Introduction This blog describes how I analysed central bank policy by means of NLP techniques in a past project. The source code is available in github repo.

**Business Context**

FOMC has eight regular meetings to determine the monetary policy. At each meeting, it publishes press conference minutes, statements as well as scripts in the website. In addition to this regular meetings, the members’ speeches and testimonies are also scripted on the website. At a meeting, the policy makers discuss, vote and decide the monetary policy and publish the decision along with their view on current economic situation and forecast, including Forward Guidance since 2012. The central banks intend to indicate their potential future monetary policy in their publications as a measure of market communication.

The objective of this project is to find latent features in those texts published by FOMC. First, I applied machine learning to economic indices to see the performance of prediction on those numerical data. Then, added pre-processed text data as additional feature in traditional machine learning technique to see if it contains the meaningful information. Finally, apply Deep Learning technique such as LSTM/RNN and BERT to see if these can better predict the rate hike/lower at each FOMC meeting.

**1. Retrieving Market**

Data Daily FED Rate and major economic indices can be obtained from Economic Research in FRB of St. Louis website called FRED: FED Rate GDP CPI / PCE Employment and Unemployment Retail Sales and Home Sales.

***Manufacturing PMI and Service PMI (***formerly known as “Non-Manufacturing Index or NMI) are published by ISM (Institute for Supply Management) website. Daily Treasury yield rates can be downloaded from US Treasury website in xml. Good to explore the details of the data on each website but it’s much more convenient to use Quandl, which provides Web APIs and Libraries to retrieve all the data in the same manner. All the data above are publicly available and free for personal use but you should always check the license terms in the original source in accordance to your objective. Once you create an Quandl Account, API Key is provided. For example, you can download data in python after pip install quandl like this:

Download the following texts from FOMC Meeting Calendar: Statements — available right after each FOMC meeting Meeting Minutes — available three weeks after each FOMC meeting, so may not be available for the latest Meeting Press Conference Transcripts — available at each FOMC meeting but only started in 2011 Meeting Transcripts — available five years after the meeting, so this cannot be used as input for the prediction while still good source to see the detail background for the old meetings Speeches — transcripts are published in this page and I used chair’s speech published between two meetings Testimony — a various testimony texts are also published in this page and I used Semiannual Monetary Policy Report to the Congress.

**3. Preliminary Analysis-**

In order to see if the texts may contain some useful insight to predict FED rate, I used Loughran and McDonald Sentiment Word List to measure the sentiment of statement. This dictionary contains several thousands words appearing in financial documents such as 10K, 10Q and earnings calls categorised to positive, negative, etc. It includes words in different forms, so stemming or lemmatising should not be applied. I applied a simply technique to flip the sentiment for negation (e.g. can’t, isn’t, no). Note that you need to obtain necessary licence for commercial use.

Also check the moving average of net sentiment with actual FED Rate decisions at each FOMC meeting. There’s a certain correlation with FED target rate, but it will not be easy to see during the Financial Crisis where the rate was at Effective Lower Boundary and quantitative easing was taken place. I treated QE announcement as a lowering rate event.

**4. Pre-processing Economic Index**

When FOMC decides the monetary policy, the difference from previous figure is also important. For each index, take difference from the previous period and the same period of the previous year for all the indices.

As a part of feature engineering, calculate taylor rules and see whether the first derivatives and difference from FED rate could be used. FED has released how policymakers use economic indices data on their website. Here, calculate Taylor Rule, Balanced-approach Rule, and Inertial Rule from raw data. The result looks to match with their publication and the correlation between these theoretical rates and actual FED Rates is quite high.

**5. Pre-processing Text Data**

There are around 200 decisions over the last two decades and a half. Depending on the models some inputs cannot be used due to missing data or available timing.

One of common issues you may face during text processing is how to handle long text in machine learning. Most of the neural net based algorithms are not capable to analyse such long texts like 10,000 words — 500 at maximum. Most of our input texts are too long to analyse as a whole document.

Another issues is data imbalance — in this example, rate decision is “hold” for more than 60% chance and available decisions are only ~200 as the meeting is taken place eight times an year. Without having enough data, machine learning can easily overfit to the training data.