Sinha, 2014). In addition, there are still qualitative information sources that have not been widely studied yet. Examples include business and political speeches, blogs, television news videos, and various social media platforms (Kearney and Liu, 2014). Many of the works to date are based on the specific news sources such as financial or corporate news (Kumar and Ravi, 2016).

2.3 Narrow Research Focus

Most of the work in this area is focused on specific companies (Kearney and Liu, 2014; Kumar and Ravi, 2016). Similar methodologies can be applied to other markets such as bonds, commodities, or derivatives which are potentially driven by public mood to a variable extent. Given the desire to address the general market sentiment, an overly narrow focus on individual markets is misleading since all markets are interfering with each other. Hence, it is necessary to not only study the U.S. markets but to consider additional markets linked to different languages such as German. To identify and track this dependence, short and long-term studies regarding sentiment changes are required. However, the majority of the research only considers the news article time of release (Kumar and Ravi, 2016).

2.4 Different Sentiment Types

Current research is not taking different types of sentiment and their interference into account. The different sentiments include public sentiments (e.g. Twitter, Facebook), media sentiments (e.g. Wall Street Journal, Financial Times), expert sentiments (e.g. analysts, investors), and company sentiments (e.g. Annual reports, CEO’s interviews). These sentiments are highly contextual and expressed in different formats (Khadjeh Nasiroussi et al., 2014).

2.5 Generally Accepted and Available Benchmark Data

The availability of experimental data and benchmarks for sentiment analysis in the financial domain is currently very limited. Most of the researchers have accumulated their own datasets
(Khadjeh Nassirtoussi et al., 2014) leading to fragmented datasets and raising issues with the reproducibility of experiments. It is not only important to choose the right data sources and experimental approach but also to consider the fundamental groundwork in this area of research. The lack of benchmark datasets which was addressed by Kumar and Ravi (2016) leads to predictions useless for comparison. Here, a common framework would provide a remedy.

3 Methodology

In order to tackle some of the issues mentioned in Section 2, the proposed research is aiming at the exploitation of different sentiments contained in different data types and sources. At minimum, two different data types, microblogs (i.e. Twitter and Stocktwits) as well as financial news articles will be used, and it will be explored whether they can be used to improve the accuracy of each other’s sentiment analysis. The proposed methodology is presented in Figure 2. The goal of this fine-grained sentiment analysis pipeline is to utilise supplementary information from and across different data sources. Microblogs and articles published close to each other in time might contain additional information relevant to both and potentially increase the accuracy of the sentiment analysis of the two. In order to exploit data type specific features, dedicated sentiment analysis methods for microblogs as well as financial articles will be developed. I believe data specific sentiment analysers lead to the most accurate sentiments analysis.

3.1 Data Stream

The data that will be used as input for the proposed experiment comprises a collection of 100 million Tweets derived from tracking 1,588 keywords on Twitter, and 25,000 financial news articles from various sources such as Bloomberg, Investopedia, and CNN Money. Initially, experiments will be based on textual data in English language, however, German data will be considered at a later stage. The collected data is currently stored in MongoDB due to its flexibility with different data structures and, therefore, storage optimisation. The size of the collection amounts to 465 GB.

3.2 Processing Stream

The processing stream consists of 3 components: the Microblog Analyser, the News Analyser, and the Sentiment Aggregator which are detailed below.

3.2.1 Microblog Analyser

The Microblog Analyser is a tool dedicated to analysing sentiment in microblogs using deep learning (DL) (LeCun et al., 2015). It will be implemented using PyTorch and generate a document-level sentiment within the range from -1 (very negative) to +1 (very positive) with 0 as neutral, using aspect-level sentiment analysis. DL has been chosen based on the performance shown in recent research in the area of sentiment analysis related to finance such as during the Semeval 2017 Task 5 (Cortis et al., 2017), in which most of the top ranked teams applied either pure DL or hybrid approaches utilising DL together with lexica or ontologies. In addition, there are still many insights yet to become concerning aspect-level sentiment analysis (Wang et al., 2016). The decision in favor of PyTorch is made due to my experience in Python and its support by Facebook, Twitter, and NVIDIA. Initially, publicly available Twitter gold standards (GS) such as the Semeval 2017 Task 5 one (Cortis et al., 2017) comprising 2494 Tweets will be used for training and testing. This GS will be replaced by a newly created, more problem-specific gold standard only focusing on entities tracked on Twitter.

3.2.2 News Analyser

The News Analyser is similar to the Microblog Analyser. It is dedicated to analyse sentiment in financial news articles also using deep learning implemented with PyTorch, for the same reasons as described in 3.2.1. Financial news articles will be annotated manually for the creation of a gold standard.

---

3http://www.twitter.com/
4http://www.stocktwits.com/
5http://www.bloomberg.com/
6http://www.investopedia.com/
7http://www.mongodb.com/
8http://pytorch.org/
standard providing training and testing data for the News Analyser.

### 3.2.3 Sentiment Aggregator

The Sentiment Aggregator uses the document-based sentiment generated by the Microblog Analyser and the News Analyser, as well as the textual data as input and generates the final sentiment for each document. It is independent from both Analysers, as they are highly specialised and dedicated to their type of textual content, therefore, they do not take the environment into account, whereas the Sentiment Aggregator identifies relationships between different documents and adjusts the given sentiments based on their relevance, the relationships between documents, and contextual information. It takes into consideration previous documents and their sentiments, thus, not only the currently given textual data. As shown in Figure 2, the sentiment of the German Tweet "Das neue iPhone ist eine Schande" derived by the Microblog Analyser is -0.243. Passing all this information to the Sentiment Aggregator, the sentiment stays unchanged since there is no additional knowledge to this point. However, at the time the second document is processed, there is additional knowledge about the previous Tweet existing. Therefore, the Sentiment Aggregator takes this knowledge into account and amends the by the News Analyser purely textual-derived sentiment from -0.343 to -0.732. The Sentiment Aggregator uses natural language processing methods to recognise and extract entities, identify relationships and determine the importance of each document. To achieve this highly complex task, a combination of knowledge bases and machine learning will be used. The knowledge bases will be based on the terms being tracked on Twitter. Besides own developments regarding the weighting and linking of information, the GATE TwitIE will be used since it is easily implementable and still one of the most powerful named entity recognition tools existing.9

9https://gate.ac.uk/wiki/twitie.html
4 Future Work

My future research will focus on targeting the previously-mentioned issues in the field of sentiment analysis for market sentiment, using the financial domain as a powerful proxy. I intend to implement the methodology proposed in this paper, revise and improve it in parallel with the additional knowledge obtained during its implementation. I also plan to extend my experiment with additional types of data containing different sentiments such as stakeholder interviews (e.g. Mario Draghi, Janet Yellen, Lloyd Blankfein, or Tim Cook) or financial reports. To increase the number of usable data sources, multi-linguality will be considered in future extensions as well. Furthermore, fundamental data and technical signals will be included in my model; this is intended to modulate the markets as accurately as possible.

References

Johan Bollen, Huina Mao, and Xiao-Jun Zeng. 2010. Twitter mood predicts the stock market. pages 1–8.

Matthew Butler and Vlado Kešelj. 2009. Financial Forecasting Using Character N-Gram Analysis and Readability Scores of Annual Reports. volume 5549, pages 39–51.

Keith Cortis, Andre Freitas, Tobias Daudert, Manuela Huerlimann, Manel Zarrouk, Siegfried Handschuh, and Brian Davis. 2017. SemEval-2017 Task 5: Fine-Grained Sentiment Analysis on Financial Microblogs and News. Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), pages 519–535.

Rohita Goonatilake and Susantha Herath. 2007. The Volatility of the Stock Market and News. Journal of Economics and Finance, 28(2):252–259.

Michael Hagenau, Michael Liebmann, and Dirk Neumann. 2013. Automated news reading: Stock price prediction based on financial news using context-capturing features. Decision Support Systems, 55(3):685–697, 6.

Colm Kearney and Sha Liu. 2014. Textual sentiment in finance: A survey of methods and models. International Review of Financial Analysis, 33(Cc):171–185.

Arman Khadjeh Nassiroussi, Saeed Aghabozorgi, Teh Ying Wah, and David Chek Ling Ngo. 2014. Text mining for market prediction: A systematic review. Expert Systems with Applications, 41(16):7653–7670.

B. Shravan Kumar and Vadlamani Ravi. 2016. A survey of the applications of text mining in financial domain. Knowledge-Based Systems, 114:128–147.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. Nature, 521(7553):436–444, 5.

Bing Liu. 2015. Sentiment Analysis. Cambridge University Press.

Tim Loughran and Bill Mcdonald. 2011. Barron’s Red Flags: Do They Actually Work? Journal of Behavioral Finance, 12:90–97.

Gil Rachlin, Mark Last, Dina Alberg, and Abraham Kandel. 2007. ADMIRAL: A data mining based financial trading system. Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2007, (Cidm):720–725.

Kim Schouten and Flavius Frasincar. 2016. Survey on Aspect-Level Sentiment Analysis. IEEE Transactions on Knowledge and Data Engineering, 28(3):813–830, 3.

Robert P. Schumaker and Hsinchun Chen. 2009. Textual analysis of stock market prediction using breaking financial news. ACM Transactions on Information Systems, 27(2):1–19.

Robert P. Schumaker, Yulei Zhang, Chun Neng Huang, and Hsinchun Chen. 2012. Evaluating sentiment in financial news articles. Decision Support Systems, 53(3):458–464.

Thomas Schuster. 2003. Meta-Communication and Market Dynamics. Reflexive Interactions of Financial Markets and the Mass Media. SSRN eLibrary, (July).

Nitish Sinha. 2014. Using big data in finance: Example of sentiment extraction from news articles. pages 6–8.

Pyry Takala, Pekka Malo, Ankur Sinha, and Oskar Ahlgren. 2010. Gold-standard for Topic-specific Sentiment Analysis of Economic Texts. Working Paper, pages 2152–2157.

Paul C. Tetlock, Maytal Saar-Tsechansky, and Sofus MacSkassy. 2008. More than words: Quantifying language to measure firms’ fundamentals. Journal of Finance, 63(3):1437–1467.
Marjan Van De Kauter, Diane Breesch, and Veronique Hoste. 2015. Fine-grained analysis of explicit and implicit sentiment in financial news articles. Expert Systems with Applications, 42(11):4999–5010.

Yequan Wang, Minlie Huang, Li Zhao, and Xiaoyan Zhu. 2016. Attention-based LSTM for Aspect-level Sentiment Classification. pages 606–615.

Yuzheng Zhai, Arthur Hsu, and Saman K Halgamuge. 2007. Combining News and Technical Indicators in Daily Stock Price Trends Prediction. In Advances in Neural Networks ISNN 2007, volume 4493, pages 1087–1096. Springer Berlin Heidelberg, Berlin, Heidelberg.