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]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.

[1] https://www.federalreserve.gov/newsevents/speeches.htm

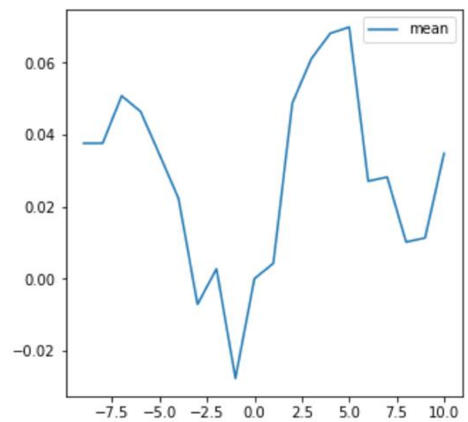
Results:

1. Bag of words:

Basic Statistics

Strategy ID	Sharp Ratio	Maximum Drawdown	Cumulative Return
20	1.117799277	2.128294673	9.025915
35	0.811062124	3.19641783	4.94178
12	0.609563784	2.217124259	5.740266
54	0.569434943	0.295962135	0.38338
56	0.566517857	2.782303302	5.254754
55	0.565095573	3.301107805	5.491214
15	0.51284976	0.325011571	0.937775
46	0.488450989	0.0588405	0.211837
2	0.409534789	0.072085727	0.132422
40	0.375522125	5.976538897	3.606798

Alpha Decay



2. Using Birch Clustering

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

$$EV = \frac{\frac{1}{\mathrm{n}}\Sigma_i^n(\mu_i - \mu)^2}{\Sigma_i \sigma_i^2}$$

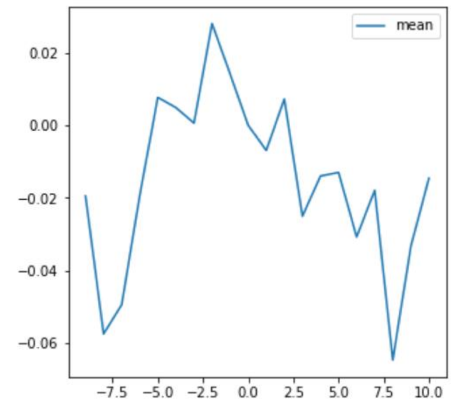
   The numerator represents the inter-group variance, and the denominator represents the intra-group variance. Thus, a high EV value indicates a better clustering result. We tried grouping numbers from 2 to 8, and it turned out the optimal group number to be 5.

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

Basic Statistics

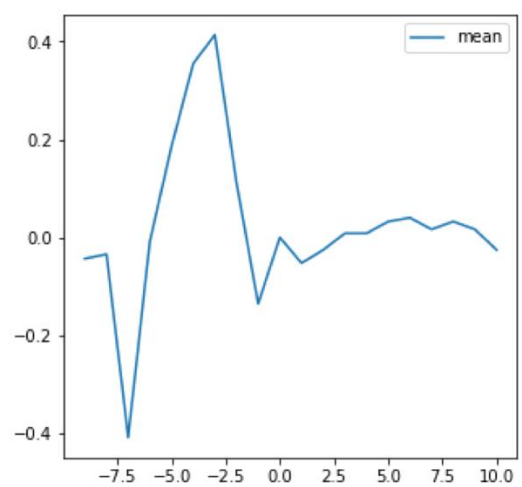
Strategy ID	Information Ratio	Maximum Drawdown	Cumulative Return
55	0.701122	4.490692	6.808737
40	0.57091	3.480486	4.8916
31	0.509244	0.039221	0.620025
50	0.445645	4.695299	2.946374
45	0.387308	0.111226	0.201752
25	0.364599	4.491992	3.416569
47	0.339173	0.058841	0.101502
48	0.329672	0.058841	0.09531
12	0.326528	5.489241	2.818283
14	0.30177	2.60515	2.504622

Alpha Decay



3. 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:

Performance

Version	Strategy ID	Information Ratio	Maximum Drawdown	Cumulative Return
Original	20	1.117799	2.128295	9.025915
Doc2Vec	6	1.013236	2.565984	8.594213
Doc2Vec	33	0.872234	3.577525	6.603261
Original	35	0.811062	3.196418	4.94178
With Clustering	55	0.701122	4.490692	6.808737
Doc2Vec	9	0.676622	0.433335	0.981856
Doc2Vec	8	0.632118	5.338426	5.227318
Original	12	0.609564	2.217124	5.740266
Doc2Vec	17	0.591079	3.283414	4.795537
With Clustering	40	0.57091	3.480486	4.8916
Original	54	0.569435	0.295962	0.38338
Original	56	0.566518	2.782303	5.254754
Original	55	0.565096	3.301108	5.491214
Doc2Vec	0	0.559261	0.526377	1.278248
Doc2Vec	1	0.519884	0.396033	0.633132

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%)