The University of Hong Kong
FINA 4350, 2020-2021

Text Analytics and Natural Language Processing in Finance
Group Name – Financial Linguists (Group 7)

FED-Watching: Reading The Minds Of Central Bankers

Financial Linguists	
Group Member	**UID**
Gopalakrishnan, Amruthraghav	3035436674
Mahajan, Arnav	3035451026
Sharma, Chhavi	3035452599
Khanna, Pareen	3035435577
Agarwal, Siddharth	3035555482

ABSTRACT

In this project, Natural Language Processing is being leveraged to read the minds of central bankers. In particular, this has been achieved by carving up the project into two parts. First, a correlation between the individual FED speeches delivered by the officials and the FED funds rate changes has been established. Second, models that predict the next FED funds rates decision using textual documents and key economic indicators are trained and tested.

BACKGROUND

The strapping influence of the Federal Reserve's FED funds rate decisions on the financial markets is common knowledge to every investor in the world. Every proprietary trading firm, every asset manager, and every investment bank is deploying state-of-the-art technologies to answer the trillion-dollar question – "*What is going to be the next interest rate decision*". In fact, financial market participants are not only looking to predict the next FED funds rate change but also getting to the bottom of every word, every expression, and every opinion delivered by the economic wizards of the US Central Bank. The FED funds rate is not only important to investors and financial market participants but also to every consumer, retailer, or business person as it affects the day-to-day activities of an average consumer. The FED funds rate acts as a benchmark for the mortgage rates at which a house buyer borrows money to buy a house or the interest rate at which an average consumer will be charged for his or her credit card spending.

OUTLINE

The report provides a step-by-step approach to building this project. It discusses the sourcing of two different types of data by building a web scraper and using additional techniques. Next, it walks the reader through the preprocessing of the data. It then explores the methodologies, observations and conclusions for finding a correlation between the FED speeches and FED fund rates as well as predicting the decision of the FED fund rate following a meeting. In the final section, it talks about the project's limitations as well as the potential future work for improvement.

DATA SOURCING

Textual data alone was unable to provide a strong enough linkage to the change in the FED interest rate thus, metadata has also been taken into consideration.

1) *Sourcing Textual Data* - Textual data consists of the minutes of the FOMC meetings, speeches of the Chair/Vice-Chair/Governor, related testimonials, statements, press conference transcripts, meeting transcripts as well as the calendar of the meetings. All the textual data was scraped from the Federal Reserve's website (https://www.federalreserve.gov/). Since more data is added over time, a sustainable scraper has been built that can be reused for scraping transcripts of upcoming events as well. The scraper handles HTML texts using BeautifulSoup and PDF files using Textract. Additional libraries used include Selenium, Pandas, Numpy and Pickle.

Overall, the data scraper has been implemented using three different python files.

- ➔ FOMCBase.py: Receives the scraped data and stores it into a text file. Although the format of the web pages belonging to the Federal Reserve is the same, the links and few attributes are different. That is why following an object-oriented approach, a base file has been built to put the common code in.
- ➔ FOMCMinutes.py/FOMCSpeech.py/etc: Files responsible for scraping the data of the web pages each specific to the components of the page. Simultaneously, it also cleans the data by removing the appendix, footer and references.
- ➔ FOMCGetData.py: Implements the different combination of files mentioned before. It specifies the years we want to scrape from and is the file that we run. It also dumps the final text into a pickle file.

2) *Sourcing Metadata* - Metadata consists of economic indices (for example, FED rates, GDP, PCE, CPI, etc.) available on the FRED website (https://fred.stlouisfed.org/) that the FOMC members take into consideration. These indices were scraped using getQuandlData.py.

PREPROCESSING

Raw data needs to be made suitable for our machine learning model. Preprocessing of the sourced data has been performed in four steps.

1) *Reorganizing the data frame* - Additional columns including “type”, “word count”, “text”, “text_section” were added to the already existing columns; “date”, “title”, “speaker”, “original_text”. The type column specifies the type of information the text contains (eg. speeches, press conferences, etc.), the word count column keeps a count of the words, the text column consists of only the speaker notes after removing extra information (eg. [Remarks by…]) and the text_section column consists of the original text separated by the [SECTION] tags.

2) *Removing text with less than 50 words* - During preprocessing, text containing less than 50 words was removed since it was less likely to make sense and make any significant contribution to the model.

3) *Breaking up the text into 200 words* - An extension of the input data frame was created where the text column has been divided into additional rows having a maximum word count of 200 words. This was done because it was easier for the model to train on and analyse shorter texts and more rows.

4) *Removing text without at least two keywords* - On conducting initial analysis, it was noticed that texts which did not have specific keywords were not supporting our analysis and making it less sensitive. The chosen keywords were rate/s, federal fund, outlook, forecast, employ and economy.

FINDING A CORRELATION BETWEEN FED SPEECHES AND FED FUND RATE

The speeches made by the Chair, Vice-Chair, and Governor are analysed by performing sentiment analysis on them to see if any common trend is visible between the overall sentiment of the speeches and the fluctuation in the interest rate. It was noted that there have only been 4 FED Chairs till now, whereas the Governor changes frequently and multiple individuals can hold that post at the same time.

Methodology - Important keywords were extracted from all speeches. These keywords were then used to analyze if the overall sentiment of the speech was positive or negative. Sentiment analysis was performed by making use of the Natural Language Toolkit in Python. The Loughran and McDonald Financial Sentiment dictionary helped categorize key financial terms from the speech into negative, positive or neutral sentiment. For the FED Funds Rate, the interest rates between the year 1996 to 2020 were used to draw a correlation between the net sentiment of each individual holding office for the position of Chair, Vice-Chair and Governor and the FED interest rate.

Exploratory data analysis was performed by plotting several graphs that supported our correlation analysis. [*Appendix - Exhibit 1*] These included:

- ➔ Number of positive vs. negative words in speeches
- ➔ Net Sentiment of speeches
- ➔ Correlation between the FED Fund Rate and the Net Sentiment of the speeches

Observations - For correlation graphs refer Appendix (additional graphs in blog or code). The Chair’s and Vice-Chair’s speeches were overall relatively neutral with balanced positive and negative vocabulary in each speech. The Governor’s speech exhibited the most polar trends. During the period between 2004 to 2008, the FED interest rate showed a steep rise, during which the speech sentiment of the Governor was very contrasting. Additionally, the graph was quite dense (along the x axis) compared to the less noisy ones of the Chair and the Vice-Chair.

Conclusion - A neutral stance is expected from the Chair as the Chair is the face of the FED and cannot allow any kind of controversy following his speech. He needs to remain composed and supportive of the FED's decisions. The Vice Chair's opinion tends to mirror that of the Chair thus, justifying the visible trend. The Governor tends to be the most liberal in terms of speech content. This trend can be explained due to the large number of governors that have held office as compared to the Chair and Vice-Chair. There have only been four chairs till now, who somewhat adapted to the "composed" speech trends.

Though there are outliers, if one must try to relate any of these personnel's speeches with the FED interest rate and wants to get a hint at the change in interest rate and how steep it is going to be, the Governor's speech would be the most suitable choice.

PREDICTING FED FUND RATE DECISION

For the second part of the project, we emphasized using various text documents released by the Federal Reserve, such as FOMC statements and press conference transcripts, to predict the FED fund rate decision. Numerical data of major economic indices were also used to increase the accuracy of the prediction.

Methodology - This part has been divided into textual data and metadata.

1) *Textual Data* - A simple EDA was carried out to gain a better understanding of the textual data that was scraped. There were about 200 documents for each category of files except for speeches, which were about 400. The meeting scripts were dropped from the dataset as these scripts were released four years after the initial meeting had taken place, hence not aiding the project to make a timely prediction. Following this, each document was linked to the outcome of the next FOMC meeting. A word ranking chart and word cloud were plotted to obtain a holistic view of the significant topics discussed in these documents. Each file has an average of about five thousand to seven thousand words per document. This was a concern as most neural network algorithms are not very efficient in analyzing long texts. Hence we split the words into 200 words sections. The tone of each text section was evaluated with negation by using the Loughran and McDonald dictionary for positive and negative words after lemmatization. For each group of words, a simple ratio was calculated to find the tone of the texts. First, we calculated the positive and negative words for a particular document, and after that, we used the following formula to compute the tone:

$$Tone \ = \frac{100\,X\,(Positive\ Words - Negative\ Words)}{Word\ Count}$$

The tone for each text document was appended in the data frame. Tokenization and computation of BoW were performed for the individual text documents, and the values were also appended to the data frame subsequently. The TF-IDF vector and Cosine Similarity was computed and appended to the data frame.

2) *Metadata* - The major economic indices taken into consideration are the GDP, PCE, CPI, Unemployment and Employment Rates, Retail Sales and New Home Sales, and finally ISM Purchasing Managers Index/Non-Manufacturing Index. A correlation graph between FED rate decisions and the different economic indices was plotted, and only highly correlated indices were selected. Taylor, Balanced and Inertial rules were also taken into account while calculating the correlations.

After careful analysis of both types of data, the Machine Learning model was trained to predict possible decisions from a meeting: Raise, Hold or Lower Interest Rate. Four models were trained, and the data was split in the ratio of 80:20 for training and testing. Testing accuracy and F1 scores were used as the primary metrics for ranking the models.

Observations - The observations for the 4 trained models are as follows [*Appendix - Exhibit 2*]:

Model	Testing Accuracy	F1 Score
Random Forest with Cosine Similarity	77.55%	.72
Random Forest with Tf-idf	72%	.33
LSTM with Document Embeddings	54%	.24
LSTM with GloVe Embeddings	65%	.41

1) *Random Forest with Cosine Similarity* - From the Confusion Matrix that was plotted for this model, it can be seen that the model has clearly overfitted on the training data and was mainly predicting "Hold" even when the actual values are "Raise" and "Lower". More data could potentially improve performance.

2) *Random Forest with Tf-idf* - This model used the Tf-idf vector to train the model and was also found to overfit on the training data.

3) *LSTM with Document Embeddings* - The vocabulary embedding was computed using the BOW earlier on along with vectorisation. The model also used Adam as its optimizer and negative log likelihood as the loss function, and finally, log softmax as the activation function. Even with neural networks, it is observed that the model is overfitting and making more predictions for "Hold" decisions.

4) *LSTM with Glove Embeddings* - This model also overfits the training data. The activation function, optimizer, and loss function were the same as the previous LSTM model. However, it was evident that the accuracy of Glove Embeddings was higher.

Conclusions - Out of all the above models, the random forest classifier with the cosine similarity managed to obtain the highest accuracy. The other 3 models were still able to predict the FED fund rate decision for most parts. However, due to the overfitting, the models were failing to predict "Lower" and "Raise" events. In hindsight, the reason for overfitting was discovered to be skewness in the decision making of the FOMC. There was a data imbalance in the decision making over the years. For instance, out of all the meetings held by the FOMC, almost 60 percent have ended with the decision of holding the FED Fund Rate. Hence, the data is biased towards holding interest rates. Without having enough unbiased data, it is very easy for the machine learning model to overfit.

LIMITATIONS

The minutes of the FED meetings and speeches of the FOMC members tend to have a neutral orientation since they refrain from mentioning exaggerated words. Lack of strong polarity scores saturates the sensitivity of the model after a point.

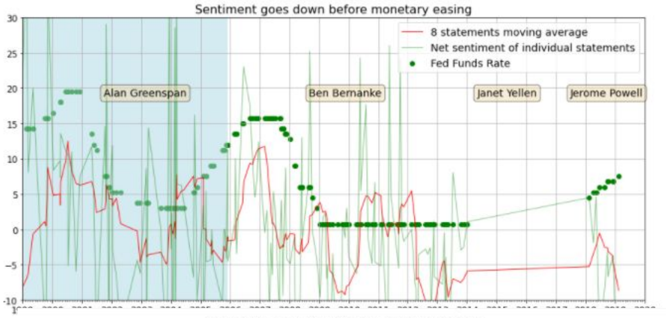
FUTURE SCOPE

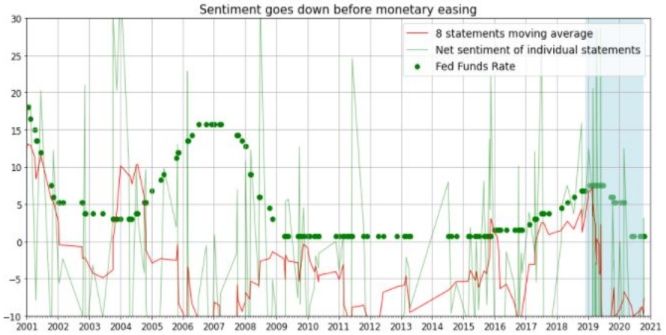
1) *Solve the problem of model overfitting* - The models had to be overfitted to train the samples due to the lack of training data even after hyperparameter tuning and imputation. Configuring the model and splitting data to augment the training data by a synthetic approach could be potentially beneficial in the future.

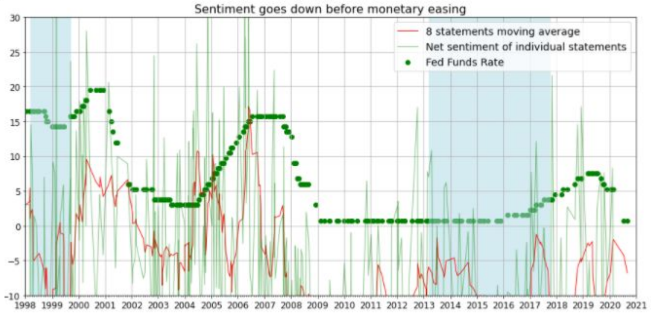
2) *Improve the quality of input text* - The input texts contained many irrelevant paragraphs that had nothing to do with FED target rate decisions, for example, information about regulations and infrastructures. Filtering out less relevant inputs will improve the accuracy of the model as well as the training efficiency.

APPENDIX

Exhibit 1: Chair, Vice-Chair, Governor Speeches vs. FED Rate







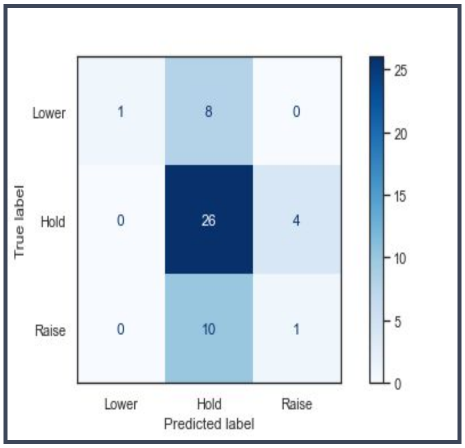
(Top-Left)Chair's Speeches vs. FED Rate

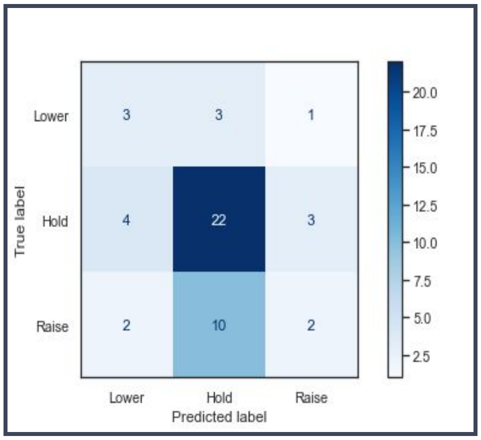
(Top -Right)Vice-Chair's Speeches vs. FED Rate

(Bottom-Left)Correlation Graph Governor's Speeches vs. FED Rate Correlation Graph

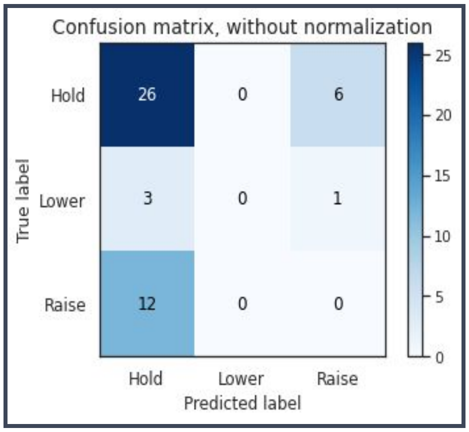
Exhibit 2: The Results of the Models from Training Model(Actual on Y-axis and Predicted on X axis)

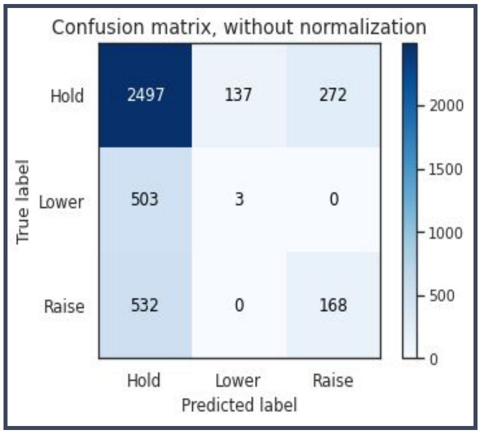
Top Row: Random Forest Classifier with Cosine Similarity(Left) and Tf-idf Vector(Right)





Bottom Row: LSTM with Document Embeddings (Left) and Glove Embeddings (Right). Refer to the jupyter notebook for more detailed graphs.


Confusion matrix, without normalization
True label
Hold
Lower
Raise
Hold
Lower
Raise
Predicted label


Confusion matrix, without normalization
True label
Hold
Lower
Raise
Hold
Lower
Raise
Predicted label