BREXIT and the News

Authors : Andres Gomez, Nicole Wu, Huaixin Wang, Habiiba Malingha

Course : Data Science Connector, Web Data Visualization (L&S 88)

Instructor : Yasmin Alnoamany

Date: December 2017

Introduction **3**

Related Work **4**

Data **4**

Gathering Data 4

Cleaning Data 5

Visualization 5

Final Thoughts **11**

References **12**

Introduction

Great Britain voted for Brexit June 23rd, 2016, and referendum was to give British people a choice to stay or to exit from EU. The referendum was proposed by David Cameron because he believed that British people should decide that issue, even though he recommended staying in the EU. The media coverage of the Brexit issue has been covered not only by all major newspapers but also by online media like Facebook and Twitter. Our work will focus on that media coverage. We hope to better understand the coverage before the vote, how people responded to the vote, and how it was covered by NYTimes. Our investigation focus could be formulated as follow: "What can we learn from a certain topic using data analysis and particularly data mining?"

Related Work

The idea of our study is based on the current literature. The study of Prismoji (2016) focuses on the reactions of Brexit through the uses of emojis on Twitter (Prismoji 2016). However this is not an academic paper and the study was done on two days (the day before and the day after brexit). The paper of Celli, Stepanov, Poesio and Riccardi (2016) uses natural language processing in order to analyse twitter feed. That analysis allow them to predicted accurately the result of Brexit. This paper confirms the idea that the uses of natural language can be use in order to predict or analyses a political event. Our paper will mainly focus on the latter. Seaton's research (2016) on the reception of Brexit in social media, which discusses political polarization manifested through the discourses of Brexit on social media. Press, especially the pro-leave right

wing press, has become more partisan, which contributes to the binary and partisan nature of the Brexit discourse.

Data

Gathering Data

We focus in data mining and particularly on the NYtimes. We uses the free version of the NYtimes API which has some limitation that are discussed in the limits. We collect the information, such as headline, snippet, location, subject, date, url and article ID, of news in NYTime regarding "Brexit". The data has no specific period cover at first. However the first use of the world "Brexit" appears in beginning of 2013 and the last use of the word was on november 2017 when we gather the data. Because of this our analysis focus on those period.

Cleaning Data

We had to convert dates in number format (i.e., 43275) to dates in date format (i.e. 2013-01-01) that we can interpret. We removed data that was from 'briefings' because the briefings were just titles that barely said things we can use to accurately visualize. They were almost the same as the snippets from the articles we were using for our analysis. We also do the tokenization, remove the punctuation and stopwords as well as stemming for the text information.

Visualization

As we discuss before our focus, this paper is to show if it is possible the main issues about a topic based only on data mining. In a first instance we computed the ten words with the most

appearance (figure 1) excluding the punctuation and stop words. An interesting fact is that the word “european” twice as much to appear in comparison to the word “britain”. Another is the appearance of the words “vote” and “leave” and the absence of the word “stay”. The word “stay” don’t even appear in the first twenty words. In order to have a better overview of the most occurred word we have created two visualisation figures. The bar chart only represents (figure 1) the 10 word with most occurrence and in the figure 2 we represent the word with most occurence in a “ word cloud”.

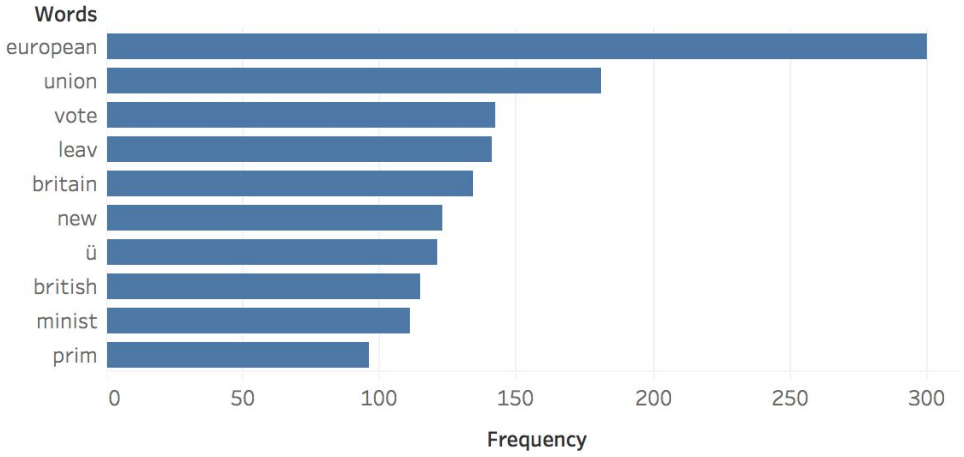

Figure 1 : Top 10 most frequent words in NYTimes

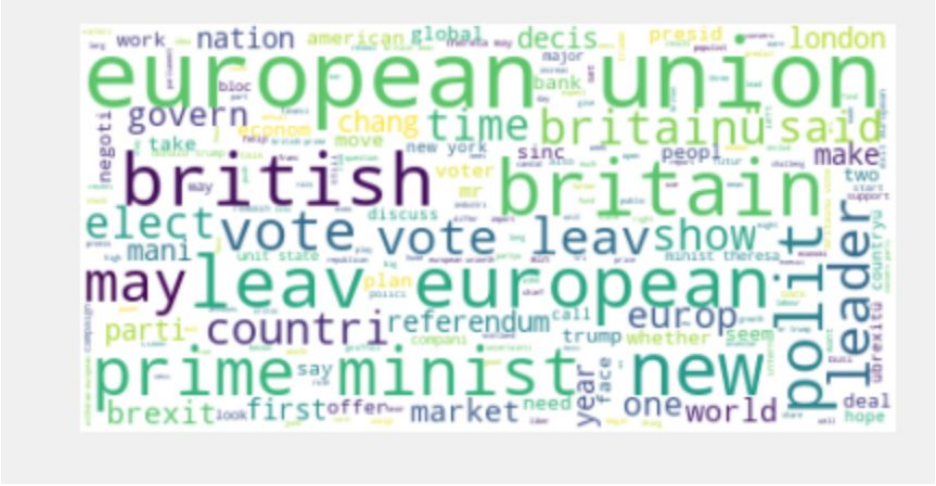

Figure 2 : Most frequent words in NYTimes

In a second instance we want to see if there was a relation between the number of articles related to brexit and specific events for Brexit in reality. The first appearance of the word "Brexit" in the new york times was in 2013 (one occurrence). We can see that the word keep appearing with few occurrences till 2016. At that point, it increased rapidly till it reaches its peak in june 2016 followed by a decrease in july 2016 and stabilize with great variation (figure 3).

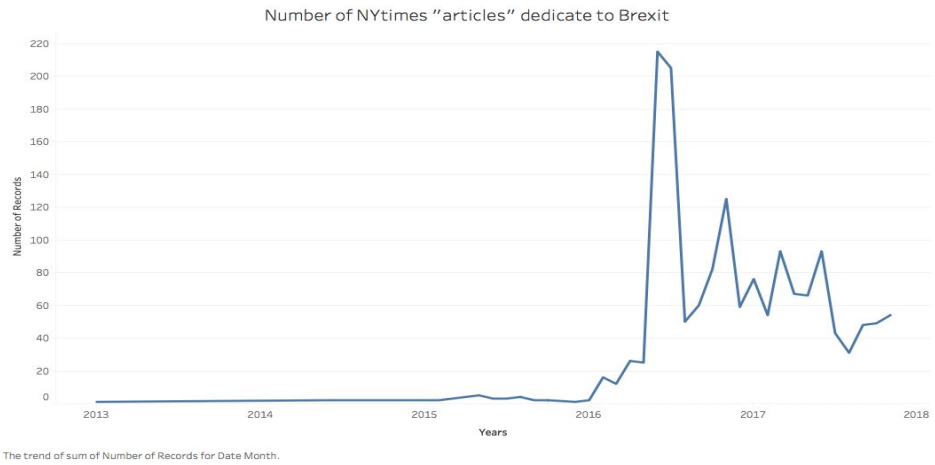

Figure 3: Articles dedicated to brexit for the period 2013-2017

We compare the number of articles over time with the Brexit Timeline (figure 4) which represent the most important political events related to Brexit according to the financial trading (2017).

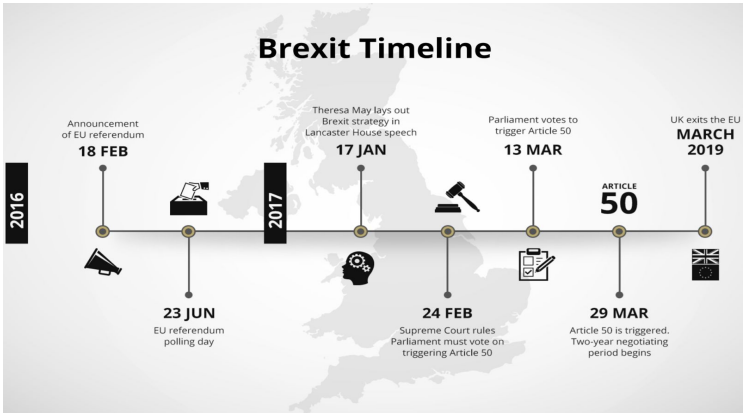

Figure 4: Brexit Timeline

The comparison between the coverage timeline (figure 3) and the political timeline (figure 4) allows us to see the relation between a political event and how it is covered by a foreign newspaper. This relation seems obvious in relation to brexit but this shows that web scraping could potentially single out important event even for topics that we don't possess any knowledge.

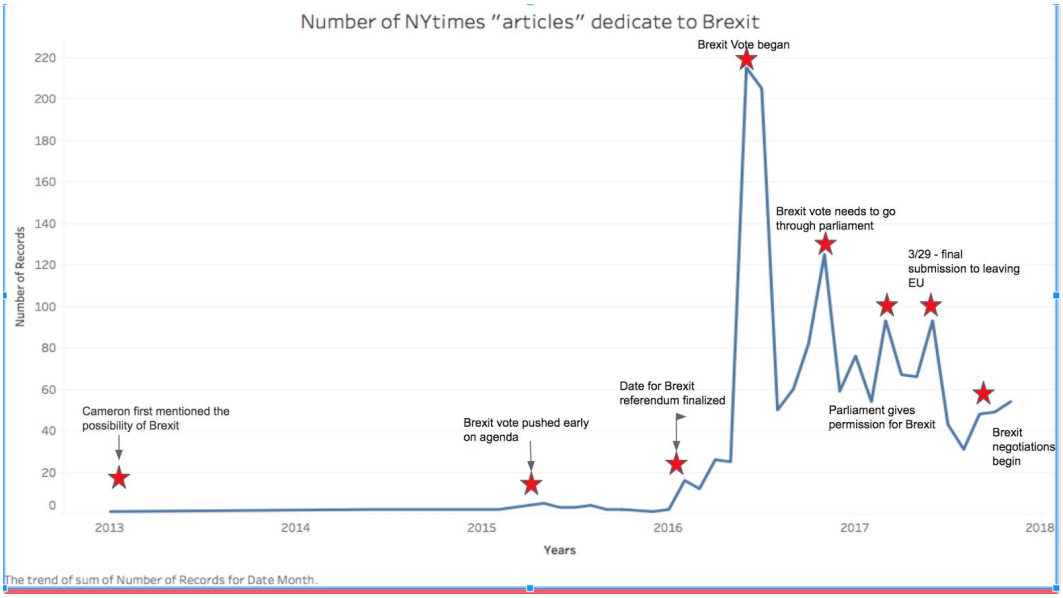

Figure 5: Articles dedicated to brexit merged with brexit timeline

In order to better understand the main concerns about brexit we create a "word cloud" for every year from 2013 to 2017. In 2013 the most common words on the New York times were "new, economy, elections" (Figure 6). Those three words match the political timeline, in 2013, when Cameron announce the referendum to leave the european union. However, because there is only one occurence for 2013 those words are based on a single article and require caution for that reason.

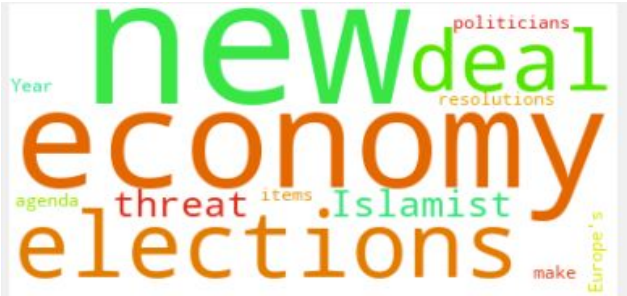

Figure 6: Most common words for the year 2013

The same is true for 2014 where the appearance of "european and union" which both refer to the European Union. Those two words always appear from 2014 to 2016 in the top 5. They disappear in 2017 where there is the occurence of words is more homogenous. This change of focus is really interesting because 2017 is the first year after "Brexit" election. Even if there is still negotiation about "brexit" between European Union and Great Britain the focus in not any more on the "European Union". The year 2017 is particular as well with the appearance of the world "Trump". This is particularly interesting because the world neither the word "obama" nor "barack" appear in the previous year. Was "brexit" only relevant under "Trump"? Looking back in the news Obama discourage "Brexit" The Guardian (2016, April 22) and Trump (The Guardian, 2017, January 16) encourage it. Why did Trump apparently received more coverage than Obama. This paper is not able to answer this question however with a deeper analysis it will be interesting to look for indicators that could explain a difference of media coverage.

Figure 6: Most common words for the year 2014

Figure 7: Most common words for the year 2015

Figure 8 : Most common words for the year 2016

Figure 9 **:** Most common words for the year 2017

Conclusion

The analysis we have done allow us to see that one of the main focus of Brexit was about the European Union. We also tried to see if we could analyses a particular stand of the New York Times on this issue however this result being an impossible task. A quantitative analysis only allows us only to frame the main topics around “brexit”. The word “leave” is a good example. The word leave has been used in many different ways, in certain context “leave” was used to report fear (ex:“many European countries fear infection by the British vote to leave their union.”), sometimes to announce as positive news in regards to the US (ex: “Britain vote to leave the European Union bolstered the dollar”) and other times it was only descriptive (ex: “England voted to leave the E.U., but Northern Ireland did not”). This example show that we cannot infer

opinion on a particular word because we have no context and show the limit of a purely quantitative analysis. We can however determine the main thematic around Brexit. From our text analysis, we can see words such as European Union, vote, Britain and leave reflect the perceived concerns of Europeans and the British since Brexit began. Tableau shows that the word Brexit was more frequently used in 2016- around the referendum year - and we can notice that this was also when US presidential campaigns were happening. As we show in our analysis the main thematics are not static and evolve through time. We have seen that European Union was the main thematic from 2014 to 2016 but data suggest a shift in 2017. This project suggest that data mining allows to explore the main themes around a specific thematic. Future projects could include data in 2018 and API of other newspaper as British Newspaper in order to show the consensus or the differences. Do New York Times and Guardian treat the information in a similar way?

Final Thoughts

All APIs have a call limit. Most APIs are rate limited to 1,000 calls per day, and 5 calls per second. The Article Search API is rate limited to 1,000 calls per day, and 1 call per second (New York Times 2017)

Brexit is a word that first appeared around 2013. However, the debate over Great Britain leaving EU is older than that, and this is only a starting date. Some other words should have been merged

in order to have a better understanding. The words “European”, “Union” or “Europe” are counted as different words when in reality they express the same concept.

There were considerable difficulties in setting a time frame while collecting data through the Twitter API when we were executing our original plan. Therefore we repositioned our topic to gather the information through the API of important newspaper, and we chose the NYTimes. We collected information through NYTimes’ API on NYTimes’ reporting on Brexit.

For future work, it will be important to analyze more texts to reflect more themes surrounding Brexit especially.

References

Celli, F., Stepanov, E., Poesio, M., & Riccardi, G. (2016). Predicting Brexit: Classifying agreement is better than sentiment and pollsters. In *Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media (PEOPLES)* (pp. 110-118).

Prismoji. (2016, June 30). Https://prismoji.com/2016/07/01/brexit-emojis/. Retrieved November 30, 2017, from https://prismoji.com/2016/07/01/brexit-emojis/

New York Times (n.d.). The New York Times Developer Network. Retrieved November 24, 2017, from http://developer.nytimes.com/

Financial Trading (2017, March 31). Brexit: Trading Opportunities. Retrieved December 01, 2017, from http://www.financialtrading.com/brexit-trading-opportunities/

Seaton, J. (2016). Brexit and the media. *The Political Quarterly*, *87*(3), 333-337.

The Guardian (2016, April 22). Barack Obama: Brexit would put UK 'back of the queue' for trade talks . Retrieved December 01, 2017, from https://www.theguardian.com/politics/2016/apr/22/barack-obama-brexit-uk-back-of-queue-for-trade-talks

The Guardian. (2017, January 16). Donald Trump's first UK post-election interview: Brexit a 'great thing' . Retrieved from https://www.theguardian.com/us-news/2017/jan/15/trumps-first-uk-post-election-interview-brexit-a-great-thing