

# Concept Term Expansion Approach for Monitoring Reputation of Companies on Twitter

M. Atif Qureshi<sup>1,2</sup>, Colm O’Riordan<sup>1</sup>, and Gabriella Pasi<sup>2</sup>

<sup>1</sup> Computational Intelligence Research Group, National University of Ireland Galway, Ireland

<sup>2</sup> Information Retrieval Lab, Informatics, Systems and Communication, University of Milan Bicocca, Milan, Italy

`muhammad.qureshi@nuigalway.ie`, `colm.oriordan@nuigalway.ie`,  
`pasi@disco.unimib.it`

**Abstract.** The aim of this contribution is to easily monitor the reputation of a company in the Twittersphere. We propose a strategy that organizes a stream of tweets into different clusters based on the tweets topics. Furthermore, the obtained clusters are prioritized into different priority levels. A cluster with high priority represents a topic which may affect the reputation of a company, and that consequently deserves immediate attention. The evaluation results show that our method is competitive even though the method does not make use of any external knowledge resource.

## 1 Introduction

Twitter<sup>3</sup> has become an immensely popular microblogging platform with over 140M unique visitors, and around 340M tweets per day<sup>4</sup>. Based on its growing popularity, several companies have started to use Twitter as a medium for electronic word-of-mouth marketing [3] [4]. There is also an increasing trend of Twitter users to express their opinions about various companies and their products via tweets. Hence, tweets serve as a significant repository for a company to monitor its online reputation; this motivates the need to take the necessary steps to tackle threats to it. This, however, involves considerable research challenges and motivates the research reported in our paper, the main characteristics of which are:

- clustering tweets based on their topics: for example, the company Apple would have separate topical clusters for iPhone, iPad and iPod etc.
- ordering tweets by priority to the company: the idea is that tweets critical to the company’s reputation require an immediate action, and they have a higher priority than tweets that do not require immediate attention. For example, a tweet heavily criticizing a company’s customer service may damage the company’s reputation, and thus they should have high priority.

---

<sup>3</sup> <http://twitter.com>

<sup>4</sup> <http://blog.twitter.com/2012/03/twitter-turns-six.html>

In this paper, we focus on the task of monitoring tweets for a company’s reputation, in the context of the RepLab2012, where we are given a set of companies and for each company a set of tweets, which contain different topics pertaining to the company with different levels of priority. Performing such a monitoring of tweets is a significantly challenging task as tweet messages are very short (140 characters) and noisy. We alleviate these problems through the idea of concept term expansion in tweets. We perform two phases of clustering and priority level assessment separately, whereby the clustering employs unsupervised techniques while supervision is used for priority level assessment.

The rest of the paper is organized as follows. Section 2 describes the problem in more detail. Section 3 presents our technique for clustering and assigning priority levels to the clusters. Section 4 describes the experiments and finally Section 5 concludes the paper.

## 2 Problem Description

In this section, we briefly define the problem statement related to this contribution. We were provided with a stream of tweets for different companies collected by issuing a query corresponding to the company name. The stream of tweets for the companies were then divided into a training set and a test set. In the training set each stream of tweets for a company was clustered according to their topics. Furthermore, these clusters were prioritized into five different levels as follows:

Alert > average priority > low priority > ‘other’ cluster > ‘irrelevant’

Alert corresponds to the cluster of tweets that deserves immediate attention by the company. Likewise, the tweet clusters with average and low priority deserve attention as well but relatively less attention than those with alert priority level. In the case of ‘other’, these are clusters of tweets that are about the company but that do not qualify as interesting topics and are negligible to the monitoring purposes. Finally, ‘irrelevant’ are the cluster of tweets that do not refer to the company.

Our task is to cluster the stream of unseen tweets (test set) of a given company with respect to topics. Furthermore, we have to assign these clusters a priority level chosen from the above-mentioned five levels.

## 3 Methodology

The proposed method is entirely based on the tweets contents, i.e., it does not use any external knowledge resource such as Wikipedia or the content of any Web page. Before applying our method we expand the shortened URL mentioned inside a tweet into a full URL in order to avoid redundant link forwarders to the same URL. Furthermore, a tweet that is not written in English is translated into the English Language by using the Bing Translation API<sup>5</sup>. In the following subsections we present the proposed strategy to analyse tweets.

---

<sup>5</sup> <http://www.microsofttranslator.com/>

### 3.1 Tweet concept terms extraction

In the first step, we extract concept terms (i.e., important terms) from each tweet so as to be able to identify a topic. To achieve this goal, we filter out the trivial components from each tweet's content such as mentions, RT, MT and URL string. Then, we apply POS tagging [5] to identify the terms having a label 'NN', 'NNS', 'NNP', 'NNPS', 'JJ' or 'CD' as a concept term.

### 3.2 Training priority scores for concept terms

In this step, to a concept term multiple weights are assigned, which describe the strength of association of a concept term with each priority level. To this aim, we employ the training data in which each cluster of tweets is labelled with a priority level. Each tweet in a cluster is associated with the label of that cluster i.e., we borrow the label from the tweet's cluster. After this, we assign a score to each concept term corresponding to its strength of association with each priority level (borrowed from the tweet's label). For example, a concept term mentioned frequently in tweets that have a specific priority level gets the highest score for that particular priority level, while the same concept term when mentioned rarely in other tweets labelled with a different priority level would get a low score for this priority level.

### 3.3 Main algorithm

In this section we describe the main algorithm that clusters the stream of tweets with respect to their topics, and which assigns them a priority level. The algorithm iteratively learns two threshold values (i.e., the content threshold and the specificity threshold) from a list of threshold values provided to it as explained in the following sections.

**3.3.1 Clustering** In this step, we cluster tweets according to their contents similarity, to the specificity of concept terms used among tweets, and to common URL mentions among the tweets. For detecting contents similarity we used the content threshold, and for determining the specificity of concept terms we used the specificity threshold. After this step we have all the tweets clustered according to their main topics.

**3.3.2 Predicting priority levels** In this step, we assign a priority level to each cluster. To this aim, we first estimate a priority level for each tweet in the corpus, and then by using the assignment of priority level to the tweets we decide a priority for each cluster. The process is explained here below.

#### *Estimate of priority level for each tweet*

First, we generate five aggregations across each priority level for a tweet. Then, the highest aggregation corresponding to a priority level becomes the priority level for that tweet. Each aggregation is computed by aggregating each

concept term’s priority score (as estimated in Section 3.2) corresponding to the priority level for that tweet.

#### *Estimate of priority level for each cluster*

Since each cluster is composed of tweets, the assigned priority level for a tweet is counted as a vote for a cluster’s priority level, and the priority level that gets the maximum number of votes (for a cluster) becomes the priority level for that cluster.

**3.3.3 Global error estimate and optimization** This step enables the algorithm to learn optimized threshold values. To this aim, we estimate the global error as follows. We first estimate the number of errors per cluster by counting the number of inconsistencies (i.e., non-uniformity) among the priority levels assigned to the tweets of a cluster. Then, we aggregate these errors estimates across each cluster to define a global error estimate. The threshold values across which the global error estimation is minimum are declared to be optimized threshold values. The output corresponding to the optimized threshold values is reported as the final output of the algorithm.

## 4 Experimental Results

### 4.1 Data set

We performed our experiments by using the data set provided by the Monitoring task of RepLab 2012 [1]. In this data set 37 companies were provided, six out of which in the training set, while the remaining 31 in the test set. For each company a few hundred tweets were provided.

### 4.2 Evaluation Measures

The measures used to the evaluation purposes are Reliability and Sensitivity, which are described in detail in [2].

In essence, these measures consider two types of binary relationships between pairs of items: relatedness – two items belong to the same cluster – and priority – one item has more priority than the other. Reliability is defined as precision of binary relationships predicted by the system with respect to those that derive from the gold standard. Sensitivity is similarly defined as the recall of relationships. When only clustering relationships are considered, Reliability and Sensitivity are equivalent to BCubed Precision and Recall [1].

### 4.3 Results

Table 1 presents a snapshot of the official results for the Monitoring task of RepLab 2012, where CIRGDISCO is the name of our team.

Table 1 shows that our algorithm performed competitively, and it is the second from the top; it is important to notice that our algorithm did not use

**Table 1.** Results of Monitoring task of RepLab 2012

Team	R Clustering (BCubed precision)	S Clustering (BCubed recall)	F(R,S) Clustering	R Prior	S Priority	F Priority	R	S	F(R,S)
UNED_3	0.72	0.32	0.4	0.25	0.3	0.26	0.32	0.26	0.29
<b>CIRGDISCO</b>	<b>0.95</b>	0.24	0.35	0.24	0.3	0.24	0.29	0.22	0.25
OPTAH_1	0.7	0.34	0.38	0.19	0.16	0.16	0.37	0.19	0.22
UNED_2	0.85	0.34	0.39	0	0	0	0.85	0.09	0.14
UNED_1	0.9	0.2	0.3	0	0	0	0.9	0.05	0.1

any external knowledge resource although sources of evidence were provided in the data set. The main reason for not using these resources was the shortage of time; this means that there is a natural room for improvement of our algorithm and for further investigation. In addition, our algorithm shows the best BCubed precision compared to other algorithms.

## 5 Conclusion

We proposed an algorithm for clustering tweets and for assigning them a priority level for companies. Our algorithm did not make use of any external knowledge resource and did not require prior information about the company. Even under these constraints our algorithm showed competitive performance. However, there is a room for improvements where external evidence could play a promising added value.

## References

1. E. Amigó, A. Corujo, J. Gonzalo, E. Meij, and M. d. Rijke. Overview of replab 2012: Evaluating online reputation management systems. In *CLEF 2012 Labs and Workshop Notebook Papers*, 2012.
2. E. Amigo, J. Gonzalo, and F. Verdejo. *Reliability and Sensitivity: Generic Evaluation Measures for Document Organization Tasks*. UNED, Madrid, Spain, 2012. Technical Report.
3. B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury. Micro-blogging as online word of mouth branding. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, CHI EA '09, pages 3859–3864, New York, NY, USA, 2009. ACM.
4. B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury. Twitter power: Tweets as electronic word of mouth. *J. Am. Soc. Inf. Sci. Technol.*, 60(11):2169–2188, Nov. 2009.
5. K. Toutanova, D. Klein, C. D. Manning, and Y. Singer. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, NAACL '03, pages 173–180, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics.