Report on Twitter topic detection and analysis

Alex Tsilingiris
Christina Pardalidou
Sotiris Karapostolakis

June 28, 2017

Contents

1 Introduction

Text

2 Model description

Section

List me TF sta hashtags? Preprocessing steps?

3 Method implementation details

Event detection

Our approach

A combination of approaches was used for event detection: Since the dataset was static and filtered based on a query (in our case the word ”trump”), terms were considered as living organisms in the given timespan, allowing us to sample a set of tweets as when each term was ”most alive”, while combining user reputation metrics to select a tweet from a reputable source.

Hashtags as terms

We solely operated on hashtags on this step since they represent an idea or a topic, which in other cases would be difficult, or not as accurate, to define using Natural Language Processing and Machine Learning.

The living organism implementation

We considered a term as most-alive at the point in which it was mostly detected in our dataset. The time window was set to 2 hours. Therefore, we had to query for the mostly found terms (using the Term Frequency statistic) and sort the results in descending order. We then generated the final that contained the original tweets with the most active terms in the dataset that were posted in the given timespan.

Selecting a tweet for the pool

We now have a pool of tweets that might contain information about an event. The approach we followed was to assume that a user with a high amount of followers represents an influential event source into a social community[1]. We sorted (descending) the tweet pool by the number of followers the original poster has and ended up with 1 tweet that was our result for each term.

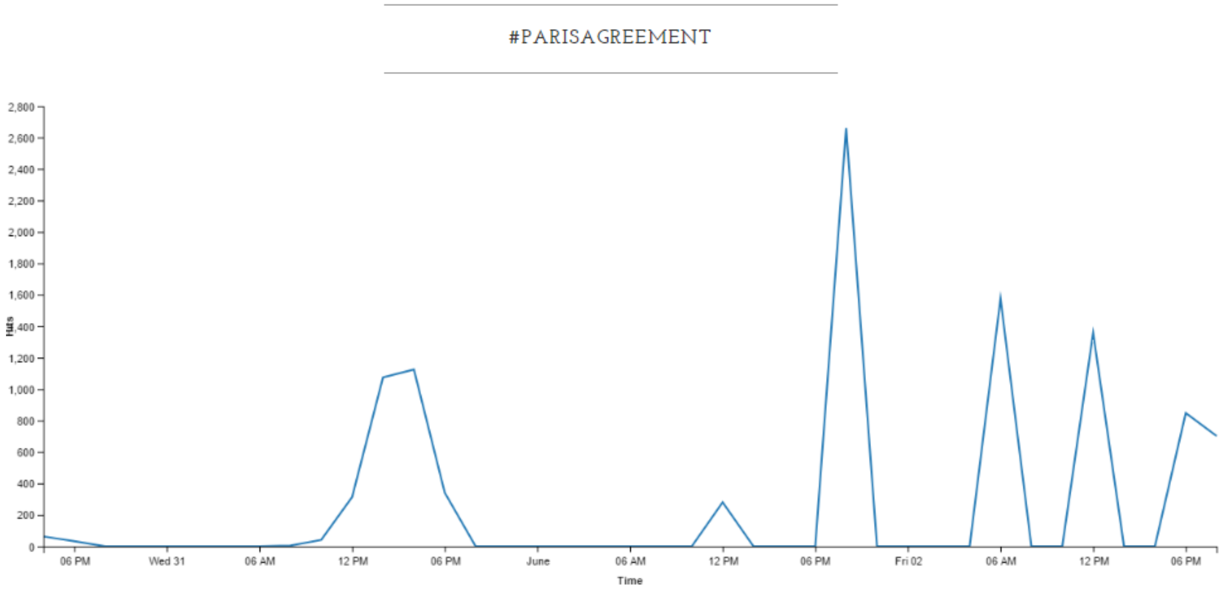
Event detection

Some top ranked terms (hashtags), their occurrences and the tweet that qualified as an event by our system will be presented in table 4.1

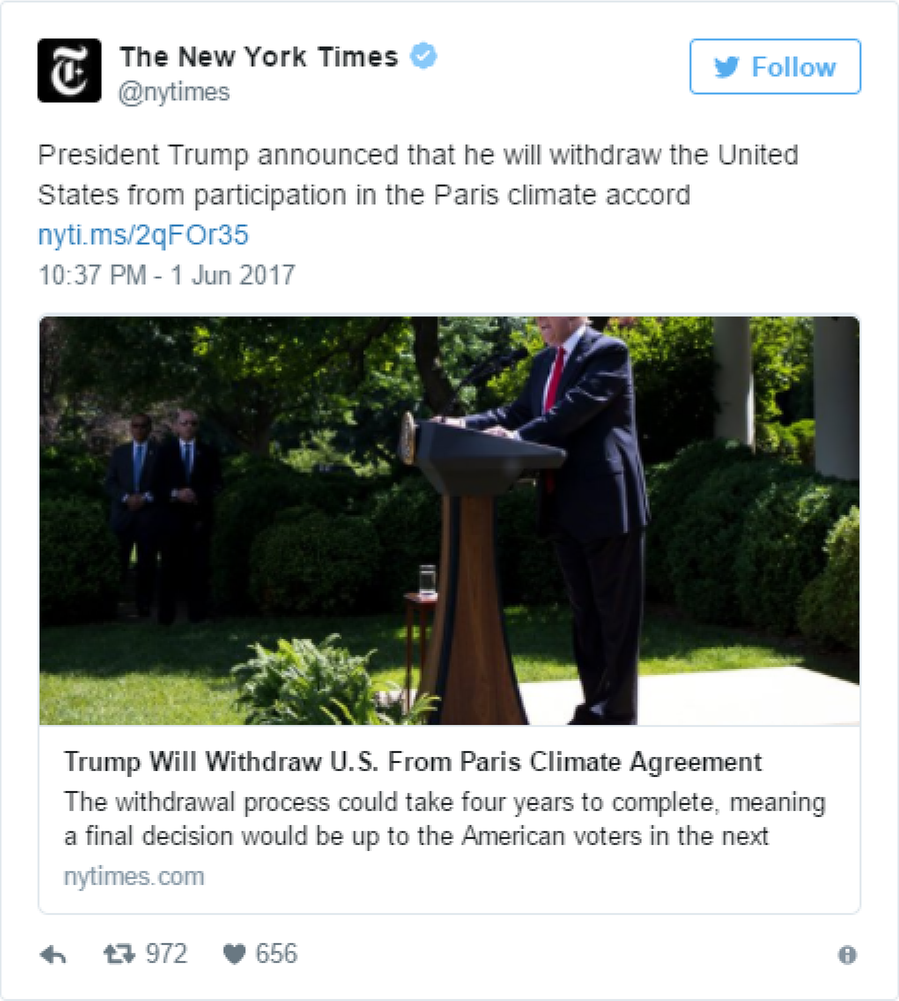
Term	Frequency
#parisagreement	10442
#covfefe	7085
#trumprussia	4229
#marchfortruth	973

Table 4.1: Terms and occurrences

The top ranked term, a graph showing its occurrence over time can be seen in the following figure:



The tweet that our system determined as an event can be seen in the following figure:



The New York Times
@nytimes

President Trump announced that he will withdraw the United States from participation in the Paris climate accord nyti.ms/2qFOr35
10:37 PM - 1 Jun 2017

Trump Will Withdraw U.S. From Paris Climate Agreement
The withdrawal process could take four years to complete, meaning a final decision would be up to the American voters in the next
nytimes.com

Additional results can be found in the project's website [2].

5 Software implementation details

Section

Test

Bibliography

[1] M. Cataldi, L. D. Caro, and C. Schifanella. Personalized emerging topic detection based on a term aging model. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(1):7, 2013.

[2] Website. Project website. `http://snf-755277.vm.okeanos.grnet.gr/`.