Massive Data Analytics' Project Proposal

Sandro Cavallari, Marco Giglio, Paolo Morettin

March 22, 2014

1 Problem description

Social networks experienced an exponential growth in the last 5 to ten years. Several people, nowadays, tend to spend many hours per day writing on their *walls*, *twitting* etc, and social networks are often be proved to be modern news aggregator.

Given that, it is important to monitor the most popular trends on social networks and to understand whether people are feeling positive or negative toward a certain topic. Our project aims in developing a methodology in order to:

1. understand when the common feeling about a certain topic shifted from positive to negative;
2. correlate this shift to news coming from newspapers and news agencies.

2 Work plan

In order to perform our analysis we need a big dataset containing tweets and another one containing the news published by popular newspapers and news agencies during the same temporal interval. Our team is planning in renting a server having low computational power but high bandwidth and availability. The server will be responsible for collecting both tweets and news 24 hours per day for some weeks. All the information collected by the server will be stored on a database.

In a second step, all the information will be downloaded on faster machines and analysed in order to discover trends and correlations.

3 Dataset

As over-mentioned the Dataset for this work will be composed by:

- Twitt downloaded by a server using twitter4j and saved on a MongoDb
- News collected by the Rss Feed of the most important newspaper websites

Our plan is to download at least 1 month of data and only after start to analyse the correlations.

References

[1] Text