Evaluating Public Company's Finacial Statements with Natural Language Processing

Anish Patel
School of Finance
Texas A&M University
College Station, United States
anishmp9@gmail.com

***Abstract*— In this paper, I present my findings on my preliminary experiment, which focuses on the possibility of applying natural language processing to extract information from the qualitative sections of a company's financial statements. We will specifically be using sentiment analysis to discover positive or negative trends throughout the report and see if the trend can dictate potential stock growth/decline. The time span for calculating the growth/decline of the company's stock will be from the release date of the report to the present. Due to the constraints on time, this initial experiment can be viewed similarly to a minimum viable product. So, the report will conclude with a discussion on potential Natural Language Processing (NLP) techniques [1] for future experiments.**

Keywords- Sentiment Analysis, 10-K, Financial Statements, Natural Language Processing, Activision Blizzard

I. Introduction

Financial Statements are written records that convey the business operations and financial performance of a company. These are publicly issued. The idea being to enhance transparency between investors and companies. Financial Statements include the company's balance sheet, income statement, changes in stockholder equity, and cash flow statement [2]. These statements are regularly audited by a third party, government agency or accounting firm, to ensure accuracy. The specific document being parsed is a 10-K. A 10-K is one of the most comprehensive reports published by a company. Public companies are required by the U.S Securities and Exchange Commission (SEC) to file these reports annually [3]. Although this report can be extensive in length, it contains valuable insight on a company's history, organization structure, subsidiaries, executive compensations, potential risks/threats in the marketplace, and current operations.

Investors rely on these reports to help make better decisions when buying or selling a company. A mix of quantitative, qualitative, and technical analysis is used to get data-driven indicators with context. Then indicators are compared and questioned till a decision is made. Although this process is being automated more and more, so investors can get ahead of the market, it still lacks the accuracy and speed to consistently beat the market. Top hedge funds still vary tremendously from year to year in performance as well as from each other. Even with a relatively stable, positive trending market for the past 8 years, top funded firms like Mangrove Partners can bring in a 50.58 % return for the year of 2016, while averaging a 23.34% annual compounding return from 2014 to 2016 [4].

The following experiment aims to test a new approach to evaluating companies by hypothesizing future stock performance can be highly correlated with text classification of a 10-K. Although many text classification methods exist, this experiment specifically focuses on sentiment analysis (classifying some item as positive or negative) for its simplicity. This desire for simplicity also explains the use of rule-based models rather than a probabilistic model. Once a value of positive, negative, or neutral sentiment is assigned for a company, and it will be compared to its overall change in stock price from the release date of the report to the present. This will be further discussed in the Results and Discussion section.

With the time available, I was able to model one company. The company of focus for this experiment is Activision Blizzard. Activision Blizzard publishes video game content and software for consoles, personal computers, and mobile devices. The market structure is somewhat analogous to the energy industry. There are producers (game studios), who design and build the games; there are wholesalers (publishers), who purchase the rights and royalties to the games to then mass produce, market, and distribute the game; and there are retailers, who sell individual copies of the game to consumers at either brick and mortar stores or through online marketplaces like Steam [5]. Deciding on Activision Blizzard was based on inexplicable stock performance. The stock is currently in decline year over year, but their statements would indicate otherwise. With solid financials, brand loyalty, and new technology adaptation, current industry metrics would suggest investors to buy the stock [6].

Note financial statements do have limitations, even though they provide an immense amount of information about the company. The statements are ambiguous in their message, and as a result, the reader might draw vastly different conclusions about a company's performance. Therefore, this makes for a challenging artificial intelligence problem, since the heuristic function is not defined well.

The remainder of the paper is structured as follows: Section II describes the framework of the proposed system and the dataset that is used for experiment. Section III discusses the experimental results of the system with current tools, and Section IV discusses improvements and challenges for future experiments with similar aims. Code is attached in the same folder as the report in the Google Doc.

II. Overview of Framework

In this section, the outline of the system is described. Text was extracted from Activision Blizzard's 10-k, which can be found on the SEC's online database. To present the corpus in a structured manner, preprocessing was done before being further analyzed by NLP packages. Pre-processing ensures that the text will be prepared in simplified format that can be read and understood by the machine. After pre-processing, sentiment of text can be determined through sentiment classification. My Sentiment classification involved a subjectivity and polarity classification for the company. Subjectivity classification was used to judge whether the text was subjective or objective. Polarity classification was used to judge whether the texts where positive or negative.

A. *Dataset*

A total of 63,322 words were inputted into the corpus, and of that number 5154 were unique. The main text from Activision Blizzard's 10-k was found in the following sections: Business, Risk Factors, Management's Discussion and Analysis of Financial Conditions and Results of Operations, and Financial Statements and Supplementary Data.

Retrieving this data was relatively painless with the following Python packages: requests, BeautifulSoup4, and pickle. Requests and BeautfulSoup4 are web scraping packages, while pickle is package to open, edit, and close data files.

I also did some data exploration. Creating a Document Term Matrix, I was able to isolate the frequency of each word and applied the sort function to this matrix to find the most used words in the 10k. The results are recorded on Table 1.

Table 1.

Rank	Word	Frequency
1	Revenues	438
2	Tax	322
3	Net	316
4	Financial	250
5	Products	245
6	December	242
7	Income	226
8	Cash	225
9	Business	211
10	Million	197

B. *System Architecture*

The general architecture of the process is illustrated by the figure below. Detailed steps on pre-processing and sentiment classification are described in Section 1 and 2.

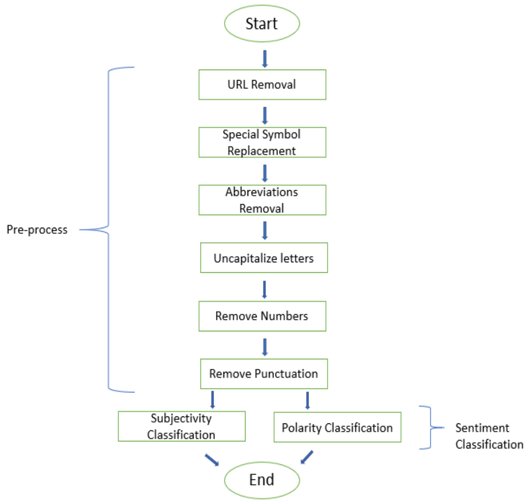

Figure 1: Flow Chart of Pre-processing and Sentiment Classification

1) *Pre-process*

Pre-processing aims to format the text in an organized, clear structure. This increases the machine understanding. This includes URLs removal, special symbol replacement, abbreviations and acronyms expansion, and capitalization. URLs and image linkage are removed from the text, since they are neutral based terms. Special symbols are replaced as spaces. The contraction, acronyms and abbreviations are removed due to ambiguity. Replacing capitalization with lower case letters is to prepare the corpus for a uniform pattern.

2) *Sentiment Classification*

When constructing a model for sentiment analysis, I used the Python package TextBlob. It was created by linguistic researchers who have already labeled the sentiment of words based on their domain expertise. Sentiment of words can vary based on where it is in a sentence. The TextBlob module allows us to take advantage of these labels. Each word in a corpus is labeled in terms of polarity and subjectivity. A corpus' sentiment is the average of these. Polarity ranges from -1 to + 1. For subjectivity, a zero would indicate a fact, while a one would indicate an opinion. Figure 2 graphs the subjectivity and the polarity.

While it was valuable to look at the overall sentiment of the corpus. I also wanted to understand the change in sentiment as one read through the 10-K. To accomplish this, I divided the corpus into 10 different sections, and gave each section a polarity, and graphed the change in polarity shown on Figure 3.

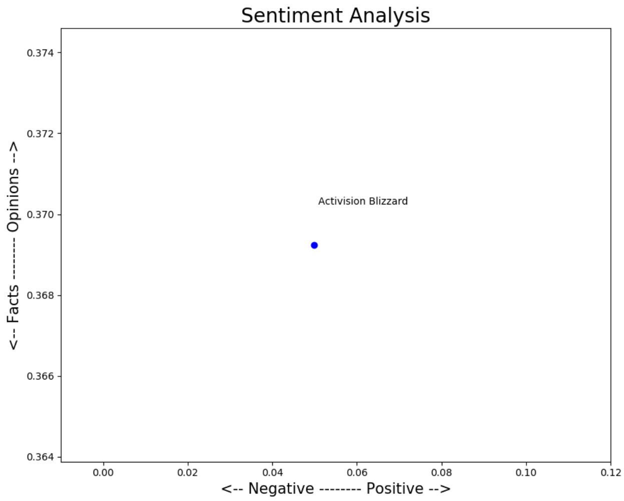

Figure 2: Activision Blizzard's Polarity vs. Subjectivity

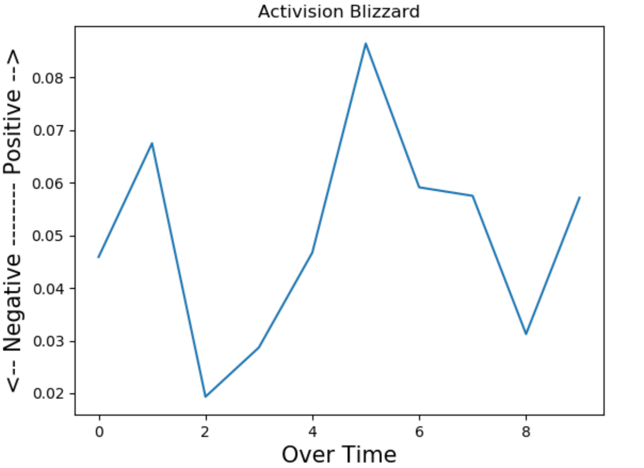

Figure 3: Activision Blizzard's Polarity over Length of 10-K

III. RESULTS AND DISCUSSION

The results for this model were quite interesting. Apparently from the metrics provided on Table 2, one can see a polarity of 0.04989 and a subjectivity of 0.36924. This means the content overall polarity was neutral, and the information given was more fact than opinion. Although it is important to note these numbers are not significant until compared with other companies to get a relative scale.

To expand on this Figure 3 models the polarity over each section in the 10-K. Sections 1 and 5 have high positive polarity relative to the sections. section 1 is the Business Section in the 10-K, while section 5 is the Selected Financial Data. Which indicates a positive polarity on qualitative sections that can be narrated by the company. Sections with low polarity in comparison are Risk Factors, Management's Discussion and Analysis (MD&A), and Financial Statements and Supplementary Data.

From the results lead to two hypotheses. The first being polarity and subjectivity of 10-K's is very low and inconsequential. It seems these factors were useless in predicting the shortcomings of Activision Blizzard's Stock. The silver lining to this analysis is highlighting the change in polarity. A theory can be made, that the greater the volatility in polarity throughout the 10-K correlates to a shorting of the company. It would also be important to isolate specific sections for their polarity for multiple companies' such as the Financial Statements and MD&A.

Although this experiment did not achieve great success in identifying correlations between 10-K sentiment analysis and stock movement, it does introduce other potential metrics that may. Overall, the experiment completed the task assigned, and introduced the concept of Natural Language Processing applications to the accounting industry.

TABLE 2.

Company	**Metrics**		
	Polarity	***Subjectivity***	***Stock % Change***
Activision Blizzard	0.04989	0.36924	-28.52%

IV. PROSPECTIVE DISCUSSION

There are some definite improvements this experiment could use, and I would like to address it in this section. First off, with the data cleaning. It comes as no surprise to me, that my polarity was close to zero due to the massive amount of words still in the corpus. Further steps after tokenization include stemming/lemmatization, parts of speech tagging, developing bi-grams/tri-grams with the adoption of a smoothing technique. Although an uni-gram model was used, I could have changed the n value to sentences or constants. Its also important to understand these n-gram models are based on the Markov Chain, so it become even more of an emphasis to really clean unnecessary text.

Note it was difficult to understand the metrics of Activision Blizzard when there were no other companies to compare it to. More companies would need to be added, maybe even separating sections could give a better correlation between stock movement and polarity of a 10-K.

From Chapter 22 of Russel and Norvig [1], I could have possible enhanced my data retrieval process if I adopted the PageRank algorithm instead of using a python web scraping package.

REFERENCES

[1] Russell, S. and Norvig, P. (n.d.). Artificial intelligence. 3rd ed.

[2] W. Kenton, "10-K," *Investopedia*, 18-Apr-2019. [Online]. Available: https://www.investopedia.com/terms/1/10-k.asp. [Accessed: 30-Apr-2019].

[3] W. Kenton, "10-K," *Investopedia*, 18-Apr-2019. [Online]. Available: https://www.investopedia.com/terms/1/10-k.asp. [Accessed: 30-Apr-2019].

[4] Barrons, "Penta Top 100 Hedge Funds," Barron's, 17-Jun-2017.[Online].Available:https://www.barrons.com/articles/penta-top-100-hedge-funds-1497665963. [Accessed: 30-Apr-2019].

[5] [Video Games in the US: About This Industry], IBISWorld; https://clients1.ibisworld.com/reports/us/industry/default.aspx?entid=1993.

[6] "Activision Blizzard, Inc (ATVI) Stock Price, Quote, History & News," Yahoo! Finance, 30-Apr-2019. [Online]. Available: https://finance.yahoo.com/quote/ATVI. [Accessed: 30-Apr-2019].

[7] W. Chong, "Natural Language Processing for Sentiment Analysis" Available:http://uksim.info/icaiet2014/CD/data/7910a212.pdf [Accessed Apr. 30, 2019]