**Technical Document - US FOMC Communication Interest Rate Forecaster**

Shawn D. Mitchell

Executive Summary

Using US Federal Reserve Open Market Committee public communications, we can forecast with approximately 80% accuracy the future 6-month interest moves. Interest rate moves are considered only as positive, negative, or neutral sentiment within a given threshold. Interest rate change amounts are not forecasted, only whether it is likely to change up or down.

Part 1: Summary and Problem Statement

Can predictions of future interest rates be made based solely on FOMC communications?

Predicting the future actions of the US Federal Reserve is a difficult task, with no method of having 100% accuracy unless one both has internal information at the FOMC and has a perfect view of the future US economic situation. Their communicated intended actions aren’t always completely aligned with the actions they take, and their real actions could be completely contrary to past communicated intentions following an economic shock.

An NLP model with reasonable accuracy, combined with additional quantitative and qualitative information, would be a highly valuable tool.

Part 2: Data

Two sources of data were required for the project: historical interest rates, and the FOMC communication documents.

The FOMC provides download links for all public communications, ranging back to the 1930’s. The types of communication documents have varied over the years, both the naming of the documents and the scope of the information contained. Communications are not released every month, and different months may have different types of communications released. Documents ranged from research articles to press releases and meeting minutes.

To obtain the data, a web scraper function was developed to loop through all requested years and download all documents available for them. Since the types of communications varied through the years, a single type of document wasn’t used (such as the currently running Beige Book). Some documents released are of very little analysis value, only containing a brief press statement or announcement that information will be released later.

There were three options for deciding what documents to use:

1. Manual process of scripting specific documents types to use for all months
2. Use all documents
3. Use a consistent pattern to pull a sub-selection for each month

A manual process would be wastefully time consuming, and require frequent updates to specify what documents to use in the future. Using all the documents would be computationally inefficient, particularly given the fact that many communication PDF’s hold no predictive value. There may also be a problem with unbalanced data features, if some months have much less information that other months.

A consistent pattern was settled on, using the PDF document for each month that has the largest file size. There are problems with this approach:

1. Largest file may not have the most text, some files may be image-heavy
2. Largest file may be a section from a multi-part communication document intended to be read as a single whole
3. What the largest file is can vary month to month. Sometimes it may be a summary of current economic conditions, other months it may be a transcript of FOMC meeting minutes.

Despite these problems, this approach was deemed the best option of the three available. A balance between computation time and manual scripting time was needed. Given more time, a more sophisticated algorithm could be developed to identify the best documents to use (word count, identify multi-file releases).

After what documents should be used was determined, a function looped through all valid PDF files and extracted the text to a pandas dataframe. Each document was assigned a date, based on the date timestamp in the files. Some files had two dates in the filename, which will cause this method some potentially non-ideal PDF assignments to month. Some documents were written by the FOMC in the month before they were released to the public. Since they have two dates, they can be considered as being relevant for both months for the purposes of this model. The loop to assign PDF documents to a specific month may assign the same document to two consecutive months. This was determined to not be a major detriment to the model, as the document may legitimately apply to both months. Some month additionally have no documents released for them, the FOMC does not release communications through all months of the year.

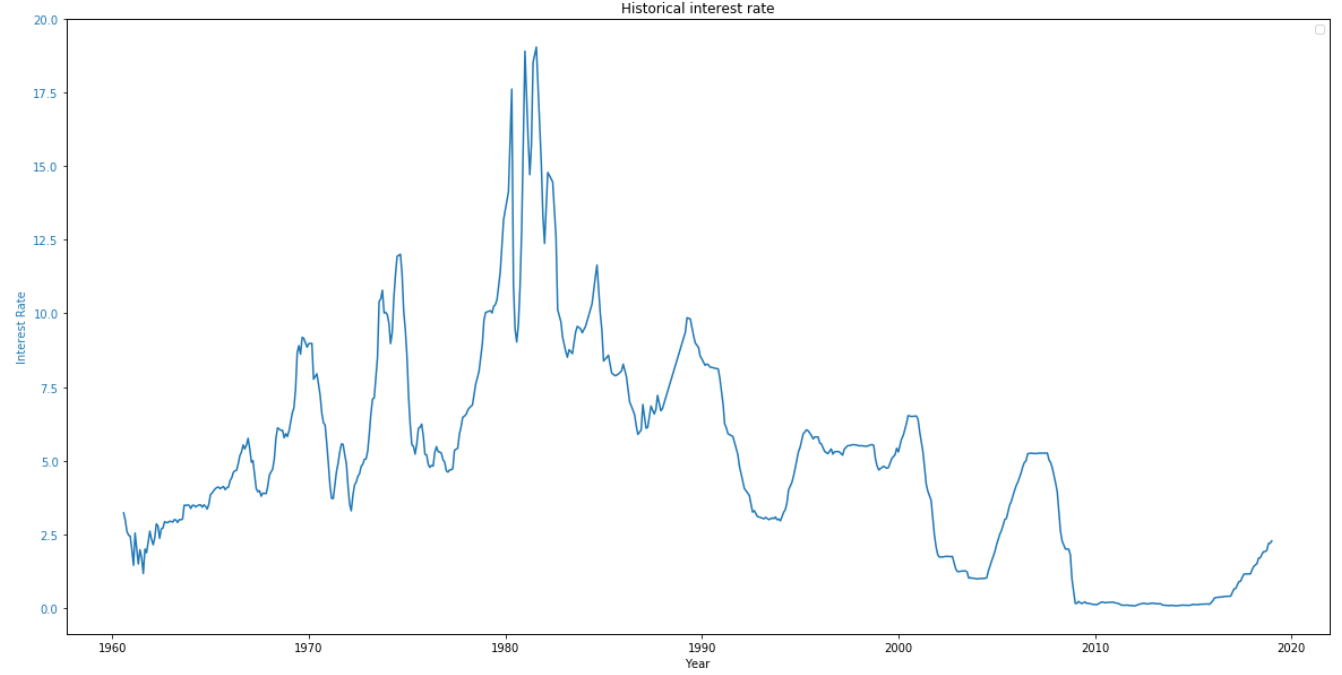
The y-target value is the interest rate change 6 months from the communication release month.

Interest rate data was collected from the US Federal Reserve FRED API, the source data was the daily real interest rate, not the target rate. The real rate was chosen for three reasons:

1. The real rate closely follows the target rate.
2. The real rate is the actual trading rate, not the target rate. The real rate is more relevant.
3. The target rate data was incomplete on the FRED API. The real rate data was complete.

Since the communication documents were grouped by month, the interest rates also needed to be grouped by month. The data collected was daily, so the monthly y-target used is the mean interest rate per month. There may be a wide spread between the lowest and highest interest rate every month, however the mean was the best option compared to using a min/max/mode. It would give the ‘middle ground’ representation of the interest rate environment for that month.  
Forecasting actual interest rate percentages would be an extremely specific, and highly difficult continuous value to forecast. The goal for the project was to create an actionable product that creates believable output. A specific rate percentage would be far more difficult to justify, rather than a hawkish/dovish sentiment prediction. The rate ranges over time change significantly, an NLP model would likely find correlations that would be very non-ideal due to the time series nature of the interest rates. For example, to predict the unusually low interest rate environment we’ve been in for the past several years, the model would likely connect features such as the names of individuals at the FOMC or dates to the interest rate prediction.

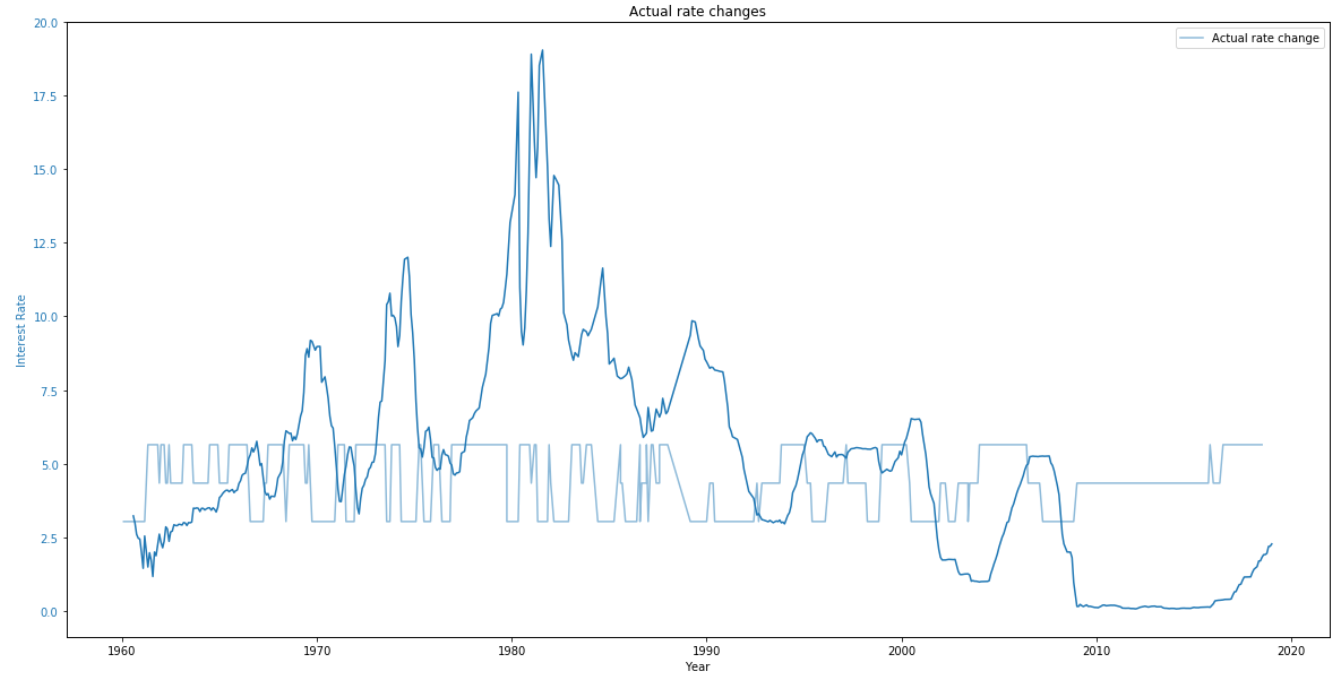
By using a hawkish/dovish sentiment prediction, we can remove some of the undesirable connections an NLP model is likely to make. This sentiment analysis allows our highly unusual low interest rate environment to be compared to past high-interest rate environments.

Initially I explored forecasting monthly sentiment (month+1 = increase, month+2 =neutral, etc). This however could lead to volatile predictions, and was a level of detail not required for the forecast. 

As seen above, the interest rate environments we have been in for the past few decades has been a rising, neutral, or falling environment. With a broader and less focused y-target, the model could give a mid-term sentiment prediction, rather than a short-term immediate action prediction.

I’ve settled on a 6-month sentiment prediction, rather than a monthly prediction. This is focused enough to be an actionable prediction, but broad enough to not throw high volatility into the model. Below is the example actual flow changes of interest rates. Middle points in the line are neutral, with rising or falling 6-month rates being above or below.

All data points are considered to be independent of each other. Time series analysis for interest rate prediction isn’t very useful – past interest rates are not a good indicator of future interest rates. They would only be primarily useful if there was a routine cyclical pattern. Recessions occur at irregular intervals, and the long-term interest rate environment can experience significant shifts over time. The past 10 year near-zero interest rate environment has no precedent, for example. A major goal of the model is to predict sudden shifts, not forecast current trend.



To summarize the data used:

1. Features are extracted from the monthly PDF communication with the largest file size. Some months do not have any communications.
2. Y-target is the following 6-month interest rate change: 1, 0, -1 for increasing, neutral, or falling. The threshold for an increase or decrease is 0.25%. A 6-month change that does not exceed that plus or minus will be considered neutral.

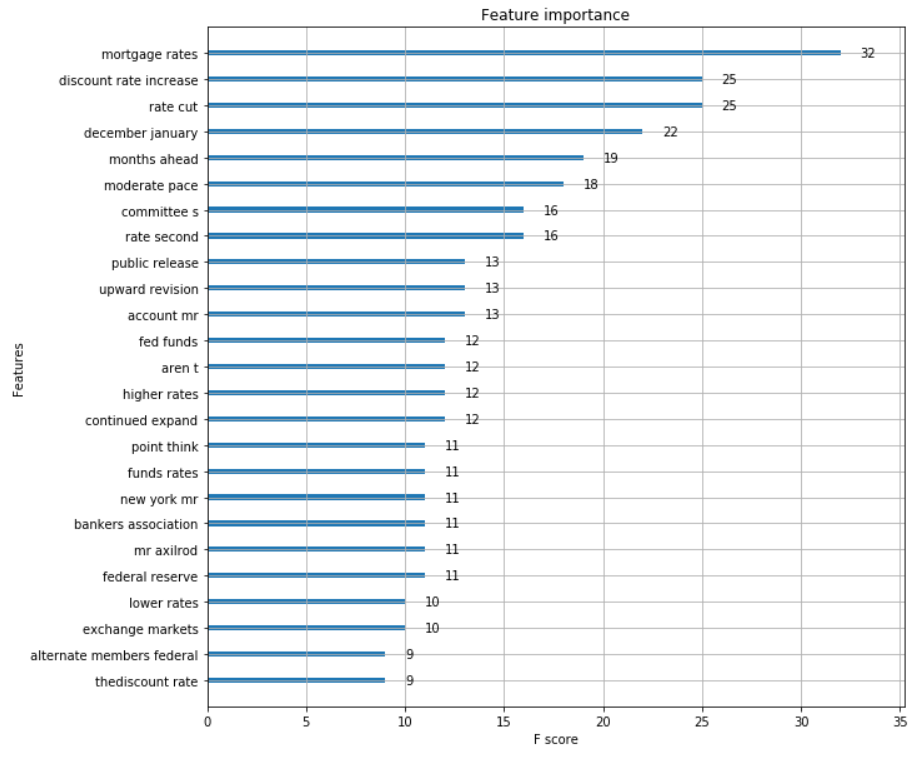
Part 3: Modelling

Preprocessing work included stemming, stop word removal, conversion to lower case. Standard process in NLP, removed differences between words with identical or very similar meanings.

Vectorizer used was TFIDF. All communications were expected to have a large common pool of common words, TFIDF vectorizer was used rather than count vectorizer to assist in identifying words and phrases that may have a higher predictive value for the model. N\_gram range was 2 to 4 inclusive, with the intent of capturing meaningful context of discussion. “Rates”, compared to “falling rates”, for example. Only characters that were alphabetical were included. When including non-alphabetical characters, features such as “000” would be extracted and potentially viewed as meaningful features.

XGBoost was chosen to the be model, based on the strength and computational speed of the model. XGBoost has a large set of possible parameters, however the only non-default setting used was setting the limit on depth to 4. XGBoost has a very strong tendency to overfit to training data, the model used for this project resulted in a 100% training accuracy but approximately 65% accuracy on testing data. A potential improvement for the model could be grid searching parameters, primarily with the intention of decreasing variance (overfit). The first parameter in XGBoost to reduce overfit is reducing the depth limit, however a further reduction past 4 was not found to significantly reduce overfit.

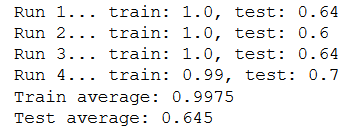
Feature importance summary:



The feature importance summary shows mostly logical features having the strongest predictive performance. A potential further improvement would be building on the stop words, adding pronouns or phrases from meeting minutes that were addressing participants. Further spell checking would assist in removing word fragments that were viewed as important n\_gram features.

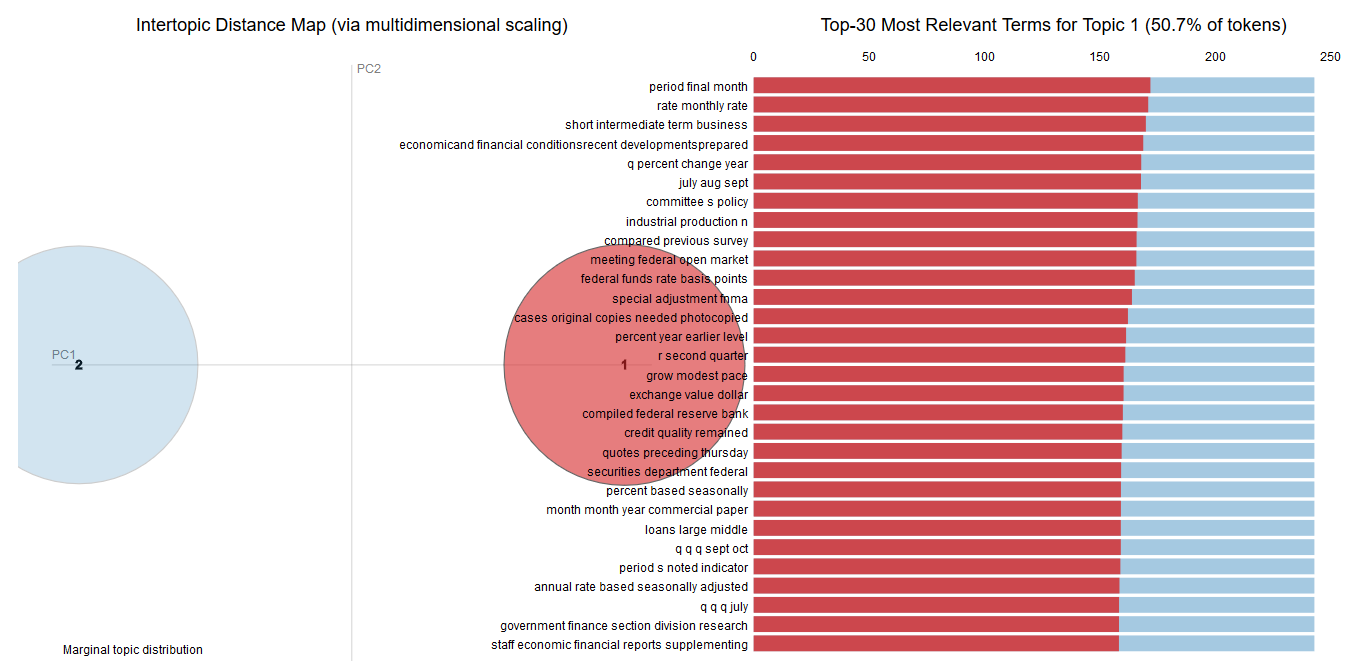
Accuracy Review

The model achieves approximately 65% accuracy currently. A higher accuracy is potentially undesirable, due to the nature of the features used in creating the predictions. The real future interest rates are not always a function of what the FOMC intends on doing. The Federal Reserve would only be able to precisely predict their own future actions if they had perfect information on the state of the economy and financial markets in the future. A model that does achieve a very high accuracy may very likely be finding non-ideal features to predict future rates, such as dates or proper nouns.

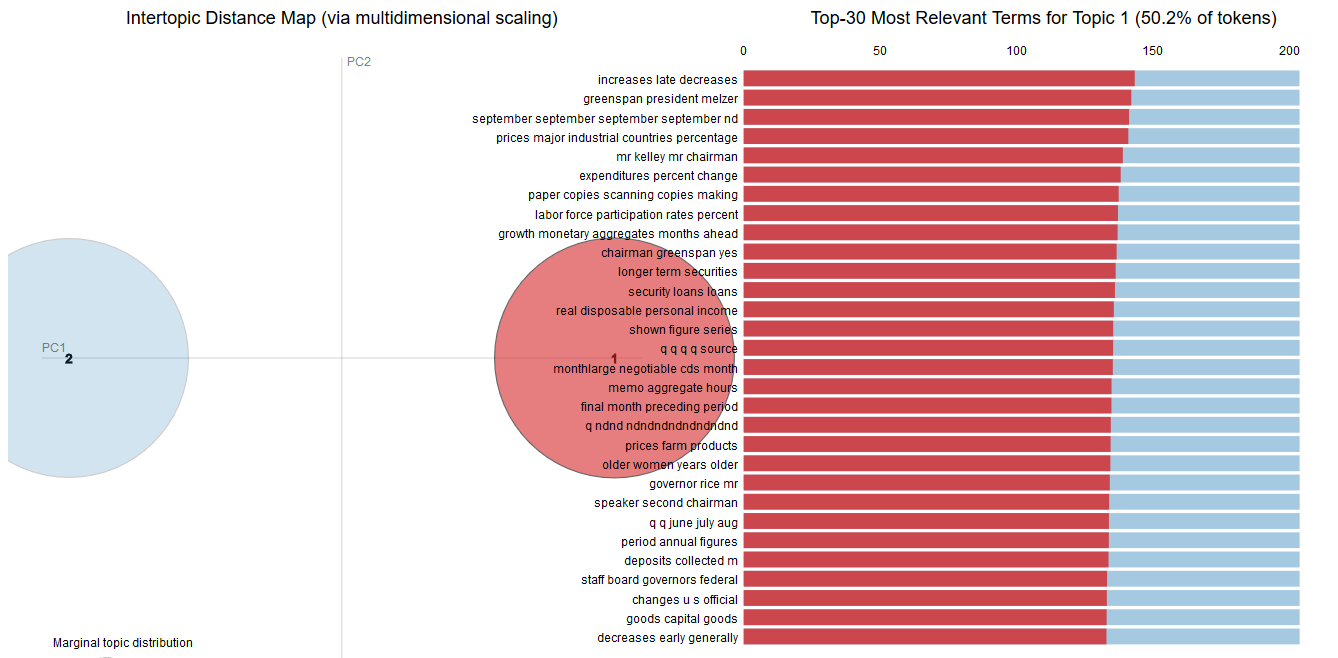


Topic Analysis

Using LDA, we can review an estimated grouping of topics that were discussed in communications that preceeded reductions or increases in actual interest rates. Reviewing the topics for the topics that preceeded increases:



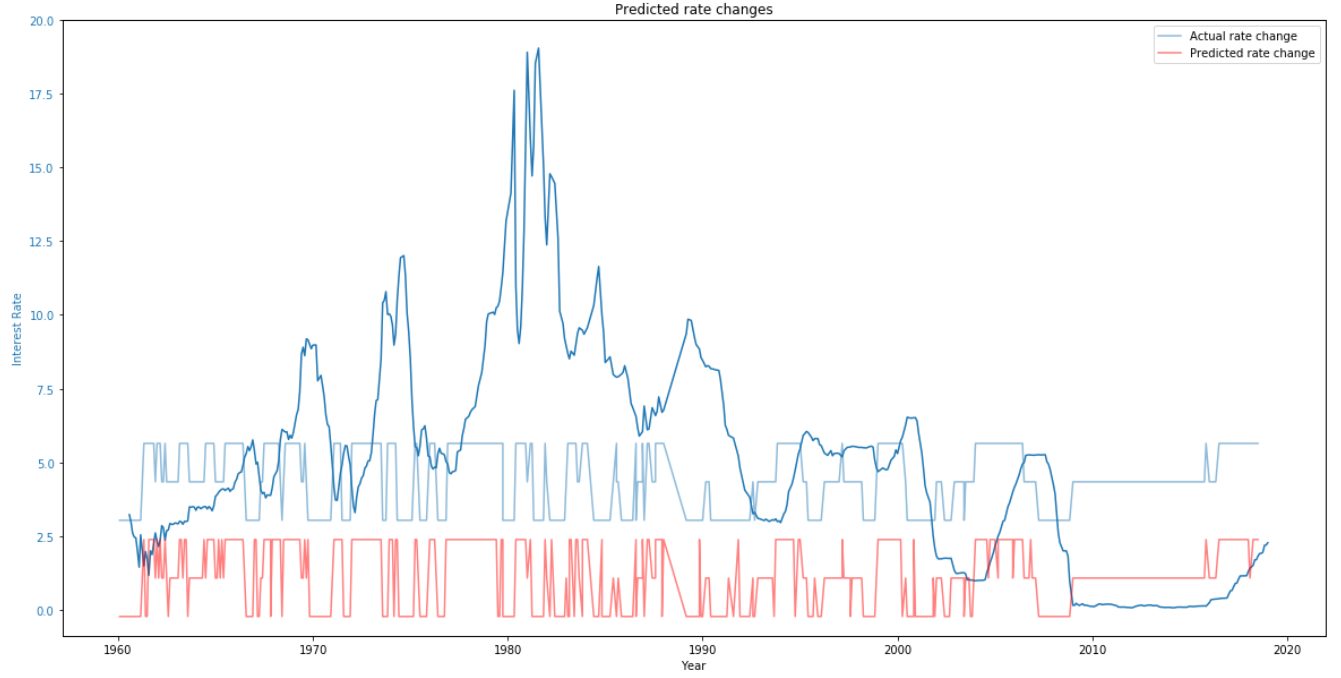
The discussion topics may not be the same as the important features for the model. In the above example LDA analysis, n\_gram range of 3 and 4 word strings were used with the TFIDF vectorizer. We can see word fragments are prevalent, again showing that further in-depth spell checking may improve the model. We do see common logical topics coming up, such as discussing manufacturing orders, financial markets, and money supply. If we compare this to LDA below on communications that preceeded rate decreases:



Here we see similar themes coming, some logical topics we would expect to be frequently discussed. Further in-depth LDA analysis may assist in determining what meaningful topics could be typically discussed before rate decreases. Typically rate decreases happen when the US economy enters a recession, so two common themes should be likely coming up, even though they may seem contradictory: high fed confidence in the economy or low fed confidence in the economy. Recessions typically follow periods of rate hikes, so the fed may be over-confident in the strength before a recession, and hike rates too far. Their communications during that time may be highly optimistic and confident. They may also begin warning of an upcoming recession, economic weakness, or over-valuations in financial markets before a recession. It may be useful to run LDA on a smaller set of pre-recession data.

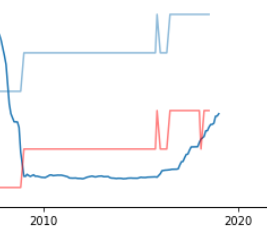
Misclassification Review

Below is a comparison of the actual interest rate, the actual 6-month change, and the model’s predicted changes. Note that the specific predictions for this are based on the last train/test sample that was taken. Specific predictions will vary based on this random split, but the results for every split will be similar.



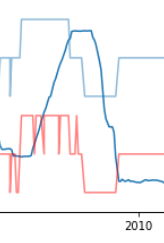
Next we’ll review a few examples of interest rate environments that it handles well, where it didn’t handle it well, and review the logic behind how it was mistaken.

Current rate environment



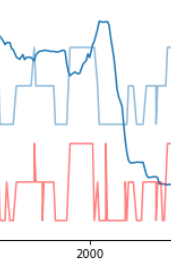
The model was able to identify the historically unusual low interest rate environment we’ve been in globally since the 2008 recession. It predicted stable rates, and continues to predict further rate increases as the year moves on. We are currently in one of the longest market bull runs in history, a recession is likely lurking around the corner for US markets as of early 2019. Next we’ll review how the model handled the 2008 recession.

2008 Recession



Again, the top line is the actual 6-month change, the bottom line is the predicted change. Leading up to the 2008 crash, as typically happens, interest rates were steadily increasing then dropped after the recession hit. In this iteration (based on the random train/test split), the model experienced confusion on just when the rate hikes would stop and plateau. It would have been late on the drop, but identified when thing would hit the bottom and become flat.

Dot-Com Crash



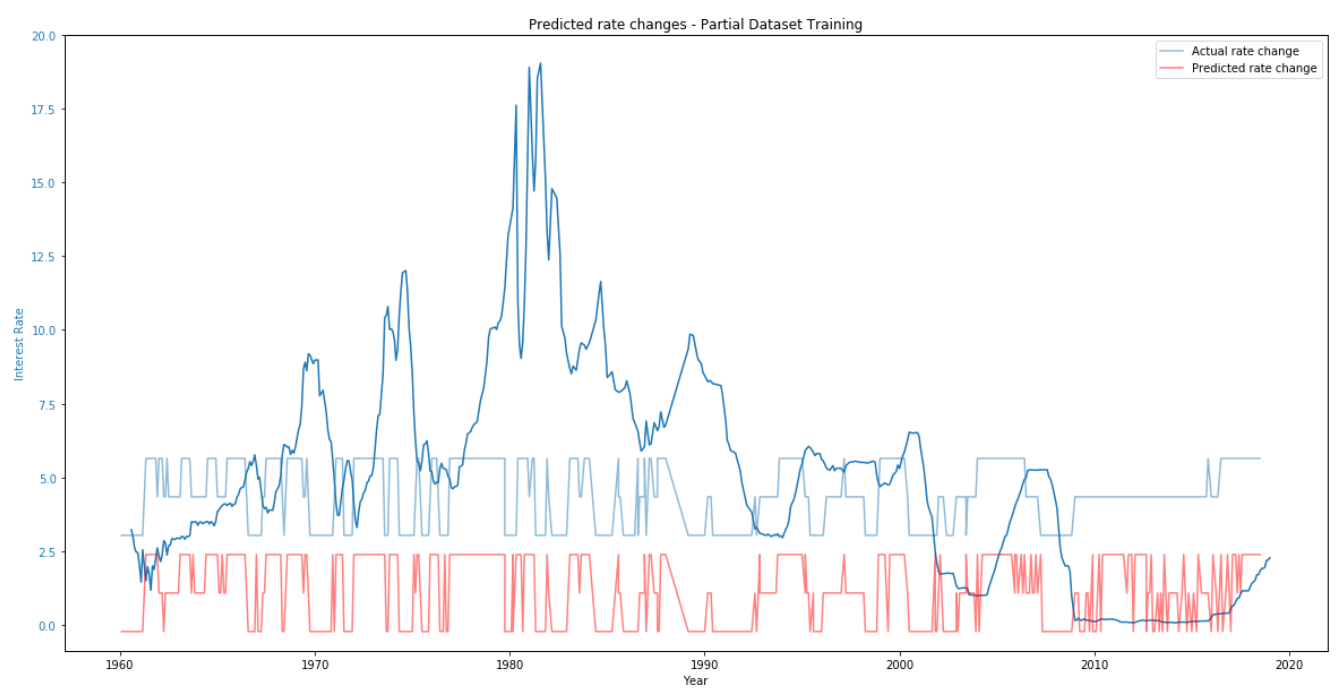
As we usually see, the interest rates were steadily increasing leading up to the dot-com crash. In this model iteration, it experienced some confusion on the way down again. Later on in this technical document, we’ll be reviewing how the model can be improved, and how this volitility can be reduced. Currently each month observation is treated as a unique and isolated occurence, the previous interest rates are not considered.

Part 4: Model Improvement

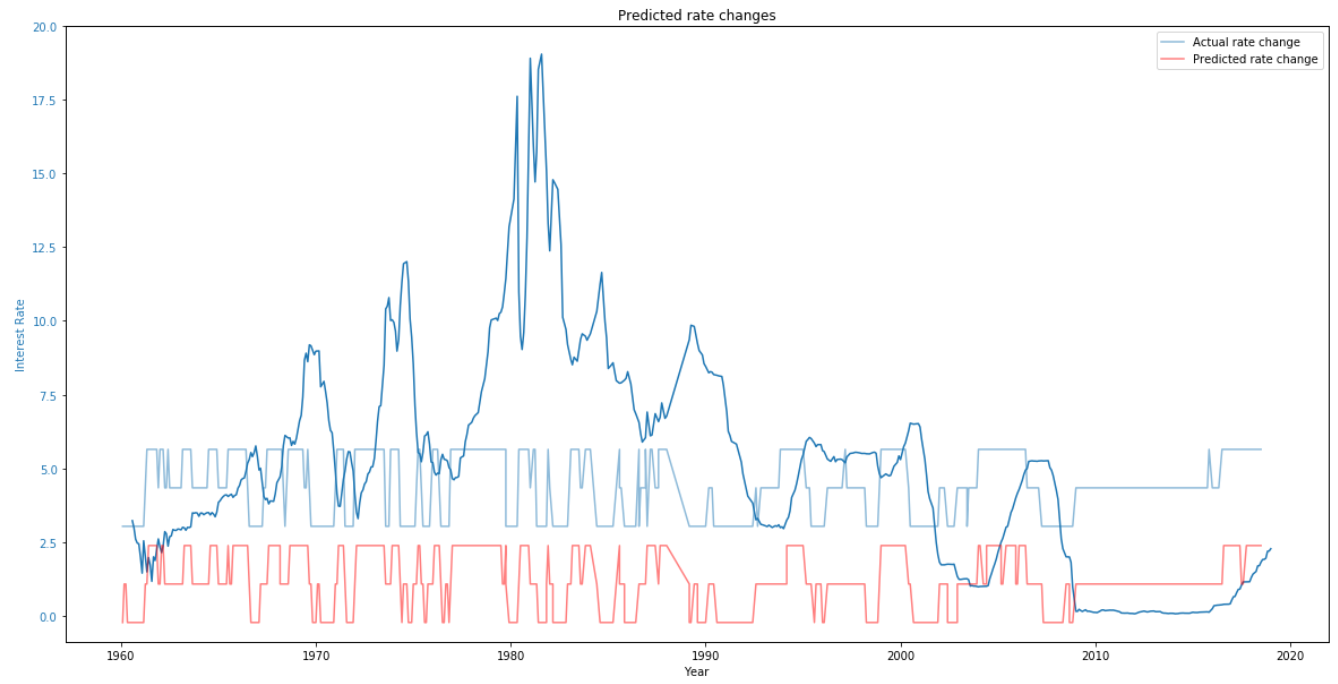
Interest rate data inherently has a time series component. In this model, each observation point is treated as an independent occurrence with no direct connection to the points before or after. Considering the data as completely independent can still create a fairly useful model, as seen above. The model was fit and tested on a random train/test split of 75%/25%.

When trained this way, the model has a view into all historical interest rate environments. This includes this current unusual low interest rate environment. The model may still be fitting to non-ideal features, such as proper nouns. Ideally, the model should be able to generalize past data into future rate changes. It should be able to find universal communication cues, not time period dependent cues.

To test this theory, testing was done with a time-series based train/test split. First, by training on data before 2008, and second training only on data before 2000. The initial results are below for the first split at the year 2008:

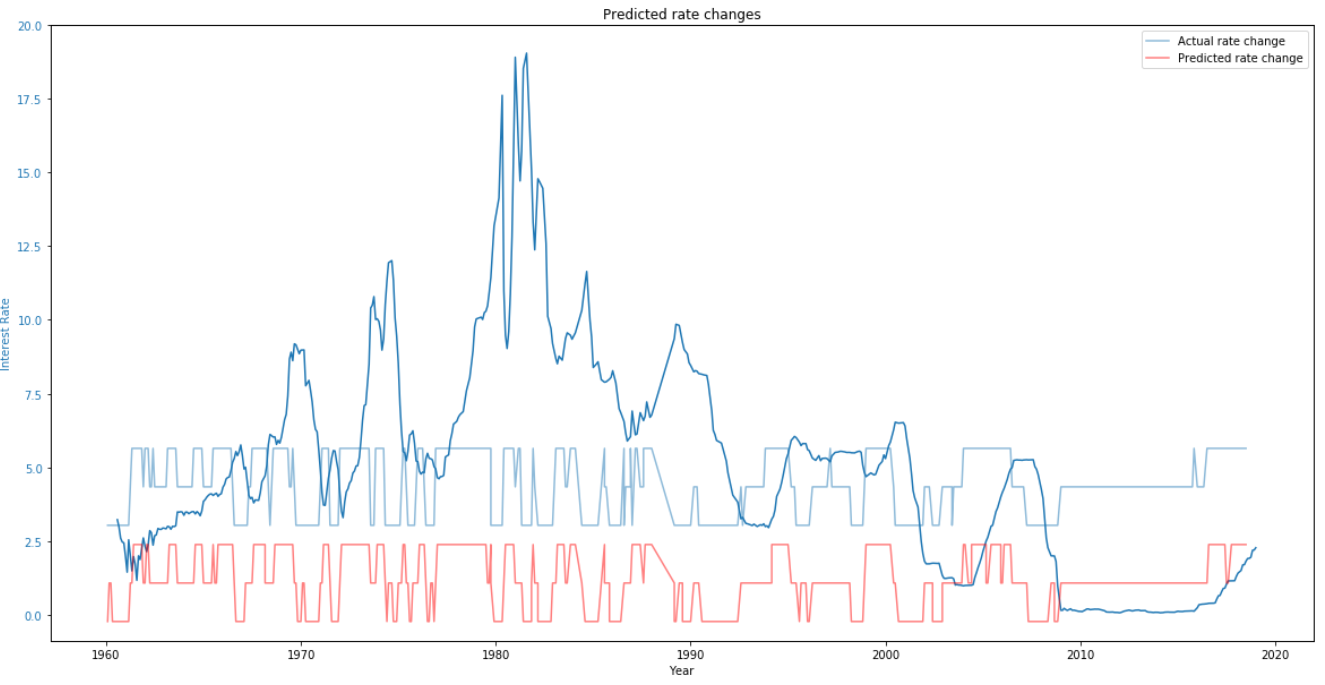


The model performs very well as expected until 2008 since it had a complete view of the data, then falls apart with erratic predictions after 2008. It moves month to month between forecasting an increase or decrease. We can correct this problem by using a single period rounded moving average. The results of the change are below:



By smoothing out the predictions, using the single period moving average, the confusion the model experienced month to month averages out to a fairly accurate long-term prediction. Again to note, the above model was only trained on data until 2008, so it had no information on proper nouns or dates to connect to the current rate environment of the past 10 years. It was not trained on the full dataset before 2008, only a random 75% partial sample.

Below are the moving average predictions with a model trained only until the year 2000:



The overall accuracy of this model was 85%. The accuracy since 2000 was still quite adequate. It correctly forecasted the rate decreases of the dot-com recession before they happened, and the rate decreases of the 2008 recession. The current quantitative easing environment was also mostly forecasted well.

Part 5: Use Case, Limitations, Future Improvements

This is not a tool that should be used on its own to make investment decisions. When used in addition to other analysis, it would however be a useful addition. The FOMC future target rates are a source of endless debate among financial professionals, for good reason. The FOMC does not always follow their own intended plans, and can change plans at any time. They are forced to react to changing economic situations, rather than dictate the course of the economy and interest rates.

With further improvements the model would be a strong analysis tool to compliment other quantitative and qualitative analysis. Improvements may include but not be limited to:

1. More thorough spell checking. The current spell checking used for this model was limited by computing power and time available. Investing more resources into spell checking models results in better accuracy.
2. Increase the quantity of text available to the model, and modify the criteria for a text to be included in the model. This model used the largest single document for each month. There are several flaws with this method, such as the fact that the largest filesize may not be the most meaningful document, and some documents have multi-part PDF’s that would only be half-covered with this current method.
3. Revising the y-target. This could include revising how far out the target is, and the threshold for considering an interest rate change to be significant enough to be considered non-neutral.

Supplemental recommended analysis would include:

1. Independent review of economic conditions.
2. Qualitative review of FOMC political objectives, political obligations, and personalities. This would include reviewing these features of the leadership, and determining an overall picture of the entire FOMC group.
3. Review of quantitative easing status.

The last point would be highly critical to review, given the unprecedented balance sheet expansion by central banks globally following the 2008 recession. Decreasing the balance sheets of central banks may have the same effect on financial markets as interest rate increases. Leading up to our next global recession, quantitative tightening will be a major central bank policy action to watch.

A challenge to developing a model for predicting interest rate changes primarily sits in the changing nature of central banking over time. Personalities involved in decisions change over time, the types of communications change over time, the political objectives change over time, the tools used by central banks changes over time, and the global economic situation changes over time. We are currently in an unprecedented global financial situation:

1. Low growth
2. Low interest rates
3. High wealth inequality
4. High debt, both government and private
5. Large central bank asset balance sheets

As a recession predictor, this model would not be advisable to use. As a supplementary fed sentiment predictor, it may be useful when combined with other tools and analysis. If multiple tools provide mixed signals, it may be an indicator of either a neutral environment and/or more review is needed. If multiple tools all raise the red flag in unison, it may be a good signal to shore up one’s financial defences.