Twitter Trending Topic Classification

MES: Maria Ludovica Costagliola - Emanuele De Santis - Serena Ferracci

Abstract— The Web Information Retrieval course allowed us to develope the described project. It is about classification of topics that are trending of twitter in a given moment. We have dealt with (i) text based classification using single tweets (ii) network based classification exploiting the graph structure of Twitter.
The used dataset is composed by tweets divided by trending topics. We classify the trending topics into 8 categories: *Event, Health, Movie, Music, Politics, Science, Society* and *Sport*.
The first approach gives good result in terms of precision and recall. Instead, the second approach does not give conclusive results because it require too many resources to be implemented on a single machine.

1 Introduction

We chose this project because Twitter is one of the most popular social network. It is used every day by millions of users expressing their opinions about several fields. When an user looks for something, the first thing Twitter displays to him is the list of trending topics of the moment. Often the user can not know what the topic is about, so it as to manually search for tweets belonging to that trending topic to better understand it.

It is interesting for the user to have a way to know what the Trending Topic is about without further searches. Our work tries to replicate the results obtained by [1]. The general categories used in this project are 8, named: *Event, Health, Movie, Music, Politics, Science, Society* and *Sport*.

2 Related Work

The paper we refer to is the work done by Lee et al. for tweets classification [1]. They have used 18 general categories to classify each trending topic. They address the problem following two different approaches

References

[1] K. Lee, D. Palsetia, R. Narayanan, M. M. A. Patwary, A. Agrawal, and A. Choudhary. Twitter trending topic classification. *2011 IEEE 11th International Conference on Data Mining Workshops*, 2011.