Final research report of NLP class:

Sentiment Analysis of Beige Book and the Prediction for GDP

Quantitative Methods in the Social Sciences

Taketo Muroya (tm3068)

Abstract

Using the NLP (Natural Language Processing) technique, this study tries to generate an economic sentiment index and make a prediction for GDP data based on the text data in the Beige Book published by FRB (Federal Reserve Board). Since this book mentions current economic conditions over time in a relatively consistent manner, the sentiment index generated from the text seems like something related to GDP data. In addition, this study translates the text data into vectors and then, nowcasts GDP data based on the machine learning model using these vectors. Even though the accuracy is still not so high because of the limitations such as serial correlation problem, the text data in the Beige Book could estimate future GDP to some degree.

Introduction

How to assess the current economic conditions using available data in a timely and accurate manner is very important for policymaking or investment. If people do not know what is going on under current economic conditions, it is hard to implement effective economic policies or lucrative investments. However, the U.S. GDP data, which is supposed to represent whole economic activities in the U.S., is usually published more than one month later than the period to which the data refers. For example, the U.S. GDP data for the first estimate of fourth quarter (Oct.-Dec.) 2019 is going to be released at the end of Jan. 2020. Thus, it is difficult to see the immediate economic conditions using traditional economic statistics.

Now, there are plenty of potential resources on the internet that could be used for assessing the current economic conditions. For instance, people might comment on something related to economic issues on SNS. Although many researchers now try to utilize these tweets data for economic or financial analysis, there could be some problems such as data representativeness and inconsistency in phrases. Since relatively younger and richer people tend to use SNS, the comment data might not represent the whole U.S. population. Moreover, the way of tweeting might vary among users, although they talk about one specific economic issue. These difficulties may cause bias in the analysis.

Thus, this study focuses on one of the economic commentaries, the Beige Book. This book is officially called “Summary of Commentary on Current Economic Conditions by Federal Reserve District” because each Federal Reserve Bank (12 banks) gathers information on current economic conditions through reports from interviews with business contacts, economists, market experts, and so on. An overall summary of the 12 district reports is compiled by a designated Federal Reserve Bank on a rotating basis. Since this report is published eight times per year and the date is usually earlier than GDP publication, the information would be useful for GDP prediction. Moreover, the 12 districts (Boston, New York, Philadelphia, Cleveland, Richmond, Atlanta, Chicago, St. Louis, Minneapolis, Kansas City, Dallas, San Francisco) would roughly represent the main regions in the U.S. and the words used in the reports would be relatively consistent within the reports over time, which would be desirable features for the NLP analysis.

Therefore, this study tries to utilize the Beige Book as a source for assessing the current economic conditions using the NLP technique. In the rest of this paper, the data and methods used for generating sentiment index and predicting GDP are spelled out, including how to get and clean the text data. After that, this study mentions its results and conclusions, including limitations such as serial correlation issue.

Data

All text data used for this research are extracted from FRB website using the web scraping technique. This technique would get much information on websites automatically, which is implemented by Python programming in this case. First, the Beautiful Soup, which is a Python package for parsing HTML and extracting the text on it, gets all URLs regarding the Beige Book from 1997.1 to 2019.9. Second, these URLs are split up into three parts because the layout of the website changes over time and it would need different coding for the web scraping. For each part, the Beautiful Soup again extracts the text data, which is the summary part of the Beige Book. During this process, the text is cleaned by regular expression function, which means that the words other than “A-Za-z0-9.” are eliminated. Finally, the text data are stored in the data frame (See Figure 1).

Figure 1: text data (e.g. 2018-2019)

time	body
201801	Overall Economic Activity Reports from the 12 Federal Reserve Districts indicated that the economy contir
201803	Overall Economic Activity Economic activity expanded at a modest to moderate pace across the 12 Federa
201804	Overall Economic Activity Economic activity continued to expand at a modest to moderate pace across the
201805	Overall Economic Activity Economic activity expanded moderately in late April and early May with few shit
201807	Overall Economic Activity Economic activity continued to expand across the United States with 10 of the 1
201809	Overall Economic Activity Reports from the Federal Reserve Districts suggested that the economy expand
201810	Overall Economic Activity Economic activity expanded across the United States with the majority of Feder
201812	Overall Economic Activity Most of the twelve Federal Reserve Districts reported that their economies expa
201901	Overall Economic Activity Economic activity increased in most of the U.S. with eight of twelve Federal Res
201903	Overall Economic Activity Economic activity continued to expand in late January and February with ten Dis
201904	Overall Economic Activity Economic activity expanded at a slight to moderate pace in March and early Apr
201906	Overall Economic Activity Economic activity expanded at a modest pace overall from April through mid Ma
201907	Overall Economic Activity Economic activity continued to expand at a modest pace overall from mid May t
201909	Overall Economic Activity On balance reports from Federal Reserve Districts suggested that the economy expanded at a modest pace through the end of August. Although concerns regarding tariffs and trade policy uncertainty continued the majority of businesses remained optimistic about the near term outlook. Reports on consumer spending were mixed although auto sales for most Districts grew at a modest pace. Tourism activity since the previous report remained solid in most reporting Districts. .....

After the web scraping, the preprocessing tools such as stop words, stemming, and lemmatization can clean the text data. The stop words eliminate useless words, such as ‘a,’ ‘the,’ and ‘at,’ for the text analysis. Since such words are not relevant to the current economic conditions in general, these words could be eliminated. While the stemming reduces each word to its word stem, the lemmatization changes each word to its word lemma. These processes would arrange different forms of a word, such as ‘go,’ ‘goes,’ and ‘going,’ into one simple word ‘go.’ These cleaning processes would be helpful for the machine to understand and classify the words in reports efficiently.

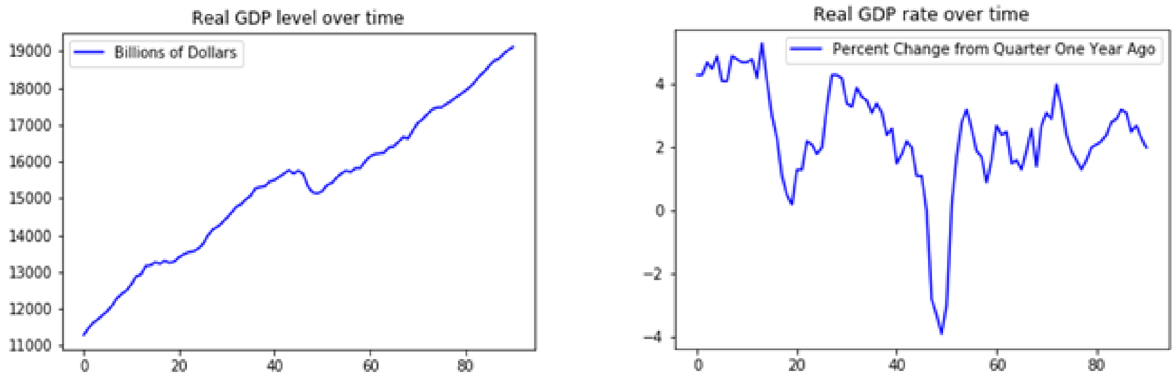
Moreover, using the word cloud package in Python, which visualizes the text by differentiating the size of words based on the frequencies of the words, it is possible to see the contents of the Beige Book visually (See Figure 2). This word cloud would make sure that this book clearly talks about economic conditions because it contains the words such as ‘price,’ ‘increased,’ ‘modest,’ and ‘growth.’

Figure 2: Word Cloud (June 2019)

In addition to the Beige Book, this study uses the quarterly GDP data. All GDP data in this paper, real GDP level, real GDP rate (percentage change from the quarter one year ago), and real GDP rate (percent change from the preceding period), come from the FRED (Federal Reserve Economic Data) database (See Figure 3). The real GDP is usually a better indicator than the nominal GDP in terms of observing the actual economic growth. The time period corresponds to the Beige Book data (1997.Q1-2019.Q3).[1]

[1] Since the frequency of the quarterly GDP data is half of the Beige Book, the GDP data is stretched into eight times per year data by reusing the same value in two periods in a row.

Figure 3: GDP level and rate over time



Methodology

For the first part of the analysis, this paper tries to generate the sentiment index from the Beige Book and then, compare it with the GDP data. Using the positive and negative dictionary, the sentiment function in Python counts a positive word such as “uplift” as +1 and a negative word such as “downfall” as -1 and then, sums up the number of these scores per a report, which would be the economic sentiment index.

For the second part of the analysis, this study tries to predict the GDP using the vectorized text data on the Beige Book as independent variables. The vectorizer packages in Python count the frequencies of words and then, generate the vectors that correspond to the text data (Count Vectorizing). When the package assigns weights on the word frequencies (tf) based on the importance of the words (idf), it could improve the transformation of text data into the numeric vectors in general (Tfidf Vectorizing).

$$Tfidf = tf * idf$$

$$tf = \frac{Term\ X\ Frequency\ in\ a\ Document}{Sum\ of\ all\ Term\ Frequency\ in\ a\ Document}$$

$$idf = log(\frac{number\ of\ all\ documents}{number\ of\ documents\ which\ include\ term\ X})$$

Since the number of raw text vectors would be too large for the machine to learn them successfully, the PCA (Principal Component Analysis) technique serves as the pre-processing method for the GDP prediction, which means that the PCA transforms the variables into reduced ones by finding the linear combinations of the variables that have the maximum variance and are uncorrelated with each other. In this paper, keeping 95% of original variances, the dimension of text vectors is significantly reduced by the PCA (from more than 6,000 vectors to around 140 vectors). After these processes, text vectors are ready to be independent variables for the GDP prediction model.

The machine learning model that this study uses for the nowcasting GDP using text vectors is the random forest method,[2] which is the ensemble algorithm of the decision tree method. The decision tree method builds up the best tree diagram that regresses the outputs well by minimizing the RSS (Residual Sum of Squares) among the results. The random forest model enhances the accuracy of this decision tree method by randomly implementing the tree build-up process again and again, and then taking the average of the prediction outputs, which would improve the accuracy of one decision tree method

[2] This is because the accuracy scores of KNN (K Nearest Neighbors), Ridge regression, and Lasso regression are actually worse than that of the random forest model for this GDP prediction model.

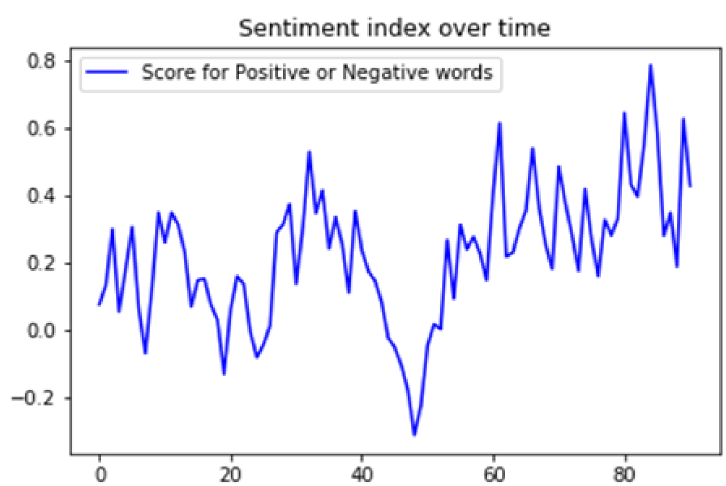
significantly. In this study, the dependent variable is real GDP rate (percentage change from the quarter one year ago) and the independent variables are the generated text vectors. Although the one decision tree regression is usually not superior to the OLS (Ordinary Least Square) regression, the random forest regression tends to score a higher accuracy as the number of decision trees increases.

$$GDP\ rate_t\ =\ f\left(X_{1t,} X_{2t,} X_{3t,} X_{4t,},,,\right) \quad where\ \ X = text\ vectors$$

Results

For the sentiment analysis of the Beige Book, this index seems like the combination of both GDP level and GDP rate because it has an upward trend and also captures some booms and recessions in GDP rate (See Figure 4). Even though the correlation between the index and GDP rate is only 0.44, this index would make it possible to assess economic conditions without reading all text in Beige Book.

Figure 4: The economic sentiment index



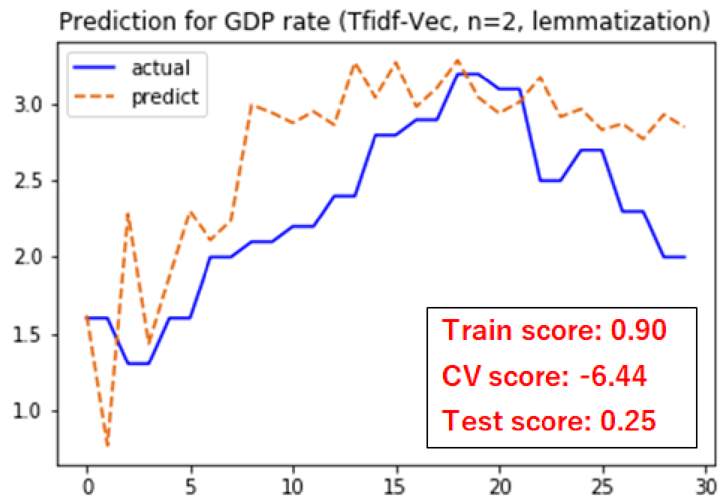
For the prediction for GDP, this paper examines the performance of the multiple random forest models (the number of trees is set as 1,000 here) by differentiating cleaning methods (stop word only, stop word & stemming, or stop word & lemmatization), n-grams (1, 2, 3, or 1-3), and vectorizing methods (Count or Tfidf). The result shows that the random forest model with stop word & lemmatization, n-gram=2, and Tfidf-Vectorizing would be the best model based on the test score (See Table 1). However, the accuracy score is not so high, even though the predicted values roughly have similar trend to the actual GDP rate during the test period (2016-2019) (See Figure 5).

Table 1: Random forest model selection[3]

	n-gram:1 Tfidf-Vec	n-gram:1 Count-Vec	**n-gram=2 Tfidf-Vec**	n-gram=2 Count-Vec	n-gram=3 Tfidf-Vec	n-gram=3 Count-Vec	n-gram:1-3 Tfidf-Vec	n-gram:1-3 Count-Vec
Stop Word	Train:0.90 CV:-7.49 Test:-2.79	Train:0.91 CV:-6.18 Test:-3.16	Train:0.91 CV:-6.82 Test:-1.91	Train:0.89 CV:-6.56 Test:-0.88	Train:0.88 CV:-8.40 Test:-2.08	Train:0.86 CV:-8.47 Test:-1.41	Train:0.92 CV:-5.55 Test:-0.87	Train:0.91 CV:-5.85 Test:-2.03
Stemming	Train:0.93 CV:-6.43 Test:-1.04	Train:0.92 CV:-7.16 Test:-0.91	Train:0.90 CV:-7.74 Test:0.21	Train:0.89 CV:-7.26 Test:-1.81	Train:0.87 CV:-8.56 Test:-0.30	Train:0.88 CV:-7.49 Test:-0.40	Train:0.92 CV:-7.47 Test:-0.24	Train:0.91 CV:-7.23 Test:-0.63
Lemmatization	Train:0.91 CV:-6.10 Test:-0.36	Train:0.92 CV:-5.77 Test:-2.28	**Train:0.90** **CV:-6.44** **Test:0.25**	Train:0.89 CV:-6.73 Test:-0.96	Train:0.87 CV:-7.83 Test:-1.08	Train:0.87 CV:-7.41 Test:-0.83	Train:0.93 CV:-5.52 Test:-1.23	Train:0.89 CV:-6.36 Test:-1.29

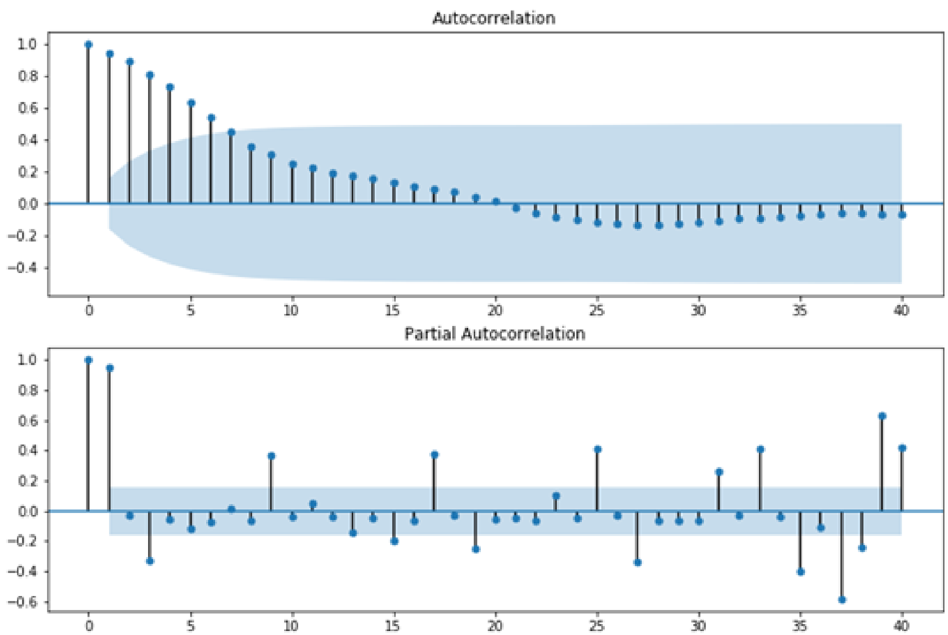
[3] In this table, Stop Word = Eliminate stop words, Stemming = Eliminate stop words + Stemming, and Lemmatization = Eliminate stop words + Lemmatization. Train = Training set score, CV = Cross Validation score, Test = Test set score.

Figure 5: GDP prediction (Random Forest)



One of the reasons that the Cross Validation and Test set score are low in this model would be the serial correlation (or autocorrelation). The serial correlation issue, which is commonly dealt with in the time-series analysis, could affect the accuracy of prediction here because all data that this paper uses are time-series data. Although the random forest method is not the simple regression model such as OLS, the serial correlation could be a problem in machine learning methods as well, which should be taken into account. The serial correlation means that residuals in the regression are correlated with each other over time, which would make estimators biased and non-consistent because the expectation of residuals conditioning independent variables would be non-zero. Actually, the graph regarding Autocorrelation and Partial Autocorrelation shows that there would be a serial correlation in the GDP rate data (See Figure 6).

Figure 6: Autocorrelation and Partial Autocorrelation in GDP rata data



To deal with this serial correlation problem, it could be important to apply a time-series model, such as the ARMA model,[4] to the random forest model. The ARMA model could fix the serial correlation problem to some degree by adding Autoregression and Moving Average terms to the model.

$ARMA\ model\ (e.g.\ AR(1)\ \&\ MA(1))$:

$$GDP\ rate_t\ =\ \beta_0 + \boldsymbol{\alpha_1 GDP\ rate_{t-1}} + \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots + e_t + \boldsymbol{\theta_1 e_{t-1}}$$

To find a desirable ARMA model, this paper examines AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) by differentiating the number of AR

[4] The ARIMA model is also one option to estimate the time-series model. However, ARIMA model which includes differencing items in the equation does not improve the accuracy in this GDP prediction model.

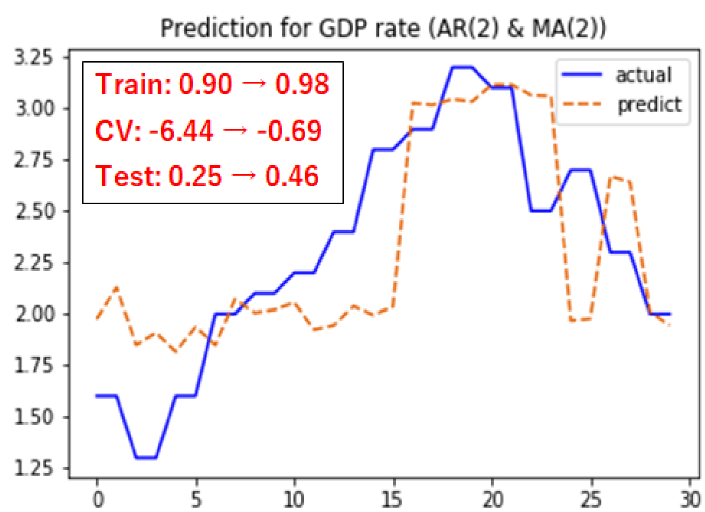
and MA using the time-series OLS regression in Python (See table 2). As a result, the AR(2) and MA(2) model would be a desirable model (lower AIC or BIC is a better model).

Table 2: ARMA model selection

	AR	MA	AIC	BIC
1	0	0	269.57	714.08
2	0	1	255.92	703.45
3	0	2	265.62	716.17
4	1	0	282.62	730.15
5	1	1	227.95	678.51
6	1	2	268.31	721.9
7	2	0	254.32	704.88
8	2	1	258.48	712.06
9	**2**	**2**	**220.54**	**677.15**

Thus, this study adds AR(2) and MA(2) terms into the random forest model, and then this model improves the accuracy scores of GDP prediction significantly (See Figure 7).

Figure 7: GDP prediction (Random Forest with ARMA model)



Conclusion

This paper tries to generate the economic sentiment index from the text data on the Beige Book and predict the GDP data using the vectorized text data on this book. As for the sentiment index, it roughly captures the features of economic conditions, which could be useful for the economic analysis. If a special dictionary of economic sentiment words were available, the accuracy of the index could improve.

As for the GDP prediction, it also predicts the future GDP rate to some degree, although the accuracy seems to be not high enough for using it for the actual decision making. There are some possible reasons for that. First, as this paper uses unlabeled data for the supervised learning, combining GDP data as target data and Beige Book as learning data, it would be hard to get the high accuracy. Second, since the Beige Book only includes the 12 districts in the U.S., it might not be sufficient to represent whole U.S. economy. In that case, the other economic or financial reports could complement the model for further studies. Finally, the serial correlation problem in time-series data could affect the accuracy. Since even the ARMA model does not have more than half accuracy for the test score, more advanced models would be needed. However, once this method is well established, it could be applicable for any other economic or financial reports, which would be very useful for assessing the current economic conditions in a more timely and accurate manner.

References

Brandyn, Bok & Daniele, Caratelli & Domenico, Giannone & Argia, Sbordone & Andrea, Tambalotti. (2017). *Macroeconomic Nowcasting and Forecasting with Big Data.* Federal Reserve Bank of New York Staff Reports, no. 830.

Yuhki, Yamamoto & Yutaka, Matsuo. (2016). *Sentiment Summarization of Financial Reports by LSTM RNN model with the Japan Economic Watcher Survey Data.* The 30th Annual Conference of the Japanese Society for Artificial Intelligence.

Thomas, B. Götz & Thomas, A. Knetsch. (2019). *Google data in bridge equation models for German GDP.* Deutsche Bundesbank, Germany. International Journal of Forecasting 35. Pages 45-66.

Daniela Bragoli. (2017). *Nowcasting the Japanese economy.* International Journal of Forecasting 33. Pages 390-402.

William H. Greene. (2018). *Econometric Analysis, 8th Edition.* Pearson.