**Project Report**

FRE-GY 7871 News Analytics and Machine Learning

Hanyuan Hu (hh1924) Binlin Chi(bc2615)

**Project Name**

5-year Treasury Yield Prediction Based on the Fed Officials’ Speech Sentiments

**Introduction & Motivation**

The purpose of this research is to investigate market impact of Federal Reserve officials’ speeches. It is usually the case that the Fed tend to communicate to the industry through speeches to hint their perspectives about the economy and potential monetary policy they would like to use. It would be helpful if we can use NLP techniques to analyze their speeches and try to produce predictive signals based on this information.

**Objectives**

We observe that the speeches cover several topics including monetary policy review, economic perspectives and also some articles less related to the market directly. Given that, the objectives of this research can be breakdown into steps as follows:

1. Prove/disprove whether some specific articles have more market impact than the others.
2. If 1 is true, try to find out the direction (bearish, bullish) of those impacts.
3. Formulate a trading strategy based on this prediction, and test different techniques to improve its performance measured by Sharp, total return and maximum drawdown, etc.

**Specifications**

1. 5-year treasury yield: This data was obtained from Quandl.
2. Speeches of the Fed officials: we consider all the speeches listed on *Speeches of Federal Reserve Officials* [[1]](#footnote-1)since 2006, totally around 800 samples. We downloaded all the full text by a web crawler we created. Source codes in: *data\_crawler.py.*

**Experiments:**

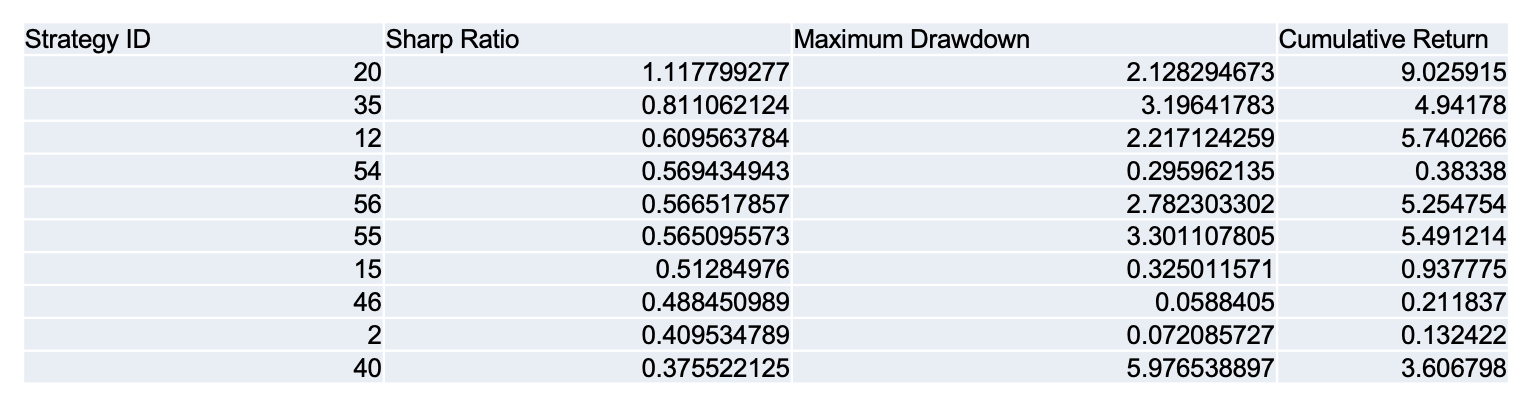
In this research, we tested two learning algorithms, Naïve Bayes with Bagging and Random Forest, on three set of features:

1. Bag-of-Words with or without title embedded;
2. Bag-of-Words with or without title embedded, also added a feature to indicate which subclass within training samples this example should belong to by clustering;
3. Doc2Vec for embedding.

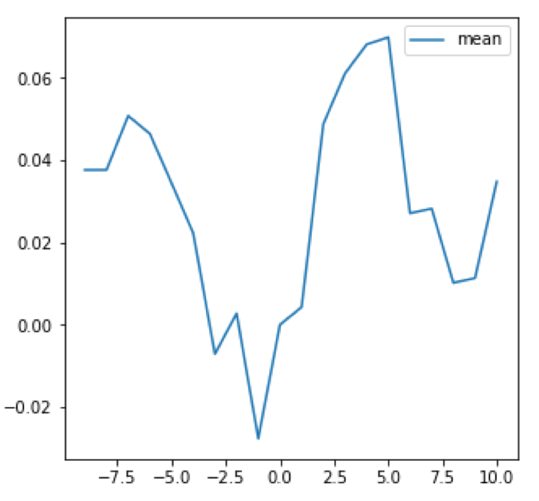
We then train and test the models within a wide range of hyper parameters on the time series dataset on a rolling basis.

**Results:**

1. Bag of words:



Basic Statistics



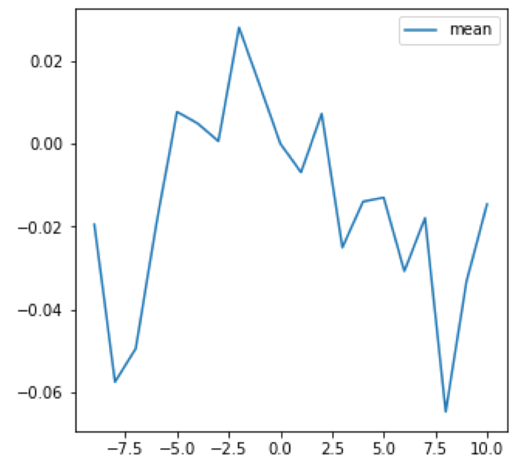
Alpha Decay

1. Using Birch Clustering

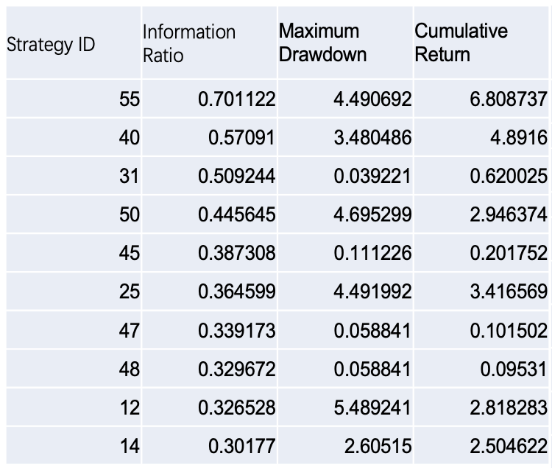
We defined a metric “Explained Variance” to determine the optimal number of groups.

The numerator represents the inter-group variance, and the denominator represents the intra-group variance. Thus, a high EV value indicates a better clustering result. We tried grouping numbers from 2 to 8, and it turned out the optimal group number to be 5.

Applying Birch Clustering method with parameter 5, we got the following result:



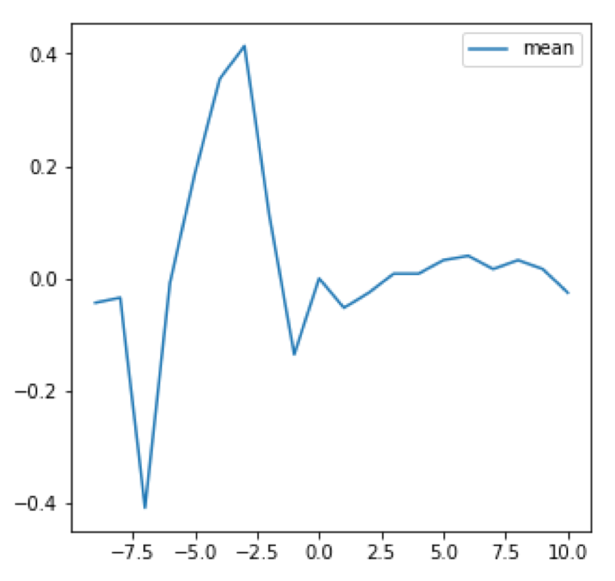
Alpha Decay



Basic Statistics

1. In the last experiment, we use continuous-valued vectors to represent sentences, rather than words. The result turns out to be:

Alpha Decay



Summary of all experiments:

A screenshot of a cell phone

Description automatically generated

The simplest model, bag of words, finally reached the best ranking, followed by Doc2Vec model.

The clustering method, however, performed unsatisfactory. A plausible explanation could be that the optimization process is undermined when we divide one step into two.

**Member Contributions:**

Binlin Chi (bc2615):

Project Infrastructure:

Web Crawler and Parsing (60%)

Model 1:

Empirical Data Analysis (100%)

Feature Engineering (40%)

Model Training and Performance Measure (20%)

Model 2:

Feature Engineering (80%)

Model Training and Performance Measure (80%)

Model 3:

Feature Engineering (20%)

Model Training and Performance Measure (20%)

Presentation (50%)

Hanyuan Hu (hh1924):

Project Infrastructure:

Web Crawler and Parsing (40%)

Back testing framework (100%)

Model 1:

Feature Engineering (60%)

Model Training and Performance Measure (80%)

Model 2:

Feature Engineering (20%)

Model Training and Performance Measure (20%)

Model 3:

Feature Engineering (80%)

Model Training and Performance Measure (80%)

Presentation (50%)

1. <https://www.federalreserve.gov/newsevents/speeches.htm> [↑](#footnote-ref-1)