Proceedings of the Fourteenth International AAAI Conference on Web and Social Media (ICWSM 2020)

The Long-Running Debate about Brexit on Social Media

Emre Calisir, Marco Brambilla
Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano, Milan, Italy
firstname.lastname@polimi.it

Abstract

Online social media platforms have become a major place where people also discuss their opinions and express their feelings about socio-political phenomena such as elections and referendums. Human-generated online content is a fruitful resource for a deeper understanding of these happenings. In this study, we present a dataset comprising 45 months (from January 2016 until September 2019) of long-running discussions on Twitter about the Brexit referendum, which can be used by social scientists and journalists for understanding the evolution of the public debate about the phenomenon. This dataset comprises 50.8 million tweets and 3.97 million users, and is also enriched with additional meta-data attributes: bot score of users, sentiment information detected by our sentiment analyzer, political stance information predicted by our stance classifier. Considering all Brexit related tweets of users during our time period, we also determine their overall stance and sentiment.

Introduction

Social media provides many opportunities to monitor and evaluate political phenomena such as referendums and elections. Citizens from all around the world, voters, politicians, private and public authorities participate and contribute to debates on social media platforms with tremendous interest. According to a survey, 66% of social media users have employed these platforms to post their thoughts about civic and political issues, react to others' postings, press friends to act on issues and vote, follow candidates, like and link to others' content, and belong to groups formed on social networking sites (Lee Rainie and Verba 2012). In this context, Twitter is known as one of the most convenient social media platforms.

In this work we discuss a dataset and a study that concerns one of the most relevant political events of recent times, which defines the process of the United Kingdom's exit from the European Union (EU), informally named Brexit. On 23 June 2016, the United Kingdom voted to leave the EU, by 51.9% for Leave, and 48.1% for the Remain side. However, the local and global impacts of the referendum have made the issue a highly active and long-standing discussion well beyond the end of the referendum. Having this dataset will provide an extensive ability to characterize the Brexit referendum, from the aspects of political stance, sentiment, and involvement of bot accounts. Additionally, the use of this data will allow political scientists and social scientists to apply further analyses for specific objectives.

The rest of the paper is organized as follows: We first describe the data collection method, and then we present a descriptive analysis of the dataset. Then, we introduce our use cases, the methods we use and details of our implementations. We visualize our findings to give the readers an intuition about how beneficial is to use this dataset. We conclude the paper by presenting the recent public datasets that are similar to our work, and we finalize the paper with the conclusion section.

Resources

Stance classification implementation is available online under Apache License Version 2.0 at the url:
https://github.com/DataSciencePolimi/CSSforPolitics.

The full dataset is available on Harvard Dataverse with assigned DOI and is accessible at the following url:
https://doi.org/10.7910/DVN/KP4XRP. (Calisir and Brambilla 2020)

Dataset

Data Collection

To collect the Brexit related posts from Twitter, we used a Web scraper framework that connects to the Twitter Search web page and collects the query results. Our search query is to download all tweets containing the Brexit keyword posted between January 2016 and September 2019. We removed the tweets posted in a language other than English from our dataset.

Data Description

Our data collection strategy resulted in obtaining 50.8 million tweets posted by 3.97 million Twitter users (Table 1). 50% of users in our dataset tweeted about Brexit only once.

Due to Twitter's privacy requirements, we share only the Tweet IDs and User IDs. By using these IDs, other attributes

Table 1: Descriptive Stats of tweets and users

Metrics	Count
`Num of Tweets`	50897533
`Num of Users`	3979965
`Num of Users having bot score in dataset*`	1735674*
`Mean of bot score`	0.20
`Median of bot score`	0.11
`Num of bot accounts(users)**`	57527
`Mean of Tweets per User`	12.7
`Median of Tweets per User`	2
`Num of users with Negative sentiment`	1061899
`Num of users with Positive sentiment`	1067209
`Num of users with Neutral sentiment`	1094410
`Num of users with Mixed sentiment`	756447
`Num of users with Remain stance`	1106434
`Num of users with Leave stance`	440971
`Num of users with Others stance`	2432560

* Due to the API limitations of Botometer service, we had bot score results only for the 43% of users
** According to the Botometer service, a Twitter account shows bot behavior if its score is higher than 0.8

Table 2: Description of *Tweet* and *User* entities provided in the dataset

Entity	Features
`Tweet`	ID, political stance of tweet, sentiment of tweet.
`User`	ID, political stance of user, sentiment of user, bot score, bot API request time, num of Brexit related tweets of user.

of tweets and users can be retrieved. In addition to Twitter IDs, we implemented several use cases to enable further analysis on our dataset.

We first tagged tweets and users with political stance label. To obtain the political stance of the tweets, we transformed the textual content of tweets into features and performed a Machine learning classification using Support Vector Machines. Secondly, we tagged tweets and users with sentiment information using an AFINN lexicon-based sentiment analyzer. Finally, in order to observe the relationship between stance, sentiment and tendency of being a bot account, we collected the bot scores of Twitter accounts available in our dataset. Table 2 describes the entities and features.

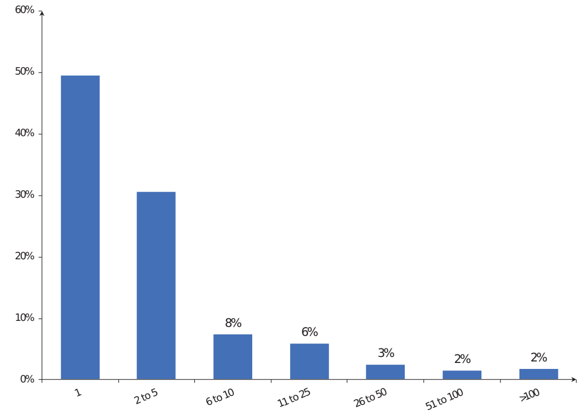

Figure 1: Tweet post frequency of users related to Brexit

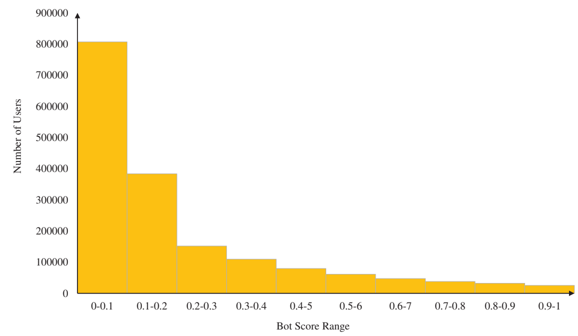

Figure 2: Bot score range of Twitter users

Use Cases

Bot Analysis

To evaluate the bot account behavior of Twitter users, we used the state-of-the-art bot detector (Davis et al. 2016) [1]. This service assigns a bot score to a Twitter account in the range (0,1) describing how likely it is to be an automated account with 1 being the maximum probability. Due to the API rate limits of the service, we collected bot score information for the %43 of users. As shown in Figure 2, most of the users do not have bot behavior, only 3% of users can be classified as bot accounts.

Sentiment Analysis

For the sentiment analysis, we used the AFINN lexicon-based sentiment analyzer which produces a score between -5 and +5. This is one of the simplest and most popular lexicons that can be used extensively for sentiment analysis. In this method, the sentiment score of a text is simply calculated by dividing the sum of each token to the number of tokens. To increase the confidence interval, we annotated a tweet as negative if the score is below -0.2, positive if above 0.2, neutral in other conditions.

To calculate the user-level sentiment from these labeled tweets, we applied the following formula:

[1]Botometer https://botometer.iuni.iu.edu/

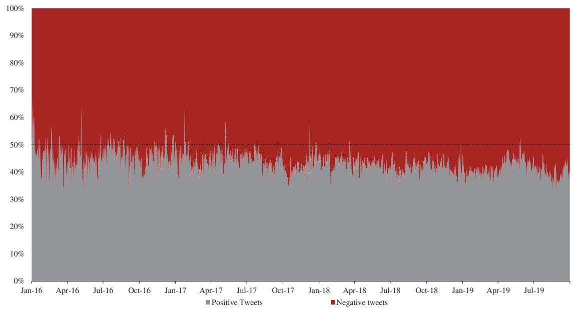

Figure 3: Number of tweets with Negative sentiment are consistently higher than the Positives, 13 percent points higher in average.

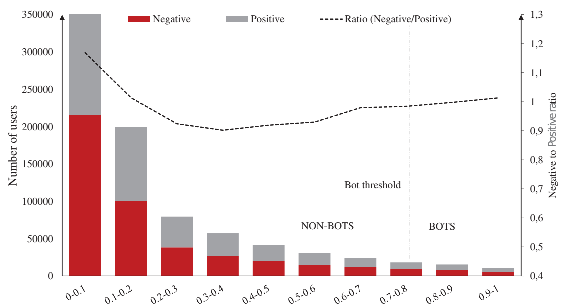

Figure 4: Bot score analysis of Twitter users regarding their user-level sentiment

- A user has Neutral sentiment if he/she has not any Positive or Negative tweets in the dataset
- A user has Positive sentiment if num of Positive tweets/(num of Positive+Negative+Neutral tweets) ≥ 0.5
- A user has Negative sentiment if num of Negative tweets/(num of Positive+Negative+Neutral tweets) ≥ 0.5
- A user labeled as Mixed for the rest of the conditions.

This method provided us to obtain 1067209 positive, 1061899 negative, 1094410 neutral and 756447 mixed sentiment users.

Temporal sentiment analysis of Brexit related tweets Our temporal analysis shows that the number of negative tweets is consistently higher than the number of positive tweets in the subjected time period (Fig. 3).

Bot analysis of users combined with user sentiment An additional analysis can be made by combining bot scores and user-level sentiment. We found that there is no strong correlation between an account's bot behavior and the sentiment of tweets of that account (Fig. 4).

Political Stance Classification

Twitter becomes a center point of discussion when the political happenings are highly polarized among people. Many users express their side by using specific hashtags, for instance, #Remain or #Leave in Brexit related tweets. For this reason, many studies in the literature are hashtag-based stance calculations. However, we found in our analysis that the number of Twitter users who use these hashtags is proportionally very small compared to the whole audience. Our method takes into account the whole textual content of text instead of only hashtags.

Stance classification is a highly challenging task since we expect from an algorithm to find the stance of a user from a short piece of text. In a recent public stance classification open task, the SVM-based model had the best performance achieving to 0.65 of F1 score (Mohammad et al. 2016). In our implementation, we first applied basic text preprocessing operations including removal of stopwords, emojis, weblinks. For the feature generation part, we followed a Tf-Idf transformation and bag of n-gram pipeline producing unigrams, bigrams and trigrams. As the machine learning classifier, we practiced several trials with Logistic Regression, Random Forest and Support Vector Machines (SVM). SVM with a linear kernel performed 0.02 points in average better than the other algorithms. Basically, SVM tries to maximize the margin between classes. In our implementation, we trained 2 different classifiers, one for the classification of Pro-Remain and one for the Pro-Leave side. Classifiers have distinct training/test datasets that are chosen randomly from our dataset and labeled by us thanks to our two years of expertise in the Brexit analysis on Twitter. Train/test datasets are balanced: Pro-Remain classifier is trained and tested with 1470 Remain, 1470 Non-Remain (735 Leave and 735 Neutrals) labeled tweets, and Pro-Leave classifier is trained and tested with 1316 Leave, 1316 Non-Leave (658 Remain and 658 Neutrals) labeled tweets. We applied 10-fold cross-validation to assess the performance of classifiers.

SVM-based Pro-Remain and Pro-Leave classifiers had 0.66 and 0,69 of F1 scores respectively. As a complementary step to the basic implementation, we also tested the classifiers with difference confidence intervals, and after validating the accuracy in the test set, we set the confidence interval thresholds as 0.3 and 0.7 (See Table 3). This approach resulted to obtain a better prediction performance. The idea of discarding 34% and 32% of vaguely predicted tweets for Pro-Remain and Pro-Leave classifiers increased the F1 scores to 0.76 and 0.78. By using this confidence interval, we performed a prediction over the whole Tweet dataset and we associated Pro-Remain and Pro-Leave labels only with the tweets having a high confidence score. Eventually, to calculate the user stance from these labeled tweets, we applied the following formula:

- A user has Pro-Remain stance if num of Remain tweets/(num of Remain+Leave tweets) ≥ 0.5
- A user has Pro-Leave stance if num of Leave tweets/(num of Remain+Leave tweets) ≥ 0.5
- A user labeled as Other for the rest of the conditions.

This method provided us to obtain 1106434 Pro-Remain, 440971 Pro-Leave, and 2432560 Others stance users. It should be noted that due to the highly challenging nature of the stance classification task, we obtained a small but highly confident subset of polarized users.

Table 3: Confidence intervals let us discard the vaguely predicted tweets, keeping in mind that having a higher confidence interval causes to data loss while increasing the performance. In this trade-off, we selected confidence intervals as [0-0.3] U [0.7-1], eventually, we labeled all of tweets as outside of this interval as Others

#	Conf.Interval thresholds to keep predictions	Precision	Recall	F1	AUC	Classifier	Data	% of Tweets dropped*	Num of Predicted Tweets as class 1**	Num of Predicted Tweets as class 0***	Num of Overlapping Tweets****
1a	Default (0.5)	0.66	0.66	0.66	0.74	Remain	1470 remain, 735 leave, 735 others	0%	1621	1302	19
1b	Default (0.5)	0.689	0.686	0.685	0.76	Leave	1316 leave, 658 remain, 658 others	0%	1426	1177	19
2a	[0-0.3] U [0.7-1]	0.74	0.74	0.74	0.80	Remain	1470 remain, 735 leave, 735 neutrals	34%	1120	818	4
2b	[0-0.3] U [0.7-1]	0.75	0.75	0.75	0.79	Leave	1316 leave, 658 remain, 658 others	32%	1026	755	4
3a	[0-0.2] U [0.8-1]	0.78	0.78	0.78	0.82	Remain	1470 remain, 735 leave, 735 neutrals	51%	851	571	3
3b	[0-0.2] U [0.8-1]	0.79	0.79	0.79	0.83	Leave	1316 leave, 658 remain, 658 others	49%	823	514	3

* Number of tweets dropped by classifier from test set based on the corresponding confidence interval threshold.
** Class 1 refers to Pro-Remain tweets for the Remain classifier and Pro-Leave tweets for the Leave classifier.
*** Class 0 refers to Neutral and Pro-Leave tweets for the Remain classifier, Neutral and Pro-Remain tweets for the Leave classifier.
*** Specifies tweets that are classified as class 1 by both classifiers (Pro-Remain and Pro-Leave) for a given confidence interval threshold

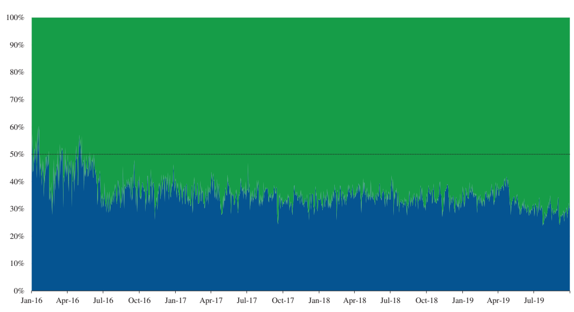

Figure 5: There is a decreasing trend in number of Leave tweets compared to Remain tweets

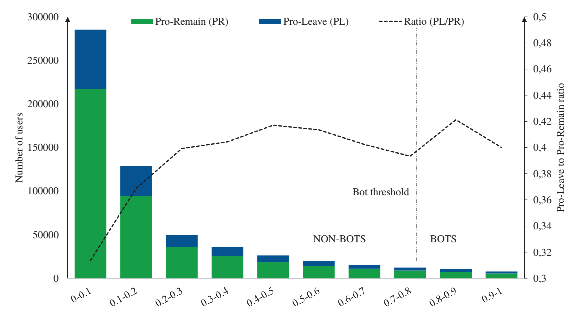

Figure 6: Bot score analysis of Twitter users regarding their user-level stance

Temporal stance analysis of Brexit related tweets Our temporal analysis show that Leave side tweets were higher only at the beginning of the time period before the referendum (Fig. 5).

Bot analysis of users combined with user stance Another aspect of our study is to combine the stance and bot behavior of a Twitter account. In our analysis, we found that there is a small correlation between an account's bot behavior and the stance expressed in the tweets of that account within our time period (Fig. 6).

Related Work

The UK's exit from the European Union, namely Brexit, has been intensively studied in the field of computational social science research. In one of the first studies on this topic, Llewellynand Cram presented data collection strategies and datasets (Llewellyn and Cram 2016). Hurlimann provided a dataset in order to define a gold standard for Brexit related tweets (Hürlimann et al. 2016). Chow et al.'s research is about detecting stance in Brexit referendum based on public survey data (Chow, Han, and Li 2019). Many other studies aim to detect the stance on Brexit from Twitter and other social media platforms (Khatua and Khatua 2017), (Celli et al. 2016), (Cesar Amador Diaz Lopez et al. 2017). Our dataset is also particularly focused on the online Brexit discussion, but covers a longer time period than other studies.

Detecting stance from text is an important challenge, for this reason, many shared tasks organized where researchers compete with each other on benchmark datasets. In the Semeval 2016 task, Mohammad et al. published 6 datasets for stance classification, and these datasets are mostly related to public policy topics (Mohammad et al. 2016). In the NLPCC-ICCPOL shared task, the participants aim to determine whether the stance of a Chinese microblogging writer is in favor or against of the given target (Xu et al. 2016).

Sentiment is a different piece of information to understand the attitude of the people. Sayadi et al. published a dataset paper containing the sentiment analysis on Twitter during the Tunisian presidential elections (Sayadi et al. 2016).

Social media users' opinions can be easily influenced by the activities of political bot accounts. Lewis recently published a dataset paper containing bot accounts on Twitter (Lewis 2019).

Conclusion

In this paper, we present our dataset related to Brexit Debate on the Twitter platform. We enriched the existing Twitter attributes with the features of Stance, Sentiment and Bot score. Considering the volume of the dataset and a variety of features, we believe researchers, social scientists and journalists will benefit from our dataset to discover new insights and have a deeper understanding of the Brexit phenomenon.
Acknowledgements. This work is partially funded by the European Commission under the H2020 framework, within the project 822735 TRIGGER (TRends in Global Governance and Europe's Role).

References

Calisir, E., and Brambilla, M. 2020. Twitter dataset about Brexit. *Harvard Dataverse* 10.7910/DVN/KP4XRP.

Celli, F.; Stepanov, E. A.; Poesio, M.; and Riccardi, G. 2016. Predicting brexit: Classifying agreement is better than sentiment and pollsters. In *PEOPLES@COLING*.

Cesar Amador Diaz Lopez, J.; Collignon-Delmar, S.; Benoit, K.; and Matsuo, A. 2017. Predicting the brexit vote by tracking and classifying public opinion using twitter data. *Statistics, Politics and Policy* 8.

Chow, W.; Han, E.; and Li, X. 2019. Brexit identities and british public opinion on china. *International Affairs*.

Davis, C. A.; Varol, O.; Ferrara, E.; Flammini, A.; and Menczer, F. 2016. Botornot: A system to evaluate social bots. *CoRR* abs/1602.00975.

Hürlimann, M.; Davis, B.; Cortis, K.; Freitas, A.; Handschuh, S.; and Fernández, S. 2016. A twitter sentiment gold standard for the brexit referendum. In *Proceedings of the 12th International Conference on Semantic Systems*, SEMANTiCS 2016, 193–196. New York, NY, USA: ACM.

Khatua, A., and Khatua, A. 2017. Leave or remain? deciphering brexit deliberations on twitter. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, volume 00, 428–433.

Lee Rainie, Aaron Smith, K. L. S. H. B., and Verba, S. 2012. Social media and political engagement. *PewInternet*.

Lewis, B. 2019. Harvard CGA Geo-tweet Archive - known bots. *Harvard Dataverse*.

Llewellyn, C., and Cram, L. 2016. Brexit? analyzing opinion on the uk-eu referendum within twitter. In *ICWSM*.

Mohammad, S.; Kiritchenko, S.; Sobhani, P.; Zhu, X.; and Cherry, C. 2016. Semeval-2016 task 6: Detecting stance in tweets. 31–41.

Sayadi, K.; Liwicki, M.; Ingold, R.; and Bui, M. 2016. Tunisian dialect and modern standard arabic dataset for sentiment analysis : Tunisian election context. In *To appear in the ACLing 2016 IEEE proceedings*. CICLING.

Xu, R.; Zhou, Y.; Wu, D.; Gui, L.; Du, J.; and Xue, Y. 2016. Overview of nlpcc shared task 4: Stance detection in chinese microblogs. volume 10102, 907–916.