In or Out?

Real-Time Monitoring of BREXIT sentiment on Twitter

Laurentiu Vasiliu Peracton Ltd. Dublin Ireland Iaurentiu.vasiliu@peracton.com

André Freitas, Frederico Caroli, Siegfried Handschuh University of Passau Germany {first.last}@uni-passau.de Ross McDermott, Manel Zarrouk, Manuela Hürlimann, Brian Davis, Tobias Daudert, Malek Ben Khaled, David Byrne Insight Centre for Data Analytics National University of Ireland Galway, Ireland {first.last}@insight-centre.org Sergio Fernández Redlink GmbH Salzburg Austria sergio.fernandez@redlink.co

> Angelo Cavallini 3rdPlace S.r.I. Milan Italy

angelo.cavallini@3rdplace.com

ABSTRACT

The SSIX (Social Sentiment analysis financial IndeXes) project is a European Innovation Project sponsored by the European Commission under the Horizon 2020 framework. SSIX aims to provide European SMEs with a collection of easy to interpret tools to analyse and understand social media sentiment for any given topic regardless of locale or language. The United Kingdom's recent referendum on European Union membership i.e. staying ("Bremain") or leaving the EU ("Brexit") was selected for the initial real-world test case for the validating the SSIX methodology and platform. In this paper, we describe the SSIX architecture in brief as well as analysis of the platforms X-Scores metrics and their application to Brexit, our initial experimental results and lessons learned.

CCS Concepts

• Computing methodologies→Artificial intelligence→Natural language processing→Information extraction.

• Computing methodologies→Machine learning→Learning paradigms→Supervised learning by classification.

Keywords

SSIX; Brexit; Natural Language Processing; Machine Learning; Opinion Mining; Twitter; Sentiment Analysis; Political Opinion Mining.

1. INTRODUCTION

The SSIX (Social Sentiment analysis financial IndeXes) project¹ is European Innovation Project sponsored by European Commission under the Horizon 2020 framework. SSIX aims to provide European SMEs with a collection of easy to interpret tools to analyse and understand social media users opinion for any given topic regardless of locale or language. The SSIX platform interprets significant sentiment signals in social media conversations producing sentiment metrics, such as sentiment dynamics, sentiment volatility and sentiment momentum.

© 2016 Copyright held by the author/owner(s). SEMANTICS 2016: Posters and Demos Track September 13-14, 2016, Leipzig, Germany

The recent United Kingdom European Union membership referendum on staying ("Bremain") or leaving the EU ("Brexit") was chosen as a first real-world test case for the SSIX consortium [1]. The goal was to stress test the SSIX platform and the methodology we have employed in order to infer opinion/sentiment from social networks. Furthermore, we employed the analysis of a set of rolling metrics called X-Scores, such as the raw aggregated sentiment, volumes, rolling averages and non-standard technical oscillators such as relative strength index (RSI) to examine their value for providing insights into sentiment behaviour. These initials tests enabled us to examine for the first time the SSIX platform in a real world scenario and provided extremely valuable feedback about both the behaviour of the technology we have employed for it and our fundamental assumptions on extracting sentiment data from social networks, which will be for various use cases, primarily for decisionmaking.

2. ASSUMPTIONS AND SSIX ARCHITECTURE

As originally foreseen, the SSIX project aims to cover the most important social networks such as Facebook, Twitter and LinkedIn. For the Brexit exercise, we started with Twitter only due to technical accessibility reasons. We note that Twitter users will not overlap exactly with the voting demographics in the UK but only a portion of it [2]. Moreover, it was not easy to identify what constitutes 'overlap' since many users do not disclose publicly their location of tweeting or residence.

However, we attempt to curtail this by, capturing English messages only. Overall, 40% of all activity can be said to come from geographical Europe (this includes GMT etc. time zones which cannot be attributed to a single country), while 18% comes from outside Europe. For 42% it was not possible to determine their location because the time zone is not set. Next, we present the location and percentage of sentiment expressed on those locations from Twitter users for some European² countries. This data represents only 33% (2.3 million Tweets out of a total of 5.9 million) from the entire data collection. Note, not all users enable their location data so it was not possible to capture this information fully.

² European here has the geographical meaning, EU and non-EU.

¹ http://ssix-project.eu/

European countries breakdown					
Country	# Tweets	%	Country	# Tweets 🛛 🖓	%
Albania	5	0.00%	Luxembourg	12	0.00%
Andorra	9	0.00%	Macedonia	1	0.00%
Austria	3,546	0.15%	Malta	3	0.00%
Belgium	28,353	1.22%	Monaco	1	0.00%
Bosnia and Herzegovina	211	0.01%	Netherlands	245,545	10.54%
Bulgaria	957	0.04%	Norway	18	0.00%
Croatia	1	0.00%	Poland	5,053	0.22%
Czech Republic	3,062	0.13%	Portugal	9,750	0.42%
Denmark	7,624	0.33%	Romania	3,094	0.13%
Estonia	475	0.02%	Serbia	8,967	0.39%
Finland	3,088	0.13%	Slovakia	652	0.03%
France	43,063	1.85%	Slovenia	14,528	0.62%
Germany	23,195	1.00%	Spain	32,364	1.39%
Gibraltar	2	0.00%	Sweden	8,603	0.37%
Greece	46,899	2.01%	Ireland	76,184	3.27%
Hungary	1,181	0.05%	Italy	34,195	1.47%
Lithuania	955	0.04%	Latvia	675	0.03%
Switzerland	12,874	0.55%	UK	1,713,732	73.59%
Total	2,328	,877		100%	

Table 1. Twitter European countries breakdown of SSIXBrexit Twitter collection.

SSIX Technical Set-up: The SSIX Platform is still currently under development; hence Brexit was the project's first validation scenario. The platform follows the Lambda Architecture principles³ where each component addresses a specific responsibility in the overall chain. The first component, based on 3rdEYE⁴, is responsible for gathering raw data from the different sources (e.g. Twitter), caching the data, performing basic filtering (e.g., spam detection), providing at the end APIs for both stream and batch processing further on in the platform. Then the data is moved to a Natural Languages Processing pipeline, where different analysis tasks (language identification, NLP analysis, etc.) are orchestrated towards providing sentiment classification in this scenario for each individual piece of content (tweet in this case). The data is then consolidated and stored, providing multiple custom metrics called X-Scores to interpret the data. The platform implementation makes extensive use of open source components: Apache Spark⁵ for computing, Apache Kafka⁶ for messaging, ElasticSearch⁷ for storing and Keras⁸ for Deep Learning, among many others.

Data were collected from Twitter using the official Streaming APIs⁹, for this environment using a dedicated server with 75 tracking keywords as well as hashtags and accounts. Incoming data were filtered based on a combination of different rules applied to the Twitter metadata (e.g. user language and number of followers). The raw data has been archived into a non-relational database for future reference¹⁰. From May 4th to June 30th, around 10.5M Tweets have been captured and stored, with an average of 175K Tweets per day.

³Lambda architecture is a data-processing architecture designed to handle massive quantities of data by taking advantage of both

batch and stream processing methods [3].

⁶ https://kafka.apache.org/

In order to train our opinion classification models (supervised classification), we manually annotated a set of Tweets. Our Gold Standard was obtained from Twitter data collected between May 4th and May 6th. A random sample of 2,000 Tweets was extracted from the filtered population for this time period. Three independent annotators, each of which had near native proficiency in English, classified the Tweets into one of five classes: "stay", "leave", "undecided", "don't care", "irrelevant". They assigned opinion strength scores between 1 (very weak) and 5 (very strong) to "stay" and "leave" Tweets. Furthermore, Tweets whose interpretation depends on an external source, such as an article or image, which are linked in the message, are flagged. A fourth annotator manually consolidated the annotations, creating the final gold standard for classification and evaluation.

For that purpose, we specifically trained a Deep Learning classifier¹¹, based on a manually annotated Twitter corpus, with an accuracy of 69%. Our classifier used a long short-term memory (LSTM) ¹²followed by a dense layer to classify the Tweets as either "stay" or "leave". Some tests were performed using all five annotated classes, but in the light of severe accuracy trade-off, we decided for the simpler, two-class classification model. This classifier was trained using the RmsProp algorithm¹³ and built using Keras.

3. BREXIT MONITORING AND ANALYSIS OF RESULTS

Our monitoring exercise had three phases: 1) **pre-voting period** (two weeks prior to the vote, then 2) **the voting day itself** and 3) the post-voting period¹⁴ (which will not be outlined in this paper). Each phase had a different monitoring pattern: for the pre and post Brexit phases, we observed the trends over 3-4 days period while during the voting day, we monitored the trends every 2 to 3 hours. Figure 1 shows the Twitter messages split between 'Remain' (blue) and 'Leave' (orange) while Figure 2 show the difference between the volume of 'Remain' (blue) Tweets and 'Leave' (orange) Tweets both a week prior to polling day (17/06 - 23/06).

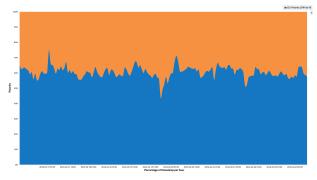


Figure 1. Share of vote week prior to voting day (17/06 - 23/06).

⁴ http://3rdplace.com/en/3rdeye/

⁵ https://spark.apache.org/

⁷ https://www.elastic.co/

⁸ http://keras.io/

⁹ https://dev.twitter.com/streaming/overview

¹⁰ https://bitbucket.org/ssix-project/brexit-gold-standard

¹¹ http://deeplearning4j.org/neuralnet-overview.html

¹² https://en.wikipedia.org/wiki/Long_short-term_memory

¹³ http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf

¹⁴ We are continuing to collect and analyse Tweets post the voting day. Further studies will be held around.

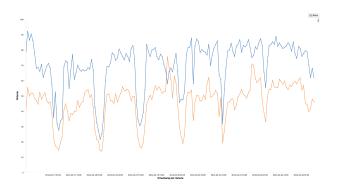


Figure 2. Remain (blue) and Leave (orange) Twitter volumes, week prior to polling day (17/06 - 23/06).

The results of the referendum were 48.1% 'Remain' and 51.9% 'Leave' [4]. The SSIX platform detected at all times a close outcome between 'Remain' and 'Leave', with some spikes on the 'Leave' side (in particular towards the closing hours of the vote). Still, the 'Remain' signal was consistently stronger, slightly above the border between 'Remain/Leave' axis, and had a slight but constant downward slope within the last 48 hours of monitoring. The volume of Tweets for the 'Remain' side was oscillating between 54.6% and 60.9% to end up/settle with 57.0% (10 pm (vote's closing hour) on the 23/06/2016) (see Figure 3).

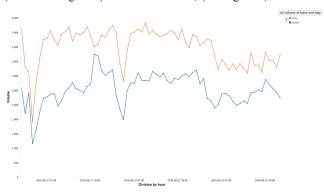


Figure 3. Leave and remain signals during the voting day and 48 hours before.

Table 2 below shows what the SSIX platform could and couldn't determine during the BREXIT monitoring exercise.

Table 2. What SSIX platform could determine and wh	at could
not to determine.	

What SSIX platform could determine	What SSIX platform could not determine	
The vote was a close call.	How close would be with the 'Remain' vs 'Leave'.	
X-Scores showed a low volatility that translated in very low movements between 'Remain' and 'Leave' side.	What was the exact overlap between UK voters and all tracked Twitter users by SSIX platform.	
A slow but constant downward trend for the 'Remain' side was detected before and during the 'Brexit' vote.	What correction factor to apply given the mismatch between UK voters and non-UK voters tweeting if any.	
The SSIX sentiment spiked in the 'Leave' area but never stabilised in this area, always returning to 'Remain'.	How many of the twitter users were for/against Brexit but didn't have the legal right to vote.	

Most of the identified Tweets come in order from UK, Netherland and Ireland.	What was the exact geographical distribution in UK
We identify ¹ / ₃ of the total volume of Tweets as coming from Europe, where UK voters but also non-UK voters were included	The geographic origin of the $\frac{2}{3}$ of the volume of Tweets, as there was no geo-location data coming from those twitter accounts
included.	How many non-UK voters tweeted in favour of 'Remain/Leave' out of the total number of collected Tweets

As mentioned in the previously, we correctly identify that the vote was going to be very close, but during the voting process and immediately afterwards we were unable to measure exactly 'how close' this was. While live monitoring of the Brexit event, according to the last day statements provided by the SSIX project [5] an ever-closing gap between the Leave and Remain was detected. This was a recurrent theme in most of our statements, due to the low levels of the sentiment strength that was trending towards zero value, having oscillations mainly between 0.1 and 0.3 within a sentiment interval of [-1;1].

Although the SSIX platform analysed that on the election day between 9am and 10pm an average of 57.5% of the people would vote "stay" - 48.1% voted stay - the results are reasonably near the target given the limitations we have faced (only one social network tracked - Twitter) and the unknowns parameters handled by the platform (not able to determine the source of 42% of total analysed Tweets). We observed the number of Twitter users supporting "Bremain" (71.3% on 20/06/2016: 5am) decreased within the three days before the referendum. Still, SSIX missed the results of the actual election by 9.4%. The question is why? There are several explanations we have identified in this respect:

Age Gap: The SSIX platform was only fed by Twitter data for the Brexit scenario which came to a huge dependence on the characteristics of the Twitter users, where even the majority of the Twitter users (35%) were in the age ranges of 15-24, the segment of users between 15 and 18 years old could not vote. Table 3 presents the age distribution of Twitter for February 2016.

Table 3. Age distribution ofTwitter users in Great Britain,February 2016 [6].

Age	% Twitter Users
15-24	35%
25-34	17%
35-44	20%
45-54	15%
55+	13%

Location Gap: Many of the Tweets collected originated from the London area. London has a large multicultural population, young and highly connected to internet and social media that voted strongly for 'Remain'. This may have introduced another bias into our final result. Also, there was a striking difference on how London voted 'Remain' compared to the rest of England population that voted 'Leave'. For example in London - Hammersmith and Fulham, 'Remain' side scored 70% while 'Leave' only 30%. In return, 'Doncaster' voted 69% 'Leave' and 31% 'Remain'.

Education Gap: according to [Guardian, 2016] there was a trend identified where UK residents with higher education voted consistently for 'Remain' while residents with no formal qualification voted consistently for 'Leave'. In the same time, according to [7] "Twitter is particularly popular among those

under 50 and the college-educated". Here this Twitter characteristic appears to have skewed the opinion/signal in accordance to the majority preference of its base users.

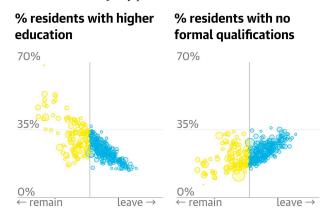


Figure 6. Higher Education voter's distribution for Brexit – Source: Guardian.

4. CONCLUSIONS AND FUTURE WORK

Our Brexit monitoring was a challenging proof-of-concept exercise for the SSIX platform. Monitoring was constrained (as explained in our introduction) by technical reasons allowing only the use of Twitter data, which generated an inevitably bias factor that we have explained in detail in Section Three.

An important lesson learned is that when analysing opinion from social networks, the demographics, geo-location, socio-economic factors have to be studied in detail with regard to the topic that is analysed. After that, a correction factor should be calculated to compensate for the errors and biases that are inevitably introduced. Corrections are even more important when there are close-call situations as it has been for Brexit. Being on the 'wrong' side of the expectation can generate consistent collateral effects, in particular for financial market applications, which is the main use case in SSIX.

In terms of improving further the accuracy of SSIX platform based upon of our proof-of-concept *Brexit* exercise, we envision adding new social networks data, as well as other web based data sources where an opinion/sentiment can be inferred by NLP techniques.

In order to avoid the bias induced by the gaps identified in the previous section, there are several actions to be taken:

- Location gap bias: Identify methods to determine the geolocation better (e.g. time zone) so a population targeted for analysis to be better pre-filtered.
- Age & Education gap bias: here a correction factor can be introduced, based on a preliminary analysis of the profile of the social network users and their benchmark against the general population.

In the same time, we consider that the difference between social media users and non-social media users will continue to reduce, so the overlap between voters and social media users will be more important. For electoral purposes, a study of this trend should be done as this would profoundly help to correct the current/future social sentiment derived signals and opinions.

Finally, our recommendation is that such opinion/sentiment analysis should never be used alone, but in conjunction with other

independent metrics and tools and so to produce a consolidated analysis from multiple inputs. Such an approach would reduce the margin of error and generate useful results for the studies in question.

In conclusion, while we did encounter a one-digit percentage error for SSIX Brexit opinion and Remain/Leave percentages, we can conclude that Twitter alone was not enough to monitor an event of such magnitude, as it was only partially covering the UK voters' characteristics. The 'Age Gap', 'Location Gap' and 'Education Gap' were three factors that deviated the opinion/sentiment analysis towards the 'Remain' side by the Twitter users, that are young, higher educated and urban. As we learned from the final result of the vote, the 'Leave' voters were on the other side of the gaps with Twitter voters, being older, having lower education and located in rural/small towns areas.

5. ACKNOWLEDGMENTS



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645425 and Science Foundation Ireland (SFI)

under Grant Number SFI/12/RC/2289.

6. DATASET

SSIX Brexit Gold Standard https://bitbucket.org/ssixproject/brexit-gold-standard

7. REFERENCES

- [1] Davis, Brian, Adamantios Koumpis, Ross McDermott, Keith Cortis, Laurentiu Vasiliu, and Siegfried Handschuh. "Social sentiment indices powered by X-scores." In 2nd International Conference on Big Data, Small Data, Linked Data and Open Data, ALLDATA 2016.
- [2] Sloan, L., Morgan, J., Burnap, P., & Williams, M. (2015). Who Tweets? Deriving the Demographic Characteristics of Age, Occupation and Social Class from Twitter User Meta-Data. PLoS ONE, 10(3), e0115545. http://doi.org/10.1371/journal.pone.0115545 Fröhlich, B. and Plate, J. 2000. The cubic mouse: a new device for threedimensional input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands, April 01 - 06, 2000). CHI '00. ACM, New York, NY, 526-531. DOI= http://doi.acm.org/10.1145/332040.332491
- Big Data, Principles and best practices of scalable realtime data systems. Nathan Marz and James Warren. April 2015 ISBN 9781617290343.
 https://www.manning.com/books/big-data Sannella, M. J. 1994. Constraint Satisfaction and Debugging for Interactive User Interfaces. Doctoral Thesis. UMI Order Number: UMI Order No. GAX95-09398., University of Washington.
- [4] EU Referendum Results http://www.bbc.com/news/politics/eu_referendum/results
- [5] 'Remain' is getting weaker http://ssix-project.eu/remain-isgetting-weaker/
- [6] Age distribution of Twitter users in Great Britain from May 2013 to February 2016.
- [7] Demographics on Key Social Networking Platforms http://www.pewinternet.org/2015/01/09/demographics-ofkey-social-networking-platform