Leveraging News Sentiment to Improve Microblog Sentiment Classification in the Financial Domain

Tobias Daudert, Paul Buitelaar, Sapna Negi

Insight Centre for Data Analytics, Data Science Institute, National University of Ireland, Galway

`firstname.lastname@insight-centre.org`

Abstract

With the rising popularity of social media in the society and in research, analysing texts short in length, such as microblogs, becomes an increasingly important task. As a medium of communication, microblogs carry peoples sentiments and express them to the public. Given that sentiments are driven by multiple factors including the news media, the question arises if the sentiment expressed in news and the news article themselves can be leveraged to detect and classify sentiment in microblogs. Prior research has highlighted the impact of sentiments and opinions on the market dynamics, making the financial domain a prime case study for this approach. Therefore, this paper describes ongoing research dealing with the exploitation of news contained sentiment to improve microblog sentiment classification in a financial context.

1 Introduction

In an increasingly complex world in which information is almost instantly available and flows with nearly no limits, people are facing a magnitude of information not always objective or unbiased. Especially with the increasing popularity of Twitter, short texts dense in information and usually rich in sentiment are becoming increasingly relevant when it comes to the education of people through news stories (Mitchell and Page, 2015). In 2017, close to 23% of the people worldwide preferred social media as the selected gateway to digital news content. The importance of digital news is also emphasized by the increasing amount of time per day that adults in the U.S. spent with digital media which grew from 214 to 353 minutes

in the last 6 years. Within the same period, the amount of time adults spent with traditional media decreased from 453 to 360 minutes. However, traditional news are still important and at minimum as influential as digital media; in 2017, 32% of the people worldwide accessed digital news directly on a news website[1] [2].

Given the importance of both news sources (i.e. microblogs and news stories), their similar instantaneous availability, and their topic intersections, it becomes relevant to study how news articles and microblogs affect each other and, in more detail, how the sentiments contained in both affect each other. This paper presents ongoing research which is dealing with this question and utilises the news-contained sentiment to improve microblog sentiment classification. This research is built on the hypothesis that sentiment carried in news articles will eventually affect the sentiment expressed in microblogs (e.g. a person develops an opinion after reading a news article and later utilises microblogs to express it).

2 Background

As the world gets increasingly connected, factors affecting peoples' sentiment rise. Research has shown the link between sentiments and the market dynamics making the financial domain an important area for sentiment analysis in text (Van De Kauter et al., 2015; Kearney and Liu, 2014). Sentiments are contained in multiple forms of text, such as news and microblogs. News can convey information regarding macroeconomic factors, company-specific reports, or political information, which can be relevant to the market

---

[1] `https://www.statista.com/statistics/565628/time-spent-digital-traditional-media-usa/`
[2] `https://www.statista.com/chart/10262/selected-gateways-to-digital-news-content/`

(Sinha, 2014). Good news tend to lift markets and increase optimism, bad news tend to lower markets (Schuster, 2003; Van De Kauter et al., 2015). Not only news are an important factor for the markets. In 2011, Bollen et al. (2011) showed that changes in public mood reflect value shifts in the Dow Jones Industrial Index three to four days later. Therefore, analysing financial text becomes progressively important and research is shifting its attention towards this topic. An example, is the Semeval 2017 Task 5 which focused on fine-grained sentiment analysis on financial microblogs in subtask 1, and news headlines in subtask 2. Given the relevance and availability of microblogs and news, both are an intriguing source for sentiment analysis. Although the existing interest in media-expressed sentiment, most of the research focuses on news, particularly news titles (i.e headlines) (Nassirtoussi et al., 2014; Kearney and Liu, 2014). This is due to three reasons, 1) annotating news titles requires less effort than full articles; 2) news titles summarise the main points of the news article, thus, it should reflect the article's content (Peramunetilleke and Wong, 2002; Huang et al., 2010); and 3) news titles are written in a way to attract readers' attention, hence, having a high load of emotional and sentimental content (Strapparava and Mihalcea, 2007; Meyer et al., 2017; Corcoran, 2006). Despite the growing attention to the sentiment classification of news, and news headlines in specific, datasets dealing with financial news titles are still rare; especially regarding a fine-grained classification in contrast to only polarity. Overall, common sources for sentiment analysis are K-10 fillings, news articles, and microblogs. A dataset linking microblogs to news articles is not existing, to the best of our knowledge. Thus far, no work investigated financial sentiments further, excluding creating new data sets, lexicons, and rule lists and applying them to retrieve better sentiment classifications.

Approaches for sentiment analysis can be grouped into knowledge-based techniques and statistical methods. Although easily accessible, knowledge-based techniques are hindered by their inability to correctly interpret semantics and nuanced concepts (Cambria et al., 2017). In the case of the statistical methods, common approaches include support vector machines (SVM) and artificial neural networks (ANN). In parallel with the momentum of artificial neural networks, the types of clas-

sifiers used in the area of sentiment analysis are shifting. While Nassirtoussi et al. (2014) report on a vast majority of the literature using SVMs and scarcely ANNs, participants of the 2017 Semeval task 5 (Cortis et al., 2017) have substantially used ANNs as well as other deep learning approaches such as Recurrent Neural Networks or Convolution Neural Networks. Artificial neural networks are powerful in terms of prediction accuracy and offer a high flexibility; however, they are arguably the least transparent models (Strumbelj et al., 2010). As interpretability comes at the cost of flexibility, accuracy, or efficiency (Ribeiro et al., 2016), the consideration of the trade-off between classifier types becomes essential. This is notably the case for automated trading and medical diagnosis (Caruana et al., 2015) where the application of a "black box" algorithm can pose a significant risk. Although potentially less powerful, machine learning approaches based on simpler algorithms allow for the identification of the components responsible for the achieved prediction.

This work is inspired by the proposal described in Daudert (2017); specifically, it exploits the idea of utilising a combination of multiple sentiments. Our work conducts the first step into a new direction by focusing on the achievement of a superior sentiment classification trough the exploitation of the relations between different sentiments.

3 Methodology

The methodology implemented in this work is based on two foundations: the creation of a suitable dataset and its use in a Machine Learning (ML) prediction model.

The dataset is a vital component of this research. As the goal is to leverage relations of sentiments in both data types, news and microblogs, a dataset linking and combining both data is compulsory. Due to its novelty, it became necessary to choose a microblog dataset and then create a novel complimentary news dataset covering the same period and entities. With these complementary datasets, the following step consisted in linking them, enriching the pre-existing microblog dataset with 1) information regarding the related news for each microblog, and 2) the related-news sentiment.

The ML algorithm chosen for this task is a Support Vector Machine (SVM). This SVM is trained and tested with the datasets explicitly created for this work, with the aim of exploring whether news-

Minimum	Maximum	Average	Total
1	40	6.4	106

Table 1: Number of news in dataset MRN per entity. News articles can cover more than one entity.

Type	Dataset M	Subset A	Subset B
Training	1990	370	221
Test	498	93	56
Total	2488	463	277

Table 2: Number of microblogs per dataset.

contained sentiment can bring an advantage to microblog sentiment classification. To investigate this, we compare a classification purely based on the microblog messages with a classification based on microblog messages as well as news sentiment.

3.1 Data

This research makes use of two datasets; an existing microblog dataset and a novel news dataset created for this work. On one hand, it utilises the microblog dataset (M) from the Semeval 2017 Task 5 - subtask 1 (Cortis et al., 2017). This dataset contains 2,488 microblogs retrieved from Twitter[3] collected between March 11th and 18th 2016 as well as StockTwits[4]. Particularly, the dataset contains the microblog message, source, as well as a manually assigned cashtag (e.g. '$AAPL' for Apple Inc), span, and continuous sentiment score. On the other hand, the newly created microblogs-related news dataset (MRN) consists of 106 news, specifically, it contains the news' titles, urls, time and date, a sentiment score, and, if available, a description for each news. The news data was gathered from multiple sources such as `wsj.com` or `bloomberg.com`.

To be selected for this dataset, two criteria have to be satisfied to ensure the relatedness to dataset M. (1) Only news published between March 11th and 18th 2016 have been considered, and (2) each news has to deal with at least one company mentioned in dataset M. To fulfill the second criteria, we automatically extracted all 871 distinct cashtags from dataset M and used those to retrieve the respective company names using Stocktwits. With this list of cashtags and the associated company names, all news have been filtered and only news containing at least one of the 871 cashtags

and/or company names have been kept. Overall, the MRN dataset covers 18 unique entities in 463 microblogs. Further information is given in Table 1.

In the following step, all news in MRN have been annotated with a sentiment score. The dataset was presented to two annotators who assigned, based on title and description, a sentiment score within the five classes [-1.0, -0.5, 0.0, 0.5, 1.0], with 0.0 as neutral. In cases when the two annotators did not agree on a particular sentiment score, an expert decided the most appropriate rating. The inter-annotator agreement on all classes achieved a Cohen's Kappa coefficient of 0.52; when using an aggregation of 3 classes [-1.0, 0.0, 1.0] it achieved a value of 0.61.

Preliminary experiments have shown that the datasets were too small to achieve adequate results on a continuous sentiment scale, thus, it became necessary to increase the data per class and decrease the possible number of classes. Therefore, sentiment scores in dataset M have been processed to cluster data in three classes by transforming sentiment scores above and lower 0.0. Scores larger than 0.0 became 1.0; sentiment score smaller than 0.0 became -1.0.

3.2 Assigning a News Sentiment to Microblogs

With the knowledge that all news in dataset MRN are dealing with companies covered by a minimum of one microblog in dataset M, a question is raised on how to convey the news-contained sentiment to each microblog. We choose an entity based approach and assume that within a certain period, sentiments regarding the same entity should be similar across different data sources. Therefore, one news sentiment was calculated for each entity mentioned in dataset MRN. The sentiments for all news dealing with the same entity have been added together and then divided by the total number of news dealing with this entity.

Although each news in dataset MRN is linked to at least one microblog in dataset M, not all microblogs have a relation to at least one news. Dataset MRN covers 18 unique entities whereas dataset M covers 871 unique entities. Thus, we created two subsets of dataset M according to the microblogs' relation to dataset MRN (Table 2). Subset A contains all microblogs (from Twitter and Stocktwits) which have a relation to at least

	Subset A		Subset B	
Measure ╲ Features	MT	MT & NS	MT	MT & NS
Micro F1-Score	0.7097	**0.7312**	0.7321	**0.75**
Macro F1-Score	0.7874	**0.8055**	0.651	**0.6938**
Weighted F1-Score	0.6997	**0.7244**	0.7111	**0.7406**
Euclidean Distance	10.3923	**10**	7.746	**7.4833**
Mean Error Squared	1.1613	**1.0753**	1.0714	**1**

Table 3: Scores as obtained by the SVM model for subset A and subset B. MT abbreviates the message text, and NS the news sentiment.

Measure ╲ Features	MT	MT & NS-3
Micro F1-Score	0.7711	**0.7731**
Macro F1-Score	0.493	**0.4948**
Weighted F1-Score	0.76035	**0.7626**
Euclidean Distance	20.8567	**20.7605**
Mean Error Squared	0.8735	**0.8655**

Table 4: Scores obtained by the SVM model for dataset M. MT abbreviates the message text, and NS-3 the news sentiment aggregated into the 3 classes [-1.0, 0.0, 1.0].

one news (e.g. the same entities are present in both the microblog and the news article); subset B contains only the microblogs from Twitter which have a relation to at least one news. Subset B is necessary as the stocktwits were not specifically collected in the same period as the tweets. All three datasets have been randomised and split into a training set of 80% and a test set of 20% to avoid any bias from the structure of the Semeval data.

3.3 Preprocessing the Data

To prepare the textual data for the ML model, the following preprocessing steps were performed:

1. URLs were replaced with $< url >$
2. Numbers were replaced with $< number >$
3. With $WORD$ representing the original hastag:

   (a) hastags in upper case were replaced with $< hashtag > WORD < allcaps >$
   (b) the remaining cases were replaced with $< hashtag > WORD$

4. Smileys and emoticons were replaced with a description (e.g.☺ becomes $slightly\_smiling\_face$) [5]

The processed text was then transformed into an unigram tf-idf representation.

3.4 Experimental Setup

The experiments use a SVM employing a linear kernel. This decision was made based on the approaches of the best teams at the Semeval 2017 Task 5 - Subtask 1. LiblinearSVC was chosen for this task (Pedregosa et al., 2012). The performance is evaluated using F1-Scores, the Euclidean distance, and the mean error squared. The SVM model is trained and tested in two distinct approaches: (1) a feature matrix representing the microblogs' messages; (2) a feature matrix representing the microblogs' messages enriched with the assigned news sentiment for each microblog. The default settings were employed, except for the maximum number of iterations which is decreased to 500 and the random state which is set to 42.

4 Results

Table 3 presents the classification results on subset A and subset B. As the table shows, utilising the news sentiment improves all measures. The weighted F1-Score for subset A is increased by 3.51% and the Euclidean distance is decreased by 7.4%; for subset B the F1-Score increases by 4.15% and the Euclidean distance is decreased by 6.66%. This suggests that the news sentiment is benefiting the classification. Applying this classification on dataset M shows similar results (Table 4). Although, it is containing unre-

lated stocktwits collected at a different period, and having only 18.6% of the microblogs with an assigned news sentiment, all measures improve; the weighted F1-Score improves 0.3% and the Euclidean distance 0.46%. However, for dataset M, it is important to notice that to make a measurable difference, the news sentiments have been aggregated into the 3 classes [-1.0, 0.0, 1.0].

5 Conclusion and Future Work

This paper presents novel research leveraging news-contained sentiment to improve microblog sentiment classification. As there are no existing datasets for this task, we created a new dataset linking microblogs and news. Our current experiments show an improvement in sentiment classification across all used measures. This insight has the potential to change the future of sentiment analysis, shifting the focus from creating continuously larger datasets to cross-data linked approaches exploiting knowledge across multiple data types. In this work, we use manually annotated news sentiment to show its impact on microblog sentiment classification. Future works must consider the quality of automated news sentiment retrieval, therefore, identifying a threshold which determines whether news sentiment has an impact on microblog sentiment classification or not. Although the promising results, tangible points for improvement exist in the limited size of the dataset as well as the noise in the data. The microblog dataset applied is outdated by two years which hindered the retrieval of relevant news stories. Moreover, it contains messages unrelated to any event identified within the news; this is predominant for the stocktwits which were not collected within a defined period. Therefore, an important future contribution is the creation of a larger dataset, limited to a given period and ideally covering the same entities. Considering the linking of news and microblogs, we believe that more sophisticated approaches beyond the occurrence of identical entities will increase the impact of news sentiment on microblog sentiment classification. News and microblogs might deal with the same company but cover different topics which are not significantly related. Furthermore, this work does not consider the importance of the news articles' source; sources with a higher credibility might be more influential than others. Although this study is not sufficiently exhaustive

to provide a conclusive answer of the benefit of incorporating news-contained sentiment for microblog sentiment classification, it suggests the potential of leveraging knowledge from across multiple data sources and builds the foundation for upcoming research in the field of sentiment analysis.

Acknowledgments

References

Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.

Erik Cambria, Dipankar Das, Sivaji Bandyopadhyay, and Antonio Feraco. 2017. Affective Computing and Sentiment Analysis. In *A Practical Guide to Sentiment Analysis*, pages 1–10.

Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad. 2015. Intelligible Models for HealthCare. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15*, pages 1721–1730, New York, New York, USA. ACM Press.

Paul Corcoran. 2006. Emotional Framing in Australian Journalism. *Australian & New Zealand Communication Association International Conference*.

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.

Tobias Daudert. 2017. Analysing Market Sentiments: Utilising Deep Learning to Exploit Relationships within the Economy. In *Proceedings of the Student Research Workshop associated with RANLP 2017*, pages 10–16, Varna, Bulgaria.

Chenn Jung Huang, Jia Jian Liao, Dian Xiu Yang, Tun Yu Chang, and Yun Cheng Luo. 2010. Realization of a news dissemination agent based on weighted association rules and text mining techniques. *Expert Systems with Applications*, 37(9):6409–6413.

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

Bradley Meyer, Marwan Bikdash, and Xiangfeng Dai. 2017. Fine-grained financial news sentiment analysis. *Conference Proceedings - IEEE SOUTHEASTCON*.

Amy Mitchell and Dana Page. 2015. The Evolving Role of News on Twitter and Facebook. Technical report, pewresearch.org.

Arman Khadjeh Nassirtoussi, 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.

Fabian Pedregosa, Gal Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and douard Duchesnay. 2012. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(Oct):2825–2830.

Desh Peramunetilleke and R.K. Wong. 2002. Currency exchange rate forecasting from news headlines. *Australian Computer Science Communications*, 24(2):131139.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?". In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, pages 1135–1144, New York, New York, USA. ACM Press.

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. Technical report, Board of Governors of the Federal Reserve System (US).

Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. *Proc. of SemEval-2007*, (June):70–74.

Eri Strumbelj, Igor Kononenko IGORKONONENKO, and Stefan Wrobel. 2010. An Efficient Explanation of Individual Classifications using Game Theory. *Journal of Machine Learning Research*, 11(Jan):1–18.

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.