Entity-Level vs Document-Level Sentiment Analysis

Luca Furst and Jonathan Pearce
McGill University
Montreal, QC H3A 0G4
robert.furst@mail.mcgill.ca, jonathan.pearce@mail.mcgill.ca

Abstract

Recent work has shown the clear advantages of performing sentiment analysis at finer scales, particularly, sentence level sentiment analysis over document level. However, research with sentiment analysis of news articles and similar text and their relationship with financial markets is still commonly conducted at the document level. Using news articles from the technology and financial sectors we evaluate the significance of sentiment difference between document and named entity level analysis, with the goal of improving data available to sentiment based stock market predictors. Performing sentiment analysis at the entity level was chosen as the relationship between entities and stock performance of companies is more clear. We find that the difference in sentiment between document and entity level analysis is significant.

1 Introduction

In this report, we investigate the difference between document level sentiment scores and local named entity sentiment scores from news articles focusing on the technology and financial sectors. Previous research on the relationship between news sentiment and stock prices have utilized sentiment analysis at the document level. We believe that assuming the overall sentiment of a news article is an accurate reflection of the true current state of a company is too strong a claim, and limits the amount of information that can be provided to a stock market predictor. There are many potential factors that could mislead a document level sentiment analysis. News articles can contain revenue/progress comparisons between quarters for a company as well as comparisons between companies operating in the same domain. Articles can reference multiple companies in the process of negotiations or recall a company's actions in the past. In general there is no guarantee as to how specific or focused a news article will be on a single company, and sentiment analysis at the document level may not be effective. We believe that performing sentiment analysis at the entity level can improve the quality of sentiment data available to stock market predictors. Our goal is to demonstrate that in fact the difference in sentiment scores between document and entity level analysis is significant. This project is the first step in confirming that entity level sentiment based stock market predictors are more successful than previously studied document level predictors.

Our analysis is performed using news articles from the technology and financial sectors. For each article in our data set, we first perform document level sentiment analysis. Next, we extract all relevant entities and find their local context in the article. We calculate the sentiment of each entities' context and take the absolute difference between that and the document sentiment. Summing over every relevant entity in the article provides us with a measure of the difference between document and entity level sentiment score for the article.

A document level model assumes that all entities mentioned in a news articles have the same sentiment. However, if most of the article is one polarity, but mentions a single entity in the opposite polarity briefly, there exists a large difference between entity and document level sentiment scores. The document level model would be unable to see and utilize this difference to its' advantage, where as an entity level model would have a more clear idea of the information being expressed in the article.

2 Related Work

Sentiment analysis is a well-studied and active research area (Pang and Lee, 2008). Extensive work and discussion in sentiment analysis has focused around document level and sentence level analysis. Pang and Lee (2004) first showed that considering sentiment at a sentence level can improve document level sentiment performance. Using a unique approach based on minimum cuts applied to text, they were able to convert subjective sentences into input for a document level classifier. Further work in this area builds off this by exploring more advanced methods of utilizing text sentiment at the sentence level to help improve the task of document level sentiment analysis (McDonald et al., 2007; Yessenalina et al., 2010). This series of work demonstrated that considering sentiment at a finer level of detail in this case, sentence level, can create more accurate sentiment based models.

Applying natural language processing methods to problems related to financial markets, in particular utilizing sentiment from news articles in order to help predict stock price movements is a relatively well studied area. In fact a significant portion of this research has used document level sentiment analysis as opposed to methods that analyze text at finer detail (Feldman et al., 2011). More recently, Kazemian et al. (2016) showed that document level sentiment analysis can be utilized to direct a relatively successful securities trading strategy. They detail both their sentiment analysis technique, which considers news articles as a whole, as well as the development and evaluation of their trading methodology. This positive result from their work helped motivate use to pursue the idea that trading techniques such as theirs could be further enhanced with a more accurate, sentiment analysis performed at the entity level.

Our hypothesis is that entity level sentiment analysis will yield the best results when analyzing news articles in the technology and financial sectors, as opposed to the more traditional sentence level analysis. Work has been done that agrees with and supports this idea. Godbole et al. (2007) constructed a large scale framework to find and track entity sentiment over time. The data source for their work was news outlets and blogs, which shares similarities with our data set. The significant portion of their paper is validating the effectiveness of their sentiment lexicon generation method. The approach taken is to define separate lexicons for each of their determined sentiment dimensions (general, health, crime, sports, business, politics, media). Their reasoning for this design decision is that each distinct news area has different standards of opinion and sentiment, this idea helps support our idea that it is important to treat business (technology and finance) news on a separate basis and filter out sports, health and political news from our work. The sentiment analysis in their work is a direct application of their built sentiment lexicon, this leaves the door open for our work to validate that sentiment for technology and financial news will be different between the entity and document level, and to examine how significant this difference is.

3 Methods

For our project we used an already existing collection of news articles from Reuters ranging from June 7 to July 12 2017 (Chen, 2017). The news articles in this data set were from many domains including technology, sports, health, finance and politics. We choose to only focus on the articles from the technology and financial sectors as those articles would feature information about companies whose stock is traded publicly. After this filtering we were left with 23,395 articles in total to conduct our analysis on. The next step was to calculate the difference between document and entity level sentiment for each article in our data set. With our project involving named entity recognition and sentiment analysis, both of which are well studied areas of natural language processing and are commonly the focus of entire academic papers, we decided to use existing libraries for our work. This decision allowed us to focus on the hypothesis we were testing and provided us with tools that most likely would have been more accurate than any models we trained ourselves. Further, developing our own models would have required time consuming or expensive data annotation and for this project this was not feasible.

For each article we began by calculating the document level sentiment score. We choose to use NLTK's VADER library (Hutto and Gilbert, 2014) as our sentiment analyzer. VADER was developed and trained for accurate sentiment analysis on a small scale, particularly tweets and short messages, which makes it ideal for the entity based sentiment analysis. In their paper, Hutto and Gilbert show that VADER generalizes beyond short text successfully and performs well on longer pieces of text such as editorials from the New York Times as well. These results made us comfortable using VADER for the document level sentiment analysis.

The next step was to perform named entity recognition on each article. We choose to use python's spaCy library. Recently a number of named entity recognition models were rigorously evaluated, spaCy performed well with respect to entity recognition accuracy (Jiang et al. 2016). For each article's main body spaCy found all the named entities of the article. We reduced the entities down to only organization and person entities, as we deemed these were the two types of entities whose context would provide significant information about companies and their current status.

For each entity that remained we approximated the context associated with the entity. Our method was to take the n words before and after each entity while remaining within the sentence barrier, as context between sentences is more likely to be different. We performed our analysis on three different values of n: 5, 10, and 15. Finally, we calculated the sentiment for each entities context.

We use the evaluation function below to determine the difference of sentiment between entity-level and document-level sentiment. Given an article D and its sentiment, sent(D) as well as all relevant entities e in the article and their respective context sentiment, sent(e).

$$dif\_n(D) = \sum_{e \in entities(D)} \mid sent(e) - sent(D) \mid$$

4 Results

Of the 22,395 technology and finance articles, 6 were not encoded correctly, causing our pipeline to incorrectly parse entities and sentiments from the article. These six articles (0.03% of the data) were removed from the resulting analysis.

	dif_10	dif_15	dif_5	d_score
count	22389	22389	22389	22389
mean	1.7714	1.7173	1.8056	0.2487
std	2.0084	2.1523	1.9205	0.3657
min	0.0000	0.0000	0.0000	-0.9892
25%	0.4101	0.2408	0.5000	0.0000
50%	1.2054	1.0490	1.3249	0.1771
75%	2.4861	2.3837	2.5519	0.5106
max	37.4897	37.2789	37.1466	0.9978

Table 1: Sentiment score statistics

	ent_len
count	22383.000000
mean	7.740741
std	8.717174
min	0.000000
25%	3.000000
50%	5.000000
75%	9.000000
max	131.000000

Table 2: Entities per article statistics

Table 1 shows the document sentiment score (d_score) and the three differences obtained by the varying context window sizes. Table 2 displays the statistics about the number of entities extracted from each article. It is important to note that the median article had only 5 entities that fall within the categories of 'organization' and 'person'. We also see that the difference between the three window sizes is negligible: the sentiment difference from using a window size of 5 and 15 words is a score of 0.0883.

Sentiment scores per entity are assigned a value between -1 and 1. Using our evaluation function (summed distance to document sentiment), we can see that there is indeed a polarity difference between entity sentiment and document sentiment. Of over twenty-two thousand articles, our results show that on average, assigning document level sentiment scores to entities can cause an entity to be assigned an incorrect sentiment. An average sentiment difference of 1.7 with 5 entities per article implies each entity deviates on average .34 points from the document score. This is a significant deviation on a -1 to 1 scale.

One possible reason for the decreasing difference in sentiment score is that as the context size increases, it approaches the true length of the sentence. Financial news articles, unlike other news, tend to be neutral in language, meaning the entity will get a net neutral sentiment score.

5 Discussion

The results from our tests are quite promising and support our hypothesis. There is a significant difference between entity-level sentiment scores and total document sentiment scores. Therefore sentiment based stock market predictors would most likely see better performance by utilizing sentiment at the entity level.

There are of course places where our methodology could have been improved and there are opportunities for future work to expand and continue what we have discussed in this report.

Our method for finding the context window for each entity was not perfect. The small n word window around each entity did not guarantee that we captured the correct and necessary context for the entities in an article. A more sophisticated method for finding accurate context windows around entities would need to be explored before using entity based sentiment to develop a stock market predictor. One example includes using part of speech trees. Find all the noun phrases in a sentence, determine if it is a desired entity, and then find all the child adjective phrases of the entity node. This approach is stronger than the word-window based context approach. Whereas words surrounding an entity do not necessarily describe the entity, adjective phrases that are children of a noun phrase necessarily describe said noun phrase. This will guarantee the polarity assigned to the entity is indeed correct.

Co-reference resolution could have improved the results of our project. SpaCy would occasionally extract several permutations of the same entity ('samsung', 'samsung electronics co ltd'), and the sentiment score could deviate between them. We attempted to use the spaCy NeuralCoref module to fix this issue, however the length of time it took to execute on a small one paragraph article made it unfeasible to scale to the size of our data set. Researchers with access to more powerful computers are encouraged to incorporate this idea into their experiments.

As briefly discussed earlier in the report our decision to use existing named entity recognition and sentiment analysis libraries was a decision based on project feasibility. Although both VADER and spaCy have been shown to be powerful tools for general purpose experiments, it is very possible that models trained specifically for this task would provide better performance. In particular a sentiment analysis model trained exclusively on news articles from the technology and financial sector may be more accurate for our experiment, since this type of news has a different standard of opinion and sentiment than other types of news (Godbole et al., 2007). A comparison between our exact methodology (with spaCy and VADER) and our methodology with a new sentiment analysis model trained on annotated technology and financial news could serve as a good example of how leveraging domain knowledge in model training can provide better results.

Changing the evaluation metric could also provide further insight. We measured the distance between entity and document polarity, however this is not the only method that could have been used. Another possible evaluation criteria could weight the polarity score of an entity with its inverse frequency. This evaluation would capture a higher difference in less frequently used entities, increasing the belief that the article is not a just representation of all the entities it contains (Entities that appear once but have a different polarity than the rest of the document should stand out).

A meaningful continuation of our work would be to validate the idea that named entity level sentiment data is better for stock market prediction. This would require one to replicate a trading strategy from literature that is based on sentiment at the document level or possibly develop a new trading strategy. From there one could use times series data from a particular stock exchange to test and compare the two trading strategies to see how much of an advantage the finer level of sentiment detail provides.

Statement of Contributions

The project design was completed collaboratively between the two of us. Jonathan sourced and cleaned the data and began the implementation of

the entity level sentiment analysis. Luca continued the implementation and finished both types of sentiment analysis. Luca collected and analyzed data for our experiments. Jonathan completed the abstract, introduction and related work sections of the paper. Luca completed the results section. Methods and discussion were completed collaboratively.

References

Lizi Chen. 2017. Big Data and Machine Learning for Finance, News-Article-And-Full-Details-Dataset. *Github.com*

Ronen Feldman, Benjamin Rosenfeld, Roy Bar-Haim and Moshe Fresko. 2011. The Stock Sonar Sentiment Analysis of Stocks Based on a Hybrid Approach. *Proceedings of 23rd IAAI Conference on Artificial Intelligence*.

Namrata Godbole, Manjunath Srinivasaiah, and Steven Skiena. 2007. Large-scale sentiment analysis for news and blogs. In *Proceedings of the International Conference on Weblogs and Social Media*.

C.J. Hutto, Eric Gilbert. 2014. VADER: a parsimonious rule-based model for sentiment analysis of social media text. In *8th international AAAI conference on weblogs and social media*

Ridong Jiang, Rafael E. Banchs and Haizhou Li. 2016. Evaluating and combining named entity recognition systems. In *Proceedings of the Sixth Named Entity Workshop*, pages 21-27. Association for Computational Linguistics.

Siavash Kazemian and Shunan Zhao and Gerald Penn. 2016. Evaluating Sentiment Analysis in the Context of Securities Trading. *Association for Computational Linguistics*.

Ryan McDonald, Kerry Hannan, Tyler Neylon, Mike Wells, and Jeff Reynar. 2007. Structured models for fine-to-coarse sentiment analysis. *In Proceedings of the Association for Computational Linguistics (ACL)*, pages 432-439. Association for Computational Linguistics.

Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *In Proceedings of the Association for Computational Linguistics*, pages 271-278.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1-135.

Ainur Yessenalina, Yisong Yue, and Claire Cardie. 2010. Multi-level structured models for document-level sentiment classification. *In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1046-1056. Association for Computational Linguistics.