ORMA: A Semi-Automatic Tool for Online Reputation Monitoring in Twitter

Jorge Carrillo-de-Albornoz, Enrique Amigó, Damiano Spina, and Julio Gonzalo

UNED NLP & IR Group
Madrid, Spain
http://nlp.uned.es
{jcalbornoz,enrique,damiano,julio}@lsi.uned.es

Abstract. We present a semi-automatic tool that assists experts in their daily work of monitoring the reputation of entities—companies, organizations or public figures—in Twitter. The tool automatically annotates tweets for relevance (Is the tweet about the entity?), reputational polarity (Does the tweet convey positive or negative implications for the reputation of the entity?), groups tweets in topics and display topics in decreasing order of relevance from a reputational perspective. The interface helps the user to understand the content being analyzed and also to produce a manually annotated version of the data starting from the output of the automatic annotation processes. A demo is available at: http://nlp.uned.es/orma/.

Key words: Online Reputation Monitoring, Social Media, Twitter

1 The Task of Online Reputation Monitoring in Twitter

The rise of social media brought serious concerns to companies, organizations, and public figures on what is said about them online; one of the main reasons is that negative mentions may affect businesses or careers. Monitoring online reputation has therefore become a necessity. An online reputation analyst typically has to perform at least the following tasks: given a stream of texts containing a potential mention to a company as input, (i) *filtering* out tweets that are not related to the entity of interest, (ii) determining the *polarity* (positive, neutral or negative) of the related tweets, (iii) clustering the strongly related tweets in *topics*, and (iv) assigning a relative *priority* to the clusters, in terms of whether a topic may damage the reputation of the entity.

Figure 1 describes the main steps carried out during the annotation process for Online Reputation Monitoring. The process starts selecting one of the entities assigned to the expert. In the system, each entity has a list of tweets that the expert has to annotate manually. The expert processes tweets sequentially: first, she decides whether the tweet does refer to the entity of interest or not. If the tweet is unrelated to the entity, the annotation process for the tweet finishes and the expert continues with the next tweet in the list. Otherwise, the polarity and topic annotations follow. Polarity annotation consists in deciding whether the tweet may affect positively or negatively to the reputation of the entity.

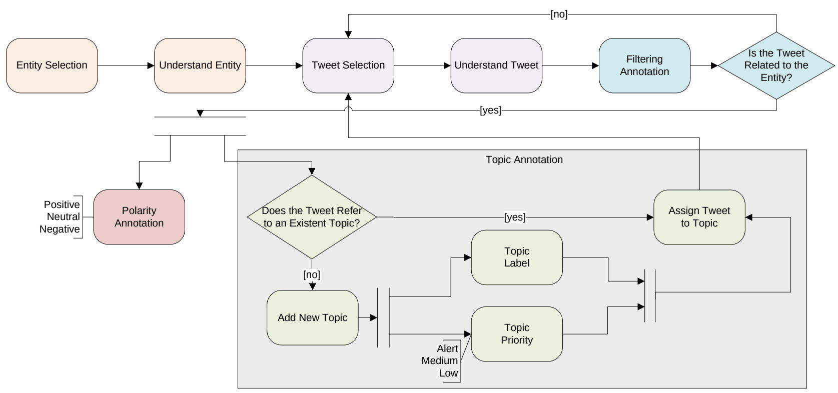

Fig. 1: Workflow of the Online Reputation Monitoring annotation process.

Topic annotation consists of identifying the aspects and events related to the entity that the tweet refers to. If the tweet refers to an already identified topic, the tweet is assigned to it. Otherwise, the expert defines a new topic. A topic receives a label that summarizes what the topic is about, and it is also classified in a priority scale (Alert, Medium or Low in our tool). When the tweet is assigned to a topic, the annotation of the current tweet is finished.

2 Architecture and Implementation

Our *Online Reputation Monitoring Assistant (ORMA)* aims to assist the daily work of reputation experts, helping them to manually process the data more efficiently.

Figure 2 shows the architecture of the ORMA demo. The system is deployed into two independent elements: the *Web Client* and the *Server*. The user interface permits the user to manually label tweets about an entity of interest, following the annotation process described above. To reduce the effort of experts, the system also proposes different automatic labels for each input data, with a confidence score indicating how trustable these labels are. The input data of the demo are tweets about different entities in both English and Spanish.

The server element processes the tweets of a given entity and proposes labels for the four subtasks. The tweets are processed in the *Tweet Labeler* using the algorithms described in [3,5,4,2]. The Tweet Labeler is divided in four components, which address each of the reputation monitoring subtasks. Each component of the Tweet Labeler can be implemented using one or multiple algorithms. The labeled tweets are then analyzed with the *Confidence Score Estimator*. For each subtask, this module analyzes the output of the different algorithms and determines a confidence score representing the degree of certainty of the system for the proposed labels. This score should be considered by the experts as a threshold that determines which tweets can be labeled automatically and which tweets need to be revised. The output of this element is stored in the database that is accessed by the user interface.

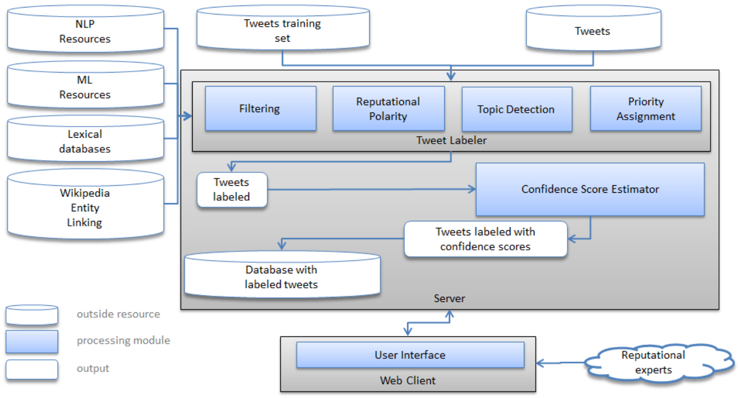

Fig. 2: Online Reputation Monitoring Assistant architecture.

The *Database* stores the tweets and the proposed labels for them (both the ones assigned by the system and manually by the user) and has two salient features: first, the tweets in the database are organized at the entity level, i.e., only the tweets associated to the entity being analyzed are loaded, which allows the database to be easily deployed in a distributed environment. Second, tweets are retrieved in a lazy manner, i.e., only tweets that need to be displayed are retrieved, which makes the system more efficient and scalable.

3 Interaction Design

Once the expert is logged in, its user context is retrieved from the database, which includes the different tasks that have been assigned, the already labeled information and the work status. The system then displays the interface for labeling content.

Once the expert has selected an entity to study, the tool shows the tweets to be analyzed in the *Tweets* table. Each row in this table is a tweet containing the name of the author (which is a link to the user profile too), the text of the tweet, the date when the tweet was published and a link to the tweet. Each tweet has also associated an icon identifying if the tweet has already received some manual annotation (reputational expert icon), if all annotations are automatic (computer icon), or a question mark icon if the tweet has not been labeled yet. On the right hand side of the Tweets table, the tool shows two combo boxes (*Related* and *Tweet Polarity*) that allow the reputational experts to annotate the selected tweet with information about its relation to the entity and its polarity.

The *Topic Management* panel groups the tweets by topics and allows the experts to work at the cluster level grouping tweets strongly related in meaning. The two buttons at the bottom of the Topic Management panel, *Add* and *Edit*, are provided to create new topics or to modify the existing ones. When annotating a tweet, the expert can create a new topic and assign the tweet to it, or may assign the tweet to one of the previously created topics. To this aim, the *Tweets in Topic* table shows all tweets previously assigned to the selected topic in the Topic Management panel. Once the topic is selected, the user must assign the tweet to the topic using the *Assign Topic* button.

The *Automatic Mode* panel situated above the Tweets table allows the expert to automatically label the tweets associated to the entity being analyzed. To this aim, the interface presents four buttons, one for each of the subtasks. The panel also contains a confidence score slide bar that allows the user to determine the desired confidence threshold of the automatic labels. Pressing any of the buttons, the system loads all automatic annotations for the appropriate subtask which have a confidence score above the specified threshold —except for those tweets which have already been manually annotated—. These automatic annotations can be manually changed by the reputational expert.

4 Testing and Prototyping

An earlier version of the ORMA annotation application (which did not include the option to automatically process the data) has been tested by thirteen experts along the preparation of the RepLab 2013 test collection [1]. Over half a million annotations for 61 different entities were performed for a total workload of 21 person month. During the exercise, the application was extensively tested for robustness and user friendliness. In particular, interaction design was significantly enhanced by many GUI changes suggested by the annotators.

Acknowledgments. This research was partially supported by the European Community's FP7 Programme under grant agreement n 288024 (LiMoSINe), the Spanish Ministry of Education (FPU grant nr AP2009-0507), 8), the Spanish Ministry of Science and Innovation (Holopedia Project, TIN2010-21128-C02) and the Regional Government of Madrid and the ESF under MA2VICMR (S2009/TIC-1542).

References

1. E. Amigó, J. Carrillo-de-Albornoz, I. Chugur, A. Corujo, J. Gonzalo, T. Martín, E. Meij, M. Rijke, and D. Spina. Overview of RepLab 2013: Evaluating online reputation monitoring systems. In *CLEF '13*, LNCS, pages 333–352. Springer, 2013.
2. H. Anaya-Sánchez, A. Peñas, and B. Cabaleiro. UNED-READERS: Filtering Relevant Tweets using Probabilistic Signature Models. In *CLEF 2013 Labs and Workshops Notebook Papers*, 2013.
3. Á. Castellanos, J. Cigarrán, and A. García-Serrano. Modelling techniques for twitter contents: A step beyond classification based approaches. In *CLEF 2013 Labs and Workshops Notebook Papers*, 2013.
4. M.-H. Peetz, D. Spina, J. Gonzalo, and M. de Rijke. Towards an Active Learning System for Company Name Disambiguation in Microblog Streams. In *CLEF 2013 Labs and Workshops Notebook Papers*, 2013.
5. D. Spina, J. Carrillo-de-Albornoz, T. Martín, E. Amigó, J. Gonzalo, and F. Giner. UNED Online Reputation Monitoring Team at RepLab 2013. In *CLEF 2013 Labs and Workshops Notebook Papers*, 2013.