Authority and Priority Signals in Automatic Summary Generation for Online Reputation Management

Javier Rodríguez-Vidal, Jorge Carrillo-de-Albornoz, Julio Gonzalo, Laura Plaza
Universidad Nacional de Educación a Distancia (UNED)

Author Note

Javier Rodríguez-Vidal (jrodriguez@lsi.uned.es), Jorge Carrillo-de-Albornoz (jcalbornoz@lsi.uned.es), Julio Gonzalo (julio@lsi.uned.es) and Laura Plaza (lplaza@lsi.uned.es), Department of Lenguajes y Sistemas Informáticos, UNED, Madrid, Spain.

We want to thank you the anonymous reviewers for their helpful comments. This research was supported by the Spanish Ministry of Science and Innovation (Vemodalen Project, TIN2015-71785-R)

Corresponding author: Javier Rodríguez-Vidal

Abstract

Online Reputation Management (ORM) comprises the collection of techniques that help monitoring and improving the public image of an entity (companies, products, institutions) on the Internet. The ORM experts try to minimize the negative impact of the information about an entity while maximizing the positive material for being more trustworthy to the customers. Due to the huge amount of information that is published on the Internet every day, there is a need to summarize the entire flow of information to obtain only those data that are relevant to the entities. Traditionally, the automatic summarization task in the ORM scenario takes some in-domain signals into account such as popularity, polarity for reputation and novelty but exists other feature to be considered, the authority of the people. This authority depends on the ability to convince others and therefore to influence opinions. In this work, we propose the use of authority signals that measures the influence of a user jointly with (i) priority signals related to the ORM domain and (ii) information regarding the different topics that influential people is talking about. Our results indicate that the use of authority signals may significantly improve the quality of the summaries that are automatically generated.

Keywords: Authority and priority signals, Extractive Summarization, Natural Language Processing, Social Networks, Social Media

Authority and Priority Signals in Automatic Summary Generation for Online Reputation Management

Introduction

With the rapid development of Internet and Social Networks in recent years, customers have more information about entities (collectivities considered as units; e.g. companies, institutions, products, etc.) that allows them to differentiate between those that are trusted from those that are not. A good reputation is difficult to gain, and it can be ruined very fast, so it is very important to keep always a good online reputation by managing it constantly. **Online Reputation Management** (ORM) comprises the collection of techniques that help to monitor and improve the public image of an entity on the Internet. The ORM experts try to minimize the negative impact of the information on the Internet while maximizing the positive material for being more trustworthy to the customers. To do that, ORM experts need to track down all the information (good or bad) related to the client in Social Networks, blogs, specialized sites, news, etc., to produce **reputational reports** which summarize the most important issues about the client. Quick identification of the problems that affect a client may avoid a reputational crisis that may translate into money loss. According to Igniyte (2018), failing in monitoring negative content cost, in the UK in 2018, between £100,000 and £500,000 for 5% of the companies.

Monitoring online information is, however, an ambitious task. Every minute a huge amount of information is published on the Internet; only Google offered over 3.4 million answers per minute in 2018[1]. There is, therefore, a need to summarize the entire flow of information to obtain only those data that are relevant and that provide novel information.

Previous works have defined what is considered a good summary for an entity that wants to track its reputation, and it is the one which includes only information that is highly relevant for it Carrillo-de Albornoz, Amigó, Plaza, and Gonzalo (2016).

[1] http://marketingactual.es/internet/tecnologia/internet/big-data-cuantos-datosse-generan-cada-minu-to-en-el-mundo

However, what is relevant for ORM purposes differs from what is considered as relevant in traditional summarization. Priority a message from the reputational point of view depends on several factors, such as the *popularity* of information (if many people are commenting on a fact), *polarity* for reputation (if the message has positive or negative implications for the entity), *novelty* (whether it is a new problem or it is a recurring one), *authority* (there are opinion makers/influencers in the conversation) and *centrality* (the entity is the focus of the conversation).

Previous works have found that knowing the authority of the Social Networks profiles is important because some of them are capable to engage many people Rodríguez-Vidal, Gonzalo, Plaza, and Sánchez (2019), the so-called influencers. Therefore, any negative message that they post about an entity may be spread immediately among hundreds or thousands of users who, in turn, transmit it to their followers, etc., potentially causing serious reputational damage Tucker and Melewar (2005). For this reason, their opinions are good candidates to appear in the reputational summaries. In Rodríguez-Vidal et al. (2019) two types of influencers are identified: (i) people whose authority is restricted to a certain domain because they possess specialised knowledge about that domain (i.e. brokers in the economy, mechanics in automotive, etc.) or (ii) people whose authority transcends to other domains, for example in the case of celebrities, sportsmen, etc. (global authorities). We define the authority of a profile as the capacity to influence other people's opinions.

In this study, we hypothesize that the information written and spread by influencers is potentially dangerous for the reputation of an entity and therefore, should be prioritized in the generation of ORM summaries. ORM analysts should be always aware of what influential people are commenting about their clients since their opinions may (1) reach a higher audience and (2) change the opinion of such audience. To validate our hypothesis, we present two different summarization approaches: the first one combines reputational priority signals from the state of the art with different users' models that characterize them as influencers or not, as well as the type of influencer (domain authority or global authority). The second method uses a traditional topic-based summarization approach and exploits the information about the users' influence to prioritize those topics that are being discussed by influencers, under the hypothesis that the conversations where influencers are involved must be quickly managed by reputational experts.

Our best results (which improve the state of the art) are achieved with the second approach that uses language models to learn domain-specific vocabulary used by influencers to generate a signal that prioritizes the different topics of conversation about the entity. These results corroborate our hypothesis that the degree of authority and influence of the person that spreads the information is crucial for the early detection of reputational crisis, and therefore, such information must be included in the reputational summary.

Related Work

Since the advent of the Internet in the 20th century, the publication of contents has grown day by day making impossible to manually process all published information in a reasonable time. There is, therefore, a need to automatically extract and summarize all relevant information and put it in a more readable way.

Automatic summarization is a field of NLP that, since the middle of the last century, pursues this objective, and that we can be defined as the production, using automatic techniques, of output texts that include the essential information of the input documents in a shorter way. There are two main types of summarization techniques: *extractive* and *abstractive* Das and Martins (2007). Extractive summarization methods extract word sequences (phrases, sentences or paragraphs) from the original documents and copy them into the summary directly. The extractive techniques have the problem of the lack of coherence between sentences of the summarized document but stands out for its computational simplicity Das and Martins (2007). Abstractive summaries are more difficult to create because they involve paraphrasing the text in the source documents and generating text by using Natural Language Generation techniques, but they address the problem of cohesion between sentences in the summary Das and Martins (2007) and produce more readable summaries. We can also distinguish between single-document and multi-document summarization Nenkova and McKeown (2012). Whereas the first approach creates automatically the summary from the information within a single document Litvak and Last (2008), the second approach uses the information obtained from different sources, talking about the same topic, to generate a single summary Lin and Hovy (2002). This last approach may introduce redundant information (content that is expressed more than once) to the summary and, therefore, some mechanisms are necessary to avoid this problem Inouye and Kalita (2011); Takamura, Yokono, and Okumura (2011).

Recent studies address the task of automatic summarizing online information using an extractive approach.The work presented in Wang et al. (2019) proposes a method which summarizes online news articles by capturing the global context at the document level using three pre-trained steps: mask (predicting a masked sentence from a pool of candidate sentences), replace (changing sentences from a document with sentences from other documents) and switch (similar to the replace step but using sentences from the same document). Also in the context of news articles, we can find the work of Zhou et al. (2018). This study joints sentence scoring and sentence selection into one task. The authors present an end-to-end neural network which obtains the representation of the sentences with a hierarchical auto-encoder and then, builds the summary extracting the sentences one by one according to their score based on the previously selected sentences and the importance of the remaining sentences. In Bouscarrat, Bonnefoy, Peel, and Pereira (2019), the authors focused not only on summarizing online news but also on summarizing judgements from the French Court of cassation. Their method takes advantage of the use of word embeddings to generate extractive summaries. It selects sentences with the closest embeddings to the projected document embedding, this projection helps to maximize the similarity between the summaries generated and the ground truth. Lately, there is increasing interest in automatically summarize information from short documents, especially from tweets. In this context, we highlight the work of Yulianti, Huspi, and Sanderson (2016). In this study, the authors utilized information from Twitter to select the more important sentences from a web document. The summaries generated are based on a query-biased concept concerning the information extracted from tweets. Another work that uses information from Twitter is Chin, Bhowmick, and Jatowt (2019). Here, the authors use Latent Dirichlet Allocation (LDA) Blei, Ng, and Jordan (2003) topic modelling to assign similar tweets to the same topic and then, generate a ranking of relevant tweets to create a summary for each topic.

In the ORM field, the generation of automatic summaries must capture the "reputational priority" of the information about the entity. The concept of "reputational priority" was first defined in the RepLab forum Amigó et al. (2013) and categorized in three different levels (unimportant, mildly important and alert), according to the probability of firing a reputational crisis. Experiments for automatically detecting the reputational priority of tweets showed that it depends on several factors: *popularity*, *polarity*, *novelty*, *authority* and *centrality* Cossu, Bigot, Bonnefoy, and Senay (2014); Spina et al. (2013). Reputational priority was successfully exploited in the task of summary generation by Carrillo-de Albornoz et al. (2016); Rodríguez Vidal (2019). The authors of this work modelled the task as a search for diversity problem Yang, Wang, Hua, and Zhang (2010). In this task, a system provides a ranked list of documents that maximizes both relevance (documents are worthwhile to the query) and diversity (documents reflect the different query intents, when the query is ambiguous, or the different facets in the results when the query is not ambiguous). According to Carrillo-de Albornoz et al. (2016), the production of an extractive summary is similar: the texts chosen to be part of the final summary must maximize both the relevance (sentences express essential information for the entity from the input documents) and the diversity (all relevant topics related to the entity should be ideally included).

Methods

In this section, we introduce the summary generation process and the different methods employed to generate, automatically, the reputational summaries. We first present the summarization system architecture, and next, we detail the different strategies used to generate the summaries.

Summary Generation System (SGS)

The summary generation system (from now on SGS) generates extractive summaries in the form of rankings of tweets. Fig. 1 illustrates each step of the system.

The system consists of three modules: the first one extracts signals from the input tweets. Then, the second module uses the previous signals, or a combination of them, to

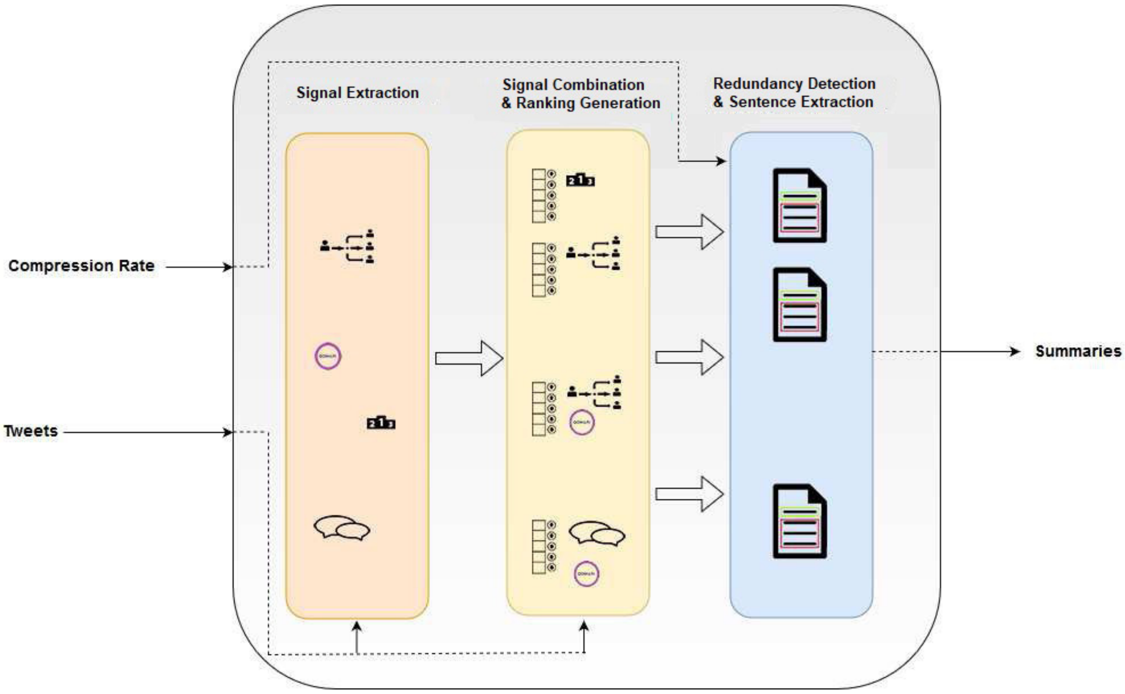

Figure 1. Summary Generation System schema

create different rankings of tweets. Finally, the third module extracts the top-k tweets from each ranking and creates summaries without redundancy until the input compression rate, which indicates the desired length of the summary is achieved. As the system output, an extractive summary is generated. Below we describe each module:

- **Signal extraction**.The first step in our system is to extract several signals of interest from a given set of tweets. Signals extracted are those presented in the next subsections.

- **Signal combination & Ranking generation**. In this step, the system receives the signals generated previously, in the signal extraction step, along with the input tweets and generates, as output, different rankings (as many as combinations of signals the algorithm performs). When the system selects two or more signals to arrange tweets, it is necessary to find a unique sorting signal that is a combination of the original ones. Combination of signals is done using one of the following algorithms: Borda voting Saari (1999) and Learning to Rank (L2R) Li, Liu, and Zhai (2009), depending on the case of study. The output of this step is a ranking of tweets sorted in descending order of priority.

- **Redundancy detection and sentence extraction**. Since the input information proceeds from multiple tweets, it is necessary to detect and remove the information that is already included in the summary, in other words, we need to detect and remove redundant information. The redundancy algorithm includes tweets from the rankings generated in the signal combination step, according to their position in it, only if the vocabulary overlap between the tweet selected and the tweets already included in the summary is less than a similarity threshold. This threshold has been experimentally set to 0.02. Similarity is calculated using the Jaccard function (Jaccard, 1901):

$$Jaccard(Tw_1, Tw_2) = \frac{|Tw_1 \cap Tw_2|}{|Tw_1| + |Tw_2| - |Tw_1 \cap Tw_2|} \tag{1}$$

Next, the system extracts the top k tweets from the ranking to be included in the summary. This parameter, k, is calculated from the compression rate provided by the user and indicates the number of tweets, from the entire input set, that must be included in the final summary. Once the draft summary is created, the system checks if its length is shorter than desired; in this case, discarded tweets are reconsidered and included in the summary by recursively increasing the similarity threshold in 0.02 until the desired compression rate is reached.

Exploiting authority and domain information to generate automatic summaries

Just as in human societies where there have always existed figures (heroes in legends, political leaders, scientists, etc.) whose ideas have been respected by their peers and their next generations, in Social Networks there are also users whose opinions influence the rest of users in the community, these users are the influencers or authorities. The authority can be circumscribed only to a certain domain, for example: banking, music, cars, etc., or it can transcend to other domains, for example, in the case of celebrities, sportsmen, etc.

The use of the global authority and the domain authority has not been previously exploited in the context of reputational summary generation. Given that the purpose of this type of summaries is to tackle possible reputational damage of an entity, and since messages from authorities can reach thousands of people, it seems reasonable to give more relevance to messages from authorities than to those from regular users (although potentially these users can also trigger reputational crises, their impact rate is lower). The textual information generated by these users is incorporated to our experimentation through the use of Language Models (LM) that model the discourses of global and domain authorities (Rodríguez Vidal (2019); Rodríguez-Vidal et al. (2019)).

This Models obtain a probability distribution of words, $p'(w)$, in which words likely to be included in an author message in the domain of authors are assigned high probability values; whereas other words, including those that are very ambiguous or not domain specific but occur in the domain of authors, receive marginalized values. This distribution of words $p'(w)$ is optimized using an Expectation Maximization procedure, in the r-th iteration, is defined as:

$$p'^{(r)}(w) = \frac{p(w|L,D) * Z(w)}{(\sum_{w' \in V} p(w'|L,D)Z(w'))} \tag{2}$$

where V is the vocabulary $w_1, .., w_{|V|}$; L and D being the background and the target domain, respectively; $Z(w)$ is the Expectation-Step and is defined as:

$$Z(w) = \frac{(1-\lambda)p'^{(r-1)}(w)}{((1-\lambda)p'^{(r-1)}(w) + \lambda p(w))} \tag{3}$$

$p(w|L)$ and $p(w|D)$ are defined as follows:

$$p(w|L) = \frac{tf(w,L)}{\sum_{w' \in L}(tf(w'))} \tag{4}$$

$$p(w|D) = \frac{tf(w,D)}{\sum_{w' \in D}(tf(w'))} \tag{5}$$

The probability of an author a belonging to the language model D is finally computed as:

$$p(D|a) = \sum_{w}(p(D|w) * p(w|a)) \tag{6}$$

where

$$p(D|w) = Z(w)$$

$$p(w|a) \propto tf(w, Y)$$

being Y the set of tweets of the author a.

After that, we compare the language of each follower with the language models of authorities (authority model) or with the language models of tweets belonging to the domain (domain model).

- *Authority model.* We hypothesise that authorities will employ a distinct way of expressing their opinions. Then, for each profile in the test set, we estimate how compatible are the profile language with the language model learned from the training set, and use one single signal to store such compatibility.
- *Domain model.* The hypothesis is that the language used, for example, to talk about football is not the same as that the language used to talk about fashion. Training is carried out with texts of the domain under consideration. Therefore, the domain signal is an unsupervised process with respect to the task, but it requires labels to assign the domain.

For a more detailed explanation about the algorithm used to calculate the different language models, see the work of Rodríguez-Vidal et al. (2019).

Exploiting priority information to generate automatic summaries

As already mentioned, previous work has defined what is considered as relevant information for ORM purposes in the form of priority signals such as polarity of information, novelty or centrality. Here we use the same priority signals used in the work of Carrillo-de Albornoz et al. (2016) and that have been empirically demonstrated to be relevant for the generation of reputational summaries. Priority signals employed are summarized in Table 1.

Signal	Definition
Author_Num_Followers	Number of Followers that a profile has
Author_Num_Followees	Number of people followed by the profile
Mentions_Count	Number of Twitter users mentioned
URLS_Count	Number of URLs in a tweet
Num_Neg_Words	Number of words with negative sentiment
Num_Pos_Emoticons	Number of emoticons associated with positive sentiment
Similar tweets in 24h	Number of similar tweets produced in a time span of 24 hours

Table 1

Set of selected priority signals

Exploiting topic information to generate reputational summaries

Topic detection has been widely used in automatic summarization. In the ORM scenario, topics give information about the different subjects of conversations (i.e., the number of robberies suffer for a bank company, factory defects appearing on a car model, etc.), that may affect to the client and should be taken into account to avoid a reputational crisis. Reputation reports must reflect these topics according to its reputational importance: first to appear in the report must be very important topics (alerts) while in the last positions of the report should appear unimportant ones. The use of topics is also useful to generate summaries automatically because tweets grouped under the same topic contain similar information and therefore, it is easier to avoid redundancy and add diversity to the final summary.

In this section, we explore the use of topics as an intermediate step to generate summaries. For this reason, we adapt the approach for topic detection in ORM, based on learning similarity functions, published in Spina, Gonzalo, and Amigó (2014). To automatically detect topics, the authors propose two different intermediate subtasks: in the first one, similarity function between tweets is learned that allows to know whether or not two tweets belong to the same cluster; the second of the subtasks uses the similarity matrix for each pair of tweets as input to an Agglomerative Hierarchical Algorithm (HAC) Schütze, Manning, and Raghavan (2008). In HAC, there is no need to specify the number of clusters a priori, it works in the following way: it first creates an individual cluster for each of the tweets; next, two clusters are agglutinated when the similarity between them exceeds a certain threshold, which acts as a condition for stopping the algorithm. According to the authors, the main drawback of this algorithm is that clusters may be merged due to single noisy elements being close to each other.

Experimental Framework

The primary focus of our experiments is to determine the importance of the signals that model the authorities and the profiles that have a broad knowledge of a domain for the automatic generation of reputational summaries. As we previously said, our hypothesis here is that the information posted by influencers (both domain-specific and global) is more important than the one published by the regular people. To do so, we perform experiments on the RepLab Summarization dataset. As far as we know, no similar resource is available for research in our task and scenario. We compare our results with state-of-the-art baselines to measure the adequateness of our approach.

The RepLab Summarization dataset

The RepLab summarization dataset[2] contains companies data from the RepLab 2013 dataset[3], where users from Twitter talk about different topics of a set of companies. Each topic consists of a different number of tweets in English and Spanish. The collection comprises tweets about 61 entities from four domains: automotive, banking, universities and music. Tweets were manually grouped by topics and, for each topic, a priority was manually assigned to it by reputational experts (possible priority values are: “Alert", “Mildly important" or “Unimportant").

To develop the RepLab Summarization dataset, only tweets from the automotive and banking domains were considered. These domains consist of large companies, i.e. Wells Fargo, Bank of America, Nissan, Fiat, etc., which are the standard subject of reputation monitoring (the annotation of universities and music bands and artists is more

[2] https://zenodo.org/record/2536801#.XDcq2lxKiUk

[3] http://nlp.uned.es/replab2013/

exploratory and does not follow widely adopted conventions as in the case of companies). As a result, the summarization subset of RepLab 2013 comprises 71,303 tweets in English and Spanish distributed as shown in Table 2.

	Automotive	**Banking**	**Total**
Entities	20	11	31
# Tweets (Training)	15,123	7,774	22,897
# Tweets (Test)	31,785	16,621	48,406
# Tweets (Total)	46,908	24,395	71,303

Table 2

Subset of RepLab 2013 used in the RepLab Summarization dataset

An annotator has presented the tweets concerning each entity grouped in topics (as in the RepLab 2013 dataset). Only “Alert" and “Mildly important" topics were considered: “Unimportant" topics were discarded as they are considered irrelevant by the reputational experts. For each tweet in a topic, the following information is available: the ID or unique identifier of the tweet, the date when the tweet was written, the number of followers of the author of the tweet, the reputational polarity of the tweet (i.e. if the tweet has positive/neutral/negative implications for the reputation of the entity), and the text of the tweet.

For each topic, the annotator was asked to generate:

- An *extractive summary*, by selecting the tweet or tweets that best summarize the content of the topic. The annotator was allowed to make no selections if she considered that no tweet is representative of the topic. We asked the annotator to be very careful not to include redundant tweets in the selection. The number of tweets selected as a representative summary ranges from 0 to 3.
- An *abstractive summary*, writing a paragraph that summarizes the content of the topic, both in English and in Spanish (note that the RepLab dataset contains tweets in both languages).

As a result, for each entity in the dataset the authors obtained (i) an **extractive**

summary consisting of a list of tweets that summarize each of the topics for that entity, ordered by the priority of the topics the tweets come from; and (ii) two **abstractive summaries** (one in English and one in Spanish), which are the concatenation of the paragraphs that summarize each of the topics regarding the entity. Fig. 2 shows the manual summaries generated for a topic (cluster) from the RepLab Summarization dataset.

```
<cluster label="K-Pax" priority="mildly_important">
    <tweet id="237940080940023809" date="Tue Aug 21 17:51:50 CEST 2012" followers="
        1835" polarity="positive"> Volvo and K-Pax: Changing the definition of a race
        car: When we're not feverishly pounding the keyboards here at... http://bit.ly
        /UVoFCJ </tweet>
    <tweet id="267133443899539456" date="Sat Nov 10 06:15:50 CET 2012" followers="83"
        polarity="positive"> Goodluck #RobertThorne at qualifying tomo! Lets go
        @KPAXracing @kpaxracingllc @volvocarsus @volvo_racing</tweet>
    <summary
      abstract_EN="Race car made by Volvo and K-Pax"
      abstract_ES="Coche de carreras de Volvo y K-Pax"
      extract="237940080940023809"/>
</cluster>
```

Figure 2. Example summaries for a RepLab topic associated to Volvo

Experiments

In this section, we explain the different experimental scenarios that we have built to generate extractive summaries automatically. Each of the following experiments takes its name based on the different signals that intervene in the generation of the ranking of tweets.

- *Authority and domain scenario*: this scenario uses the authors' authority and domain signals explained in Section **Exploiting authority and domain information to generate automatic summaries**:

    1. **Authority**: we only use the Authority signal to generate the summary. This signal estimates the likelihood of a given profile to be an authority and is obtained as explained in Rodríguez-Vidal et al. (2019).

2. **Domain**: we only use the Domain signal to generate the summary. This signal estimates the degree of domain knowledge that users have, and is obtained as explained in Rodríguez-Vidal et al. (2019).

3. **Authority+Domain**: we use the Learning to Rank (L2R) algorithm to combine Authority and Domain signals, the score obtained is used as a ranking signal.

- *Priority scenario*: this scenario incorporates the signals, described in Section **Exploiting priority information to generate automatic summaries**, for priority detection to the different cases explained in the authority and domain scenario.

  1. **Authority+Priority**: Authority and priority signals are combined using a Borda voting step Saari (1999). This voting mechanism generates a final signal which is a combination of the input ones that is later used to rank the tweets.

  2. **Domain+Priority**: Domain and priority signals are combined using a previous Borda voting step, as previously explained.

  3. **Authority+Domain+Priority**: Authority, Domain and priority signals are combined using Borda.

- *Topic scenario*: this scenario aims to exploit topic information to generate summaries. We combine this information along with authority and domain signals.

  1. **Authority+Topic**: tweets are grouped into clusters using the method explained in Section **Exploiting topic information to generate reputational summaries**. Then, each cluster is ranked according to its priority value (adding the Authority value of its tweets) and then, the tweets under the same topic are sorted according to their priority (Authority signal). The top element of each cluster is selected for the summary. Since the summary must satisfy the desired compression rate, if the number of elements in the summary is less than this compression rate, the system includes in the summary the next top element that is less redundant with the

content already included in the summary, starting from the most priority cluster, until the compression rate is achieved.

2. **Domain+Topic**: same as before, but now the priority of the clusters is assigned using the Domain signal.
3. **Authority+Domain+Topic**: same as before, but now the priority of the clusters is assigned using the L2R score of combining the Authority and Domain signals.

Metrics

Since the purpose of a summary is to condense the relevant information in the input documents, we must select a metric that is more recall-oriented than precision-oriented. For this reason, we use Recall-Oriented Understudy for Gisting Evaluation (ROUGE) Lin (2004) metric to evaluate our task. From all ROUGE variants, we selected ROUGE-2 due to its high correlation with human judges shown in many test collections. In our case, the evaluation is carried out by comparing our system outputs against both the extractive and the abstractive manual summaries provided by the RepLab Summarization dataset (see **The RepLab Summarization dataset** Section). This evaluation was done using ROUGE 2.0 tool Ganesan (2015).

Baselines

We have collected different baselines summaries, using different compression rates (5, 10, 20 and 30 %), for comparing our results.

- **Followers**: the number of followers of the user that writes the tweet is a basic indicator of priority because things said by people followed by a high number of users are more likely to be spread all over the network. Thus, in this baseline, tweets are ranked according to the number of followers of the author who wrote them. The baseline summary is built by choosing the top-ranked tweets, avoiding redundancy as in the proposed system, until the compression rate is reached.

- **LexRank**: this algorithm Erkan and Radev (2004) is one of the most popular centrality-based methods for multi-document summarization. The algorithm uses a graph, where the nodes are the candidate sentences to be included in the summary, and two nodes are connected if the similarity between them is above a given threshold. Once the graph is built, the system finds the most central sentences performing a random walk on the graph and include them in the summary until the desired length is reached.

- **Signal Selection & Voting-priority**: this baseline system uses the signals showed in Table 1 to produce several rankings of all tweets for a given test case (an entity) and then combines the rankings using Borda. Redundancy is removed as in the other approaches.

- **L2R-priority**: this baseline uses the same initial set of signals than SSV. The L2R approach makes use of a machine learning (ML) algorithm (we have evaluated several ML algorithms and finally selected random forest Breiman (2001)) and an optimization function to generate several rankings to maximize the optimization function (here we optimize nDCG metric due to its similarities with the evaluation of the proposed problem). We refer to this baseline as L2R.

Results & Discussion

In this section, we present the results of the experiments and discuss such results. Table 3 shows the results of the different experiments when (i) extractive and abstractive manual summaries are used for evaluation (see description of the RepLab Summarization dataset), and (ii) different compression rates are used.

	ROUGE-2							
	Extractive				***Abstractive***			
	Compression Ratio				***Compression Ratio***			
	5%	*10%*	*20%*	*30%*	*5%*	*10%*	*20%*	*30%*
Authority	0.12	0.21	0.36	0.48	0.09	0.16	0.28	0.35
Domain	0.15	0.24	0.41	0.52	0.11	0.18	0.29	0.36
Authority + Domain	0.14	0.24	0.40	0.53	0.10	0.18	0.29	0.38
Authority + Priority	0.15	0.24	0.40	0.51	0.12	0.17	0.30	0.37
Domain + Priority	0.16	0.25	0.42	0.53	0.12	0.18	0.30	0.37
Authority + Domain + Priority	0.18	0.26	0.43	0.54	0.12	0.18	0.30	0.38
Authority + Topic	**0.36**	**0.64**	**0.72**	**0.73**	**0.20**	**0.35**	**0.41**	0.43
Domain + Topic	**0.36**	**0.64**	**0.72**	**0.73**	**0.20**	**0.35**	**0.41**	**0.44**
Authority + Domain + Topic	**0.36**	**0.64**	**0.72**	**0.73**	**0.20**	**0.35**	**0.41**	0.43
Baseline-LexRank	0.20	0.29	0.40	0.50	0.09	0.12	0.17	0.22
Baseline-Followers	0.19	0.31	0.49	0.60	0.09	0.15	0.23	0.28
Baseline-SSV	0.24	0.36	0.52	0.64	0.12	0.17	0.25	0.30
Baseline-L2R	0.18	0.28	0.45	0.57	0.09	0.14	0.22	0.27

Table 3

ROUGE scores for the different summarization and evaluation strategies

- If we analyse the results of the *authority and domain scenario*, we can see that the use of the domain signal is better than the use of the authority signal for all compression rates and both types of reference summaries. This seems to indicate that, in specialized domains, people with some knowledge about the domain concern more to the clients, because their specialized opinion is more valuable and more valued for the general public and, therefore, is more likely to cause reputational damages. One example of a specialized domain is banking. Here the clients, i.e. financial institutions, are interested in knowing the opinion of economic gurus (such as the President of the International Monetary Fund, for instance) because their messages could affect global economy and their investments. When the evaluation is done against the extractive model summaries, these results are still far from the best baseline (SSV). When the evaluation is done against the abstract, results are similar. This seems to indicate that the information about the authority and domain expertise of the people talking about a company/entity is not the only that is relevant for ORM, and that other aspects need to be taken into account.

- From the analysis of the *priority scenario*, we can see that combining the authority and domain information with the priority signals from Carrillo-de Albornoz et al. (2016) slightly increases the evaluation results, and this is true for both types of evaluation (abstractive and extractive). This reinforces the idea of the importance of the polarity of what is written and the novelty of the information. In particular, the combination of authority and domain expertise information with the priority signals seems to be the best choice but is still far from the best baseline (SSV) in the case of the extractive evaluation.

- The results of our final approach, the *topic information scenario*, show that, for reputation monitoring, it is crucial to know the topics that people with authority and knowledge about the domain talk about. Including the topic information is important because it provides diverse content to the final report and gives a bigger picture, to the ORM experts, of the issues that may affect the entities. Combining the information about the different topics that the people are talking about with the influence of the people that take part in the conversations helps to select for the summary only those topics that may affect the reputation of the entity.

- Concerning differences between the two evaluation strategies (evaluation against extractive summaries and evaluation again abstractive summaries), it must be noted that the vocabulary overlap of the automatic summaries with the extractive summaries is expected to be higher, and therefore the absolute evaluation values are higher, in terms of ROUGE-2, than those achieved by comparing with the abstractive summaries.

Conclusions

In this paper, we have studied the effect of including authority and domain knowledge information in the automatic generation of reputational summaries. We have developed a Summary Generation System (SGS) to select the most representatives tweets about a given entity (company, product, etc.) exploiting information related to authority and knowledge about the domain of the users that spread the information in conjunction with other information typically used in automatic summarization such as priority or topic information.

Our experimental results have allowed us to arrive at the important conclusion that, in the field of marketing, being able to identify the opinions and conversations of influential people is crucial, since they can convince others and therefore to influence opinions. Therefore, good reputational summaries must take into account the influence of the people and give priority to the topics of conversation for which influencers are active. In particular, in specialized domains such as automotive and banking, the opinions of people with knowledge about the domain seems to be more influential than those of global authorities, and thus are more relevant for inclusion in the summary.

Concerning the future work, we want to: (i) enrich our reports by including some statistical information regarding the entities; (ii) creating automatic reports according to different aspects chosen by the user of the system (e.g. topic of interest, geographical area, etc.) and (iii) test the usefulness of these reports by conducting user studies.

Appendix

Example of a reputational summary

```
1  "Santander may sell U.S. car finance arm to raise cash http://bit.ly/WCi6Za "
2  "Santander planea absorber Banesto http://www.telecinco.es/informativos/economia/
     Santander-absorber-Banesto-CNMV-cotizacion_0_1526175033.html "
3  "Sernac ofició al Banco Santander por nueva falla http://bit.ly/RfDthz "
4  "Inditex, Mercadona y Santander lideran el ranking de mejores empresas para trabajar
     en España #empleo #trabajo http://ow.ly/fo7Bh "
5  "Elmo: 6 de diciembre - 5.00 Santander aumenta las alarmas sobre Salfacorp: duda que
     pueda cumplir sus compromisos de http://goo.gl/bwn65 "
6  "Santander cerrará 700 oficinas tras la integración de las filiaes Banesto y Banif.
     http://bit.ly/U6ZCy7 #economia #finanzas #bolsa #forex"
7  "Banco Santander despide a 1.200 empleados de Brasil por el pinchazo ... http://bit.ly
     /Unyo31 "
8  "El #SERNAC pidió antecedentes al Banco Santander por nuevo fallo en sus sistemas http
     ://ow.ly/fV1PT "
9  "¿Financieros?: compras en CaixaBank y Santander, ventas en Mapfre y Popular http://
     bit.ly/Tx780W #finanzas #economia"
10 "Concurso FotoTalentos`13 Fundación Banco Santander y Universia http://ow.ly/g1YEA "
11 "Santander y la burbuja: ""Algunas comunas de Santiago presentan alzas que no ... -
     Diario inmobilia... http://bit.ly/XjLdAI #inmobiliaria"
12 "Negative outlook for Santander UK says S&P: Santander UK has been taken off
     CreditWatch negative by Standard and... http://bit.ly/T5kdUT "
13 "Ingresa unos 11,9 millones Emilio Botín vuelve a cobrar todo el dividendo de
     Santander en efectivo http://www.cincodias.com/ "
14 "Anuncia Banco Santander en España cierre de 700 sucursales http://mile.io/YbODpB "
15 "Santander plans to invest in Spain's bad bank http://dlvr.it/2VSJ4K #forex"
16 "Santander y Aegon se alían para potenciar el negocio de bancaseguros en España | http
     ://Diarioelaguijon.com http://www.diarioelaguijon.com/noticia/12280/ECONOMIA-Y-
     EMPRESAS/Santander-y-Aegon-se-alian-para-potenciar-el-negocio-de-bancaseguros-en-
     Espana.html "
17 "Segunda convocatoria del programa Becas Santander. http://buzz.mw/-SJp_y "
18 "Get a Car - Enter your zip code to find dealers near you that offer financing with
     one of Santander programs. http://bit.ly/pZGfh0 "
19 "Santander considers absorbing Banesto - http://FT.com - Financial Times http://
     tinyurl.com/d2ked9s "
20 "El Santander cerrará 700 sucursales al integrar Banesto en su estructura http://ow.ly
     /g9VSJ Banesto se dispara en Bolsa"
21 "VIDEO Un grupo de jóvenes arremete contra una sucursal del Santander y revienta el
     escaparate con una valla http://www.youtube.com/watch?v=x2QevygFits #14N"
22 "Conveyancing Top solicitor pulls off Santander mortgage fraud - Bridging and
     Commerical: Top solicitor pulls off... http://bit.ly/WSWfYV "
23 "#Colombia Santander tiene un programa de tecnología para mujeres empresarias http://
     bit.ly/Wqorr5 "
24 "Santander says to close 700 bank branches after Banesto buyout: MADRID, Dec 17 (
     Reuters) - Spain's largest bank ... http://bit.ly/SDNW76 "
25 "#Spain's #Santander studying how to absorb #Banesto: http://bit.ly/Zci9x1 | #MADRID #
     Banco"
26 "Mirad gráfico al final del post y entenderéis como uno puede convertirse en banquero
     casi gratis #Santander # Banesto http://www.gurusblog.com/archives/banco-santander
     -absorber-banesto/17/12/2012/ "
27 "La absorción de Banesto por parte del Santander pone fin a 110 años de historia de la
      entidad: http://www.telecinco.es/informativos/economia/absorcion-Banesto-
     Santander-historia-entidad_0_1526175166.html "
28 "Santander México es reconocido como Banco del Año - http://bit.ly/SsTE9p "
29 "Santander invertirá 660 millones y Caixabank, 470 millones en la primera fase del
     banco malo: Santander y CaixaB... http://bit.ly/Scq4Hp "
```

References

Amigó, E., Carrillo De Albornoz, J., Chugur, I., Corujo, A., Gonzalo, J., Martín, T., ... Spina, D. (2013). Overview of replab 2013: Evaluating online reputation monitoring systems. In *Proceedings of the 4th international conference on information access evaluation. multilinguality, multimodality, and visualization - volume 8138* (pp. 333–352). Berlin, Heidelberg: Springer-Verlag. Retrieved from `https://doi.org/10.1007/978-3-642-40802-1_31` doi: 10.1007/978-3-642-40802-1_31

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993–1022.

Bouscarrat, L., Bonnefoy, A., Peel, T., & Pereira, C. (2019, July). STRASS: A light and effective method for extractive summarization based on sentence embeddings. In *Proceedings of the 57th annual meeting of the association for computational linguistics: Student research workshop* (pp. 243–252). Florence, Italy: Association for Computational Linguistics. Retrieved from `https://www.aclweb.org/anthology/P19-2034` doi: 10.18653/v1/P19-2034

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Carrillo-de Albornoz, J., Amigó, E., Plaza, L., & Gonzalo, J. (2016). Tweet stream summarization for online reputation management. In *Advances in information retrieval: 38th european conference on ir research, ecir 2016, padua, italy, march 20-23, 2016. proceedings* (pp. 378–389). Cham: Springer International Publishing. doi: 10.1007/978-3-319-30671-1_28

Chin, J. Y., Bhowmick, S. S., & Jatowt, A. (2019). On-demand recent personal tweets summarization on mobile devices. *Journal of the Association for Information Science and Technology*, *70*(6), 547–562.

Cossu, J.-V., Bigot, B., Bonnefoy, L., & Senay, G. (2014). Towards the improvement of topic priority assignment using various topic detection methods for e-reputation monitoring on twitter. In E. Métais, M. Roche, & M. Teisseire (Eds.), *Natural language processing and information systems* (pp. 154–159). Cham: Springer International Publishing.

Das, D., & Martins, A. F. (2007). A survey on automatic text summarization. *Literature Survey for the Language and Statistics II course at Carnegie Mellon*

University, *4*, 192–195.

Erkan, G., & Radev, D. R. (2004). Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research*, *22*, 457–479.

Ganesan, K. (2015). Rouge 2.0: Updated and improved measures for evaluation of summarization tasks. *arXiv preprint arXiv:1803.01937*.

Igniyte. (2018). *The reputation report 2018 - igniyte.* (https://www.igniyte.co.uk/wp-content/uploads/2018/03/The-Reputation-Report-2018-Igniyte.pdf)

Inouye, D., & Kalita, J. K. (2011). Comparing twitter summarization algorithms for multiple post summaries. In *Institute of electrical and electronics engineers third inernational conference on social computing privacy, security, risk and trust* (pp. 298–306).

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bull Soc Vaudoise Sci Nat*, *37*, 547–579.

Li, H., Liu, T.-Y., & Zhai, C. (2009). Learning to rank for information retrieval (lr4ir 2009). In *Association for computing machinery special interest group on information retrieval forum* (Vol. 43, pp. 41–45).

Lin, C.-Y. (2004). Rouge: A package for automatic evaluation of summaries. *Text Summarization Branches Out*.

Lin, C.-Y., & Hovy, E. (2002). From single to multi-document summarization. In *Proceedings of the 40th annual meeting of the association for computational linguistics.*

Litvak, M., & Last, M. (2008). Graph-based keyword extraction for single-document summarization. In *Proceedings of the workshop on multi-source multilingual information extraction and summarization* (pp. 17–24).

Nenkova, A., & McKeown, K. (2012). A survey of text summarization techniques. In *Mining text data* (pp. 43–76). Springer.

Rodríguez Vidal, J. (2019). *Authority and priority signals in online reputation monitoring* (Unpublished doctoral dissertation). Escuela Internacional de Doctorado. Programa de Doctorado en Sistemas Inteligentes.

Rodríguez-Vidal, J., Gonzalo, J., Plaza, L., & Sánchez, H. A. (2019). Automatic detection of influencers in social networks: Authority versus domain signals.

Journal of the Association for Information Science and Technology.

Saari, D. G. (1999). Explaining all three-alternative voting outcomes. *Journal of Economic Theory*, *87*, 313–355.

Schütze, H., Manning, C. D., & Raghavan, P. (2008). *Introduction to information retrieval* (Vol. 39). Cambridge University Press.

Spina, D., Carrillo-de Albornoz, J., Martí, T., Amigó, E., Gonzalo, J., & Giner, F. (2013). Reputation monitoring team at replab 2013..

Spina, D., Gonzalo, J., & Amigó, E. (2014). Learning similarity functions for topic detection in online reputation monitoring. In *Proceedings of the 37th international association for computing machinery's special interest group on information retrieval conference on research & development in information retrieval* (pp. 527–536).

Takamura, H., Yokono, H., & Okumura, M. (2011). Summarizing a document stream. In *European conference on information retrieval* (pp. 177–188).

Tucker, L., & Melewar, T. C. (2005, Jan 01). Corporate reputation and crisis management: The threat and manageability of anti-corporatism. *Corporate Reputation Review*, *7*(4), 377–387. Retrieved from `https://doi.org/10.1057/palgrave.crr.1540233` doi: 10.1057/palgrave.crr.1540233

Wang, H., Wang, X., Xiong, W., Yu, M., Guo, X., Chang, S., & Wang, W. Y. (2019, July). Self-supervised learning for contextualized extractive summarization. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 2221–2227). Florence, Italy: Association for Computational Linguistics. Retrieved from `https://www.aclweb.org/anthology/P19-1214` doi: 10.18653/v1/P19-1214

Yang, K., Wang, M., Hua, X.-S., & Zhang, H.-J. (2010). Social image search with diverse relevance ranking. In *International conference on multimedia modeling* (pp. 174–184).

Yulianti, E., Huspi, S., & Sanderson, M. (2016). Tweet-biased summarization. *Journal of the Association for Information Science and Technology*, *67*(6), 1289–1300.

Zhou, Q., Yang, N., Wei, F., Huang, S., Zhou, M., & Zhao, T. (2018). *Neural document summarization by jointly learning to score and select sentences.*