

Abstract

1 Introduction

2 Related Work

3 Dataset

The dataset consists of a collection of tweets from the time period May 2012 to December 2014. We focus on the political organizations of Latin America and identify a set of 63 organizations that include individual politicians and political groups, for example, anarchists. The politicians are mainly from countries like Venezuela, Columbia, Mexico, and so on. The tweets were collected using the twitter API, based on these 63 organizations of interest. The tweets included :

- Tweets by the organizations
- Tweets that mention these organizations

The data has the following information :

1. Time of the tweet
2. Retweet or not ?
3. User details : Location, counts of tweets posted, followers, friends, klout score
4. Mentions : list of screennames and corresponding ids of users mentioned in the tweet (When a tweet by a user, includes the username of another user, it is called a 'mention'.)
5. Retweet count : Number of times the tweet has been retweeted
6. GeoLocation Enrichment
 - Latitude, longitude
 - Location/country
7. Basis Enrichment
 - The tweets are tokenized and tagged with Parts-of-Speech tags
 - Entities identified if any, i.e., if an expression in the tweet is an organization or an individual or a URL.
 - Noun-phrases identified.

8. Sentiment score : computed by a third party¹. These scores range from -24 to 24 and are mapped to values between [0,1], as detailed in the experiments section.

Approximately 6.9 million tweets were collected, that had about 10,400 tweets by the organizations, and the rest being the tweets mentioning these organizations.

4 methods

5 Models

6 Empirical Evaluation

7 Conclusion

¹<http://datasift.com>