Sentimental impact of world events on the Twitter social network

Nuno M. Mota Gonçalves
nuno.motagoncalves@epfl.ch

Matteo Y. Feo
matteo.feo@epfl.ch

Lucía Montero Sanchis
lucia.monterosanchis@epfl.ch

Abstract

The aim of this project is to quantify the *sentimental impact* a certain world event had on the Twitter community. Taking into account the emotional value associated with a specific event - found through *Language Recognition (LR)*, *Name Entity Recognition (NER)* and *Sentiment Analysis (SA)* - we are now able to define the *impactfullness* of conflicts from the Uppsala Conflict Data Program (UCDP) on a scale from 0.0 to 1.0. At this stage, country-wide events are the most viably described by our system, although with further developments (and time) it could be fine-tuned with more sources of data. The preliminary results seem to be pretty consistent and, further down the line, we can easily foresee the development of a *sentimental impact* prediction model for future events.

1 Introduction

For our project we initially decided to use both UCDP and Twitter-leon datasets. Because of time-constraints, however, we ended replacing the latter with our own Twitter dataset created using a Markov Chains based model. This model, trained on stories, geo-political and religious texts, basically generates Tweets on several different subjects.

From the combination of the UCDP and Twitter datasets we were now using, we wanted to understand how we could match certain Tweets to a given event and then evaluate the emotional value of the event itself - based on the retrieved information. By using *LR*, *NER* and *SA*, although limited to the English language, we manged to draw the desired correspondence. It's important to note we made the assumption that the content of English Tweets alone is a good representation of the world's overall opinion (which may introduce a Western Country preference to the analysis).

As to how we will make that information clear, we decided to showcase scenarios in which we pipeline our methodology and discuss the pros and cons of our approach to the problem - through statistics and plotting. In the end, every scenario will have its *sentimental impact* evaluated and compared to other scenarios.

To describe our work in more detail, we split our report into several sections: (2) Data Collection - how we got a hold of our data; (3) Data set description - where we both state which datasets we used for our project and describe our initial treatment of their data; (4) Methodology - showing how we made use of certain libraries and our datasets to achieve our goal; (5) Results and Findings - contains some considerations about certain parts of the project; and (6) Conclusions - our final remarks and future improvements on this idea.

2 Data Collection

Since we ended up not using the Twitter-leon dataset, all the data we used came from locally downloaded datasets. All of them are public and where acquired in either CSV or JSON format.

3 Dataset Description

3.1 UCDP Dataset

The UCDP dataset contains data for 133012 worldwide conflicts from 1989 until 2016. Other information available for conflicts relevant for our analysis is:

- Type of violence - Categorical variable that classifies conflicts into *State-Based*, *Non-State* and *One-Sided* violence.

- Location - Coordinates, region and country are available. We are only interested in the country.

- Start and end dates - They can be used to calculate the duration of the conflict.

- Number of casualties - Both the total number and the casualties on each side involved are available. We consider only the total number.

UCDP Dataset Filtering

We first assumed the more recent year's Twitter population should be more representative of nowadays reactions to the same events we are studying. Moreover, since the number of worldwide active users on Twitter has increased throughout the years (Arrojo, 2015), we ultimately decided to focus on conflicts that took place during 2016 - last year contained in the datasets. On top of that, it is also more likely that the public reacts to certain types of violence over others - and therefore we opted to analyze *one-sided violence* events. This type of violence represents 13.7% of the conflicts in 2016, as well as 10.7% of casualties in that year. Table 1 contains the statistics of these conflicts.

	Casualties	**Duration**	**N. per country**
mean	6.5	3 days	22.9
std	19.9	21 days	33.8
min	0	0 days	1
25%	1	0 days	2
50%	2	0 days	6
75%	5	0 days	25.5
max	324	334 days	106
count	893		

Table 1: Statistics for one-sided conflicts in 2016

As shown in the table, the number of casualties per conflict varies from 0 to 324. We assume that the public reaction will be higher for conflicts with casualties and, as a result, only conflicts with casualties are considered.

Table 1 also includes the number of conflicts per country, since we are assuming that the general public is likely to lose 'interest' (meaning they probably won't show drastic changes on their behavior) on the violent events that take place in a country where conflicts are very frequent. Therefore we also only consider the countries where there were, at most, three one-sided violent conflicts with casualties during 2016.

The statistics for the final filtered conflicts are shown in Table 2. The result is 30 conflicts from 18 different countries.

	Casualties	**Duration**	**N. per country**
mean	8	0 days	1.67
std	16.9	1 days	0.69
min	1	0 days	1
25%	1	0 days	1
50%	2	0 days	2
75%	5	0 days	2
max	86	5 days	3
count	30		

Table 2: Statistics for filtered conflicts

3.2 Other Datasets

More than identifying entities in a text using our NER model, we want to identify which countries they might be referring to specifically. To achieve this, we basically used several public datasets, to create bilateral word associations between: countries, cities, nationalities, religions, religious affiliations and currencies.

4 Methods

4.1 Language Recognition

For this stage we basically use an underlying library, langdetect, to filter out foreign tweets.

4.2 Named Entity Recognition

As stated before, our main goal is to find the country a certain Tweet is talking about. To this end we use the Spacy library and its NER models trained on OntoNotes 5 (giving us plenty of labels to work with). Even though the association from a country's name or nationality might be straightforward, religions and currencies tend to reference many different countries. As such, our NER module calculates the overall probability of each country in three steps:

1. For each identified identity, use the word association to find all the countries the word might be referencing and assign each a probability of $\frac{1}{N}$, where N represents the number of total countries associated.

2. From all the country-probability sets created in the previous step, we now generate a single set based on their union. For countries that exist for several entities, we add their probabilities.

3. Finally, we do a normalization of the probability set p:

$$p_i = \frac{p_i}{\sum_{j=1}^{|p|} p_j} \tag{1}$$

From these probabilities, we assume that the Tweet is talking about a certain country IFF:

$$p_i = \max_{1 \leq j \leq |p|} p_j \wedge p_i \geq threshold \tag{2}$$

where $threshold \in [0, 1[$.

4.3 Sentiment Analysis

For sentiment analysis we used *NLTK's Vader sentiment analyzer*. We then associate the composite value from its computation with the corresponding Tweet. This value will always be within $[-1, 1]$.

4.4 Daily Sentimental Strength

Before we analyze the impact of an event, we wanted to measure the *sentiment strength* on a daily basis, using all of the Tweets for that specific date. For this step, we want to penalize outliers. We mainly want to focus on the 'clustering' of the *sentiment strength* over that day so random, overly-emotional tweets should not affect our final measure. To do that, we define an *outlier* following (Iglewicz, 1993) For a given set of values d:

$$f_i = |d_i - \tilde{d}| \tag{3}$$

$$g_i = 0.6745 * \frac{f_i}{\tilde{f}} \tag{4}$$

$$h_i = \begin{cases} 1 & : g_i > threshold \\ 0 & : g_i \leq threshold \end{cases} \tag{5}$$

$$r = \frac{\sum_{i=1}^{|d|} d_i * h_i}{\sum_{j=1}^{|h|} h_j} \tag{6}$$

where $r \in [-1, 1]$ and represents the final result for our *daily sentiment strength*.

For instance, for the set of points shown in Figure 1 the daily average obtained is -0.483.

4.5 Sentimental Impact Measurement

On this stage, however, we do not want to penalize outliers. Since we are going to calculate *sentimental impact* over the *daily sentiment strengths*, we assume that big differences between successive values means they were triggered by an event. Instead, we actually want higher values to strongly

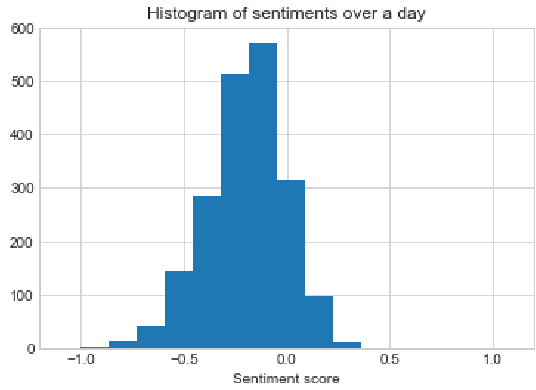

Figure 1: Skewed normal distribution

define the final result we get (which means there was a big impact). For a set of daily averages, evenly spread around a certain date of an event, we split them into b and a sets - each representing the values before and after the date of the event respectively. The exact day of the event is always considered to be the first day in the a set.

$$f = |\max_{1 \leq i \leq |b|} b_i - \min_{1 \leq i \leq |b|} b_i| \tag{7}$$

$$g = |\max_{1 \leq i \leq |a|} a_i - \min_{1 \leq i \leq |a|} a_i| \tag{8}$$

$$r = |g - f| \tag{9}$$

where $r \in [0, 1]$ and represents our final measurement for *sentimental impact*. We are aware that this measurement completely ignores sentimental 'aftershocks' (emotional fluctuations that might be visible after the occurrence of a given event) but this is a desired characteristic since we are comparing single day conflicts with multiple-day conflicts. The same behavior shifted within the time-frame (maybe a delayed response) will represent the same final value but conflicts that impact people several times (multiple-day) will not be considered more *impactfull* just because of their duration.

For the example in Figure 2, the impact measured is 0.735.

5 Results and Findings

Although the NER module is pretty accurate, we identified a lot of cases in which word capitalization had direct impact on whether or not it would be able to find a specific word. Although expectable, due to the training provided by OntoNotes 5, applying capitalization operations to

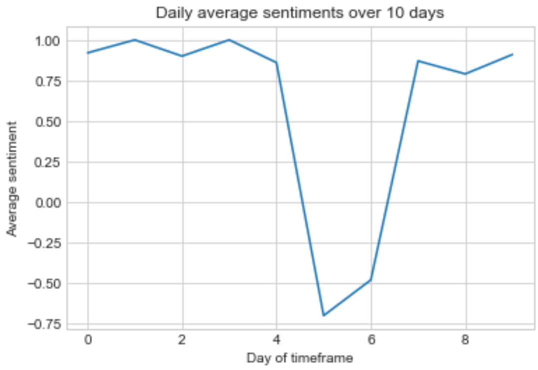

Figure 2: Daily average sentiments

the whole text (e.g uppercase, lowercase and titlecase) before passing it onto the module did not improve its accuracy in any type of entity. Capitalizing only nouns would improve our results but, since the model can't identify those nouns in the first place, we would have to rely on other sources in order to do it.

6 Conclusions

Looking at all our results, we ended up with a pipeline that would allow us to achieve our objective - to measure the *sentimental impact* of an event on a Tweet dataset. Although using an automatic Tweet generator (along with manually created Tweets) greatly helped us testing and improving our implementation, for the most part it represents synthetic data. To take this project to the next level, we would need to gather enough real-world data to fully prove our methodology. Moreover, we strongly believe that following this line of thought one could use events' features to train a predictive model based on its *impactfullness*.

References

Maria Jose Arrojo. 2015. *International Journal of Social Science Studies*, volume 3. Prentice-Hall, Englewood Cliffs, NJ.

Boris Iglewicz and David Hoaglin. 1993. *The ASQC Basic References in Quality Control: Statistical Techniques*, volume 16. Edward F. Mykytka, Ph.D., Editor.