**Articles for testing:**

Optimistic Outlook:

<https://oilprice.com/Energy/Energy-General/Brent-Oil-Price-Could-Double-By-December.html>

Pessimistic Outlook:

<https://nymag.com/intelligencer/2020/04/oil-prices-coronavirus-russia-saudi-uso-iraq-nigeria.html>

<https://www.cnn.com/2020/04/28/investing/oil-prices/index.html>

Computational Finance

FNBU - 4433 - EL1

Final Group Project Write Up

Peter Warzecha, Jason Hackett,

Minjie Shen, Jainan Liu, Eileen Jiminian

Clearly Defined and Novel:

In class, we have used an Emotion Lexicon in order to conduct a sentiment analysis. A sentiment analysis is the interpretation and classification of emotions ( i.e positive, negative and neutral) within text data using text analysis techniques. Using that sentiment analysis, we derived the emotion of the article in which the article showed anger, joy, disgust, fear, sadness, surprise, and trust. If the code was done correctly, then the results would show how often these emotions were shown in the article. In addition, this code represents how some emotions would stand out more than others.

What made our code clearly defined and novel was the approach to the article. We conducted a sentiment analysis in order to determine how biased the article was due to the author’s stance. The code contained three categories: pessimistic, optimistic and indifferent. If the code had a more pessimistic stance, the result would say that the article was pessimistic and so on. In addition, we created our own lexicons to suit our needs in order to help us calculate the results of the article. As a result, we had to look up each category and their synonyms to create a proper outlook. In addition, a category like pessimism is not the same as negative. A pessimistic person expects negative outcomes more often and vice versa. This is clearly defined because it is concise and straight to the point. Anyone can understand what is going on and what is being done. There is clarity in the steps as shown below. Overall, this sentiment is both clearly defined and novel.

Source of Textual Information & Results:

With the target of our project being to conduct sentiment analysis, our code was used to determine whether the author of an article is demonstrating an optimistic, pessimistic, or indifferent stance on the current standing global oil market. To effectively perform validity checks on our sentiment analysis tool, we used a variety of news sources, converted into a readable text file. The first source of textual information is an article taken from oil price.com titled; “Brent oil price could double by December”. The article is essentially broken up into three sections, first discussing the current state of the global oil market, the second pertaining to what can be expected in terms of recovery in both the short and longer-term, and finally, a section discussing how companies are strategically maneuvering through the complexities of the current market state. After running this article through our tool, we determined the overall perception of the article to be optimistic, a connection made through our three dictionary sets.

The second article ran through the code comes from the New York Intelligencer. In a sense, this article is positioned in more of an editorial view while the previously mentioned article from oil price.com demonstrates more of a factual, economic-based outlook on the current situation. Titled “oil is crashing again. It could take whole governments down with it.”, this article translates the current oil market crash into broader stroke terms of intergovernmental dependencies, partnered with the global pandemic of COVID-19. What we see from this article is a grim outlook on the state of the world, connecting the potential crash of the oil market in conjunction with a fragile global economy, amidst the outbreak of COVID-19, calling for a potential crash of international banking that is tied to oil markets. Essentially this article is stating that if the oil companies go under, banks reliant on oil will also go under, which has the potential to collapse governmental systems worldwide. After running this article through our code, we saw that our key-value dictionaries matched up with the overall tone of this article, classifying the author’s perspective as pessimistic.

The final article that we processed through our sentiment analysis tool comes from CNN business, titled “US oil drops as much as 20% as oversupply concerns keep roiling markets”. What we see from this article is a unique perspective, compared to the two previous articles, as CNN business discusses the current global oil market from a finance perspective. This article is primarily driven through facts and figures relating to percent decreases and barrel price, dollar figures, and market point fluctuations. Essentially what we are seeing is a completely new type of article, where the tone is set from a numeric standpoint. While reading the article denotes a negative perspective, explaining market drops end dangerously low prices, a run through the sentiment analysis code proved otherwise. Our results from the code designated an “indifferent” outlook, being that this article is driven from a figure standpoint. This brings up a weakness in using sentiment analysis for an article of this caliber. What we are seeing here is not an article driven by emotion or perception, but strict financial values. Perhaps a future enhancement for tools such as this can be coupling textual sentiment reading with previously defined mathematical value functions related to percentage and dollar amounts. Notably, this enhancement would prove a new set of unique challenges but does seem feasible.

To obtain the most comprehensive understanding of the flexibility in our sentiment analysis tool, we ran three very different articles related to the same situation. In the first article from oil price.com, the tool tested perception from an overarching economic-based standpoint. The second article focused more on author motion, giving a longer text with a personal touch, perfectly read by our sentiment analysis tool. Finally, we ran the tool with a financial article, which, as expected, was most difficult, but relatively accurate. While the CNN article possessed an overall pessimistic outlook, our tool classified it as indifferent, which from a financial standpoint makes sense. With a better understanding of the types of articles used to conduct our report, the next section will go into the details of how we constructed the sentiment analysis tool itself.

The next step in the process of our project was to create a coding program that can slice through the articles, taking out key words that can be used to identify the overall tone of the writing. As mentioned before, the purpose of the program is to discover the overall market perception of the oil industry given the latest news of major declines in oil price. Our plan is to let the program read each article we collected and determine the sentiment, identifying it as either optimistic, indifferent, or pessimistic. Each article will have its own different sentiment value based on libraries that were created by us manually. Once we reach a value, it will correlate to a specific tone that each article places an emphasis on. The steps to construct the measure are given as the following:

Step 1: Read in the textural information

* Direct the program to go to where we store our resources
* Ask the program to read every txt file in the folder; the program will go on and on until all the news is read.

Step 2: Negation handling (Input: the txt file in the folder)

* Define three patterns: n’t + word / no + word / not + word
* Find everything that fits the patterns and save them into a list
* Scrape these words from the text
* Let the function return two variables: a text with negating words combinations removed and a list of negating word combinations
* Convert the words within the list to their original forms

Step 3: Cleaning up the text (Input: a text with negating words combinations removed)

* Remove the numbers. This is because it is hard to determine what sentiment a number contains.
* Remove punctuations and convert characters to lowercase
* Replace multiple white spaces with a single whitespace; remove leading and trailing white spaces
* Split the text into multiple pieces and add them into a list; avoid splitting words like “New” and “York” because they are only meaningful when they are together
* Filter out the stop words from the list and convert the rest of words to their original forms
* Let the program return a list with those words

Step 4: Read in the Excel Spreadsheet

* Direct the program to read every spreadsheet we created (pessimistic, indifference, optimistic)
* Add the words in each spreadsheet into a list and convert them to lowercase and their original forms
* Create a negating version of pessimistic and optimistic word lists
* Get final pessimistic word list by combing stemmed pessimistic word list with optimistic negating word list; get final optimistic word list by combing stemmed optimistic word list with pessimistic negating word list

Step 5: Get the sentiment of each article

* Count the number of words in the text
* Measure the frequency of pessimistic, indifference and optimistic words
* Calculate each sentiment: frequency divided by number of words in the text
* Get the outlook of each article, which is determined by the sentiment that has the largest value
* Print out the result and update the number of each type of articles

Step 6: Determine the market perception

* Print out the number of each type of articles
* Get the market perception, which is determined by the type of article that has the highest value

Real-world implementation:

Through the sentiment analysis of the oil price-related online reviews, we are able to know the market’s overall attitude towards future oil price trends. We specify the trend into three categories: pessimistic, indifferent, and optimistic. After running the sentiment analysis on three online reviews, we see each of them demonstrate different attitudes towards the future oil price. However, if we are able to run a large number of online reviews through our sentiment analysis, for example, say that 20% of online reviews have an optimistic attitude, another 20% of results shows that the market has an indifferent view on the oil market, and the rest 60% of the market view holds a pessimistic view on the future oil price. This is only a projection; the real results need more data of online reviews to have a solid statistical trend. From the three online reviews we have now, we cannot make any certain prediction about the market’s attitude toward the oil market.

Once we have a large amount of data, we might use this information to predict the future oil price movement and do some arbitrage trading actions. Our sentiment analysis is based on recent online reviews, so the effect of the trend might be a short-term effect. We narrow the effective period to one quarter. It is because any random event or news will change the overall fear index and sentiment of the whole market. For instance, the recent COVID-19 breakout in the world dragged the stock market in the US very quickly. To put our analysis in the real world, we think the trends and percentages could tell us the future market direction. In our case, the majority of online reviews show a pessimistic view on the future oil price, and the fear index is still very high. As we all know, the market price shows the equivalence point between people’s supply and demand. When people have a pessimistic view of the future oil price, the demand gets lower while the supply does not change. This will push the price to an even lower low within a short time period.

With this information on the mind, we can carry out two trading strategies: shorting the short-term oil price, and mapping for the long-term oil price growth in the future. This strategy is based on the idea that “buy when everyone else fears, sell when everyone else is greedy.” We already know that the overall market had a pessimistic view on the short-term oil price, and when people have a bearish view, they would sell their stocks, which leads the oil price to drop lower. At the same time, these online reviews have some influence on other investors in the market. When other investors see those reviews online, they will be affected to sell their oil stocks. And the chain reaction begins. With this information in mind, we can short the short-term oil market price to profit on the emotional selling trend. If through other financial analyses, we think the oil price is low enough to buy any dip, we can also set up a long-term portfolio to long the oil market.