IBEREVAL 2017:

Task: Classification Of Spanish Election Tweets (COSET)

Maite Giménez [1]

Paolo Rosso [1] Roberto Paredes [1]

Tomás Baviera [2] Dafne Calvo [2] Germán Llorca [2] José Gámir Ríos [2]

coset2017@gmail.com

March 8, 2017

[1]PRHLT Research Center
Universitat Politècnica de València

[2]Mediaflows Research Group

Table of contents

Task Description

Motivation

- Political conversation in Twitter are common, and it increases when a General Election comes close.
- Analyzing the topics discussed by users provides interesting insights of this growing public conversation on politics.
- Nowadays, political campaigns are monitoring this in semi-automatic fashion. The focus of this task is to be able to at least classify tweets depending on the political topic discussed.

Task description

Therefore, we face a classification task. In COSET, the aim is to classify a corpus of political tweets following these five categories:

1. **Political issues** Related to the most abstract electoral confrontation.
2. **Policy issues** Tweets about sectorial policies.
9. **Campaign issues** Related with the evolution of the campaign.
10. **Personal issues** The topic is the personal life and activities of the candidates.
11. **Other issues.**

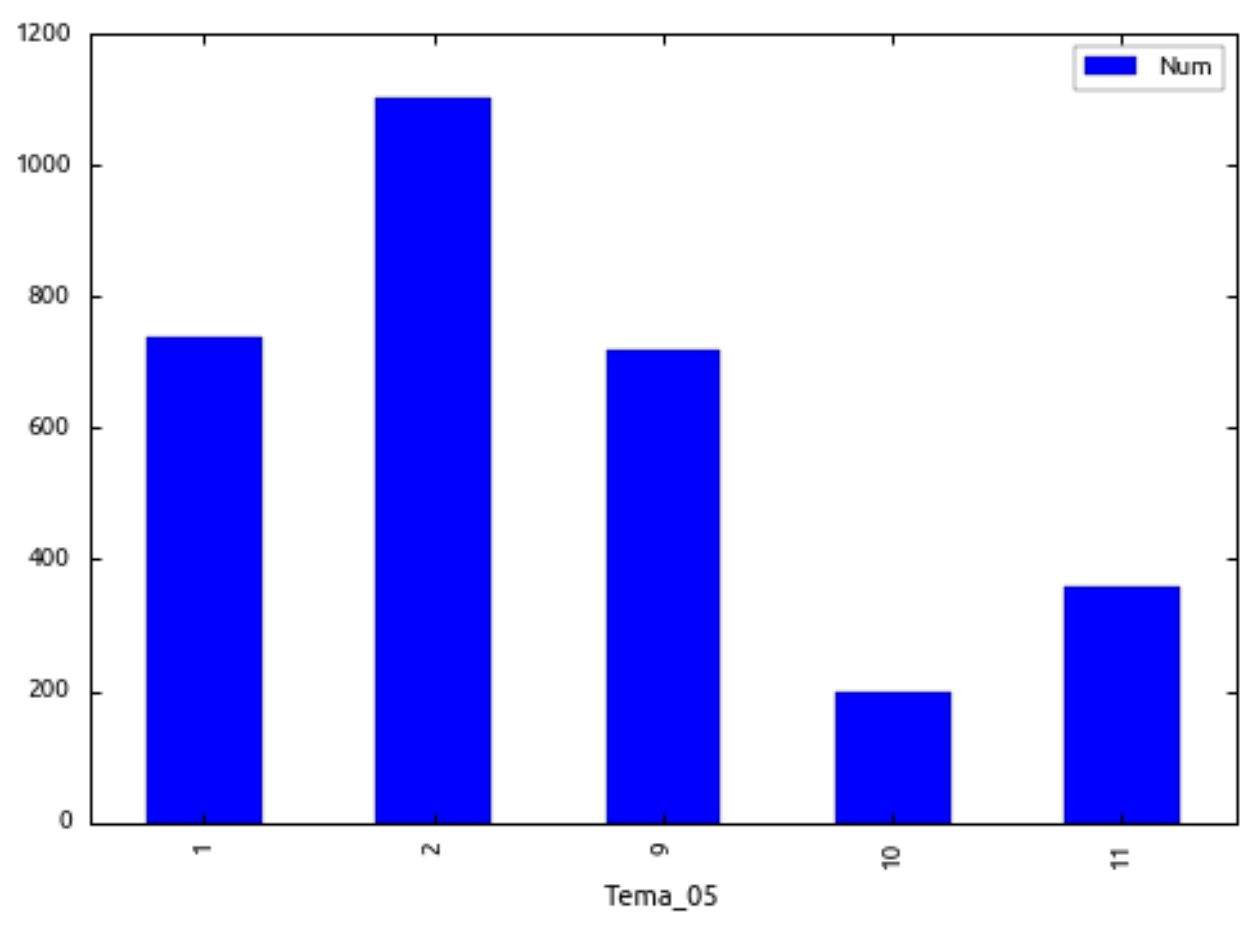
Given a tweet the system should be able to automatically predict its topic (an integer value within the set [1, 2, 9, 10, 11]).

Task description II

- The tweets are written in **Spanish**
- The tweets were gathered during the **2015 Spanish General Election**.

Corpus

Corpus insights: Topic Distribution



Corpus: Train, Dev and, Test partitions

Table 1: Dataset partitions

Partition	# Tweets
Train	2242
Dev	250
Test	624

Corpus distribution

The corpus will be distributed password-protected following one of these two formats.

Plain file

A plain file with the following structure:

`<tweet_id> \t <text> \t <topic_label>`

CSV

A unix-based csv file with the same structure

`<tweet_id>,<text>,<topic_label>`

Potential Problems: Excel might truncate the `<tweet_id>`

Evaluation

Submission details

- Participants should provide a plain text file with this structure: `<tweet_id> \t <topic_label>`
- The running code developed in order to infer the topics. The organization will verify the reproducibility of the results of the systems.

Evaluation metrics

$$F_{1-macro} = \frac{1}{|L|} \sum_{l \in L} F_1(y_l, \hat{y}_l) \tag{1}$$

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{2}$$

$$precision = \frac{1}{|L|} \sum_{l \in L} Pr(y_l, \hat{y}_l) \tag{3}$$

$$recall = \frac{1}{|L|} \sum_{l \in L} R(y_l, \hat{y}_l) \tag{4}$$

Important Dates

Important Dates

March 20th, 2017 Release of training data.

April 24th, 2017 Release of test data.

May 08th, 2017 Submission of runs.

May 15th, 2017 Evaluation results.

May 29th, 2017 Working notes due.

June 12nd, 2017 Review to authors.

June 26th, 2017 Camera ready due.

Contact

Contact

Mail: coset2017@gmail.com

Mailing list: coset@googlegroups.com

Join the shared task!

Questions?