Research Report Elias

Maria-Hendrike Peetz

May 23, 2013

1 Introduction

The intention of my visit of UNED and Llorente & Cuenca was to understand what *reputation polarity* is. This important to ensure that we are using the right measures and methodology in ongoing online reputation benchmarks. The research questions I am asking with respect to this question are:

RQ1 What are the global measures used to annotate the reputation of a company and how can they be translated into computational terms?

RQ2 What are the measures on a tweet level used to annotate the reputation of a company and how can they be translated into computational terms?

RQ3 What are the procedures of brand analysts in the annotation of reputation polarity of tweets?

In the following I lay out how the visit helped me in answering the three questions.

2 Reputation Measures

Social media analysts measure reputation intuitively every day. In order to understand their measures, I filmed and interviews analysts at Llorente y Cuenca. In particular, I filmed four people in their daily monitoring task. People were asked to explain every step they do and every annotation out loud. Afterwards I interviewed the people, asking for different measures between sectors. The experts were encouraged to answer freely. In total we collected about 4h of video material. The results still need to be evaluated, but they should help to answer **RQ2** and **RQ2**.

Furthermore, with my host Julio Gonzalo, we further developed a questionnaire for employees at Ll& C that will allow us to understand what their measures for reputation polarity are. The questionnaire is currently in the process of being filled out.

So far, we found that analysts want help with tedious annotation tasks, such as annotations for relevance to an entity. They also would like to have assistance. On a tweet level, the analysts' most important measure seems to be the number of followers the author of the tweet has and the number of retweets.

3   Annotation Procedures

In order to answer **RQ3**, I improved the a collection of log data of the annotation software of Ll& C. This annotation software allows analysts to create company profiles and compare those profiles with competitors. In order to do so, documents from different data sources (Twitter, Google, Facebook, Youtube) are being annotated for their relevance to the company, the reputation polarity (i.e.: what is the impact on the reputation of the company), and the topical dimension.

I was provided with a recent database dump of the annotations themselves and log data. I learnt about the process model that the developers have for the users. We identified two important future research questions in the field:

RQ4 The number of documents that are relevant for an entity is the number of Google hitsfor the company name as query. Obviously, this is very prone to errors (i.e.: Apple). How, with maximally 100 documents being annotated for relevance, can we project this knowledge onto all retrieved documents?

RQ5 How can we include active learning to rank into the assessment of relevance and reputation polarity of documents?

An answer to **RQ4** will help finding computational models for global measures, as asked for in **RQ1**. Findings of processes as asked for in **RQ3** actually lead to the new research question **RQ5**, as to how we can improve the current process.

4 Further collaboration

Additional collaboration with the hosts at UNED has been done on the participation in RepLab 2013. We intent to submit an active learning based system.

5 Future collaboration

Future collaboration will involve writing a journal paper on *What is Reputation* and is estimated to be finished by the end of the year.

6 Conclusion

The results of my visit are data sources that can be used to define reputation polarity. Future work and ongoing collaboration is required for documentation and evaluation of the data sources.

First preliminary insights gained from the visit have been included in a recent CIKM submission.