Intelligent Environments 2020
C.A. Iglesias et al. (Eds.)

doi:10.3233/AISE200063

Assessing the Impact of Tweets in Flood Events

Carlos PERIÑÁN-PASCUAL,[a1] José M. CECILIA,[a] Alicia SEPÚLVEDA-MUÑOZ,[a] Francisco ARCAS-TÚNEZ[b] and Nicolás José FERNÁNDEZ-MARTÍNEZ[b]
[a] *Universitat Politècnica de València (Spain)*
[b] *Universidad Católica San Antonio de Murcia (Spain)*

Abstract. Social sensing can provide useful information to help detect, manage and solve problems related to people's lifes and physical surroundings. Because of the huge amount of content generated on social media, the problem of social sensing is the varying quality of data, so it is necessary to filter out the irrelevant content returned by search requests. The goal of our research is to develop a knowledge-based system that is able to analyse tweets in Spanish to select the most salient posts with respect to a given problem (e.g. flood events). The main contribution of this article is to describe a measure that computes the salience of tweets by integrating the text-oriented perception of the problem with the network-oriented impact of the message. The system was tested with the natural disaster of a DANA that struck Spain in September 2019.

Keywords. Twitter, social sensor, problem detection, topic categorization, sentiment analysis

1. Introduction

Social sensing leverages user-contributed data from social media for crowd intelligence extraction. As explained by [1], social sensors may serve as a complementary or an alternative source to physical sensors. On the one hand, social sensors are complementary because they are able to explain why or how specific events occurred. On the other hand, physical sensors may not be available in scenarios where user-generated data are essential, e.g. emergency situations. As stated by [2], "social media has the potential to provide actionable intelligence to emergency services during a crisis". In this context, research aimed at analysing social-media content for disaster-management purposes has increased during the last decade, but "the field of natural hazard monitoring using Twitter remains fairly under-studied" [3]. The goal of this paper is to describe a knowledge-based system for social sensing where the impact of a given tweet with respect to a given problem is computed by taking into consideration not only how reliable we can feel that the message actually describes the problem (i.e. the text-oriented perception of the problem) but also how influential the message was

[1] Corresponding author: Universitat Politècnica de València, Escuela Politécnica Superior de Gandía, Paranimf 1, 46730 Gandia (Valencia), Spain; E-mail: jopepas3@upv.es

to other users (i.e. the network-oriented impact of the message). The remainder of this article is organized as follows. Section 2 describes some works related to social sensors for the detection of flood events. Section 3 provides an account of the implementation of our model to detect micro-texts describing problems. Section 4 evaluates the research and, finally, Section 5 presents some conclusions.

2. Related Work

Heavy rainfall can lead to severe floods that can cause disruption of critical infrastructures and human activity. Physical sensors in the form of gauging devices can only measure the amount of precipitation or the height of floodwater but not the impact on people's lifes, so social sensors become a valuable source of information. Harnessing social media to create situational awareness among citizens, emergency responders and governmental agencies during natural disasters in general, and flood events in particular, has become a relevant research topic over the last few years, where most of these studies have focused on the processing of English micro-texts from a supervised approach.

Two main types of models have been used for detecting flood events in Twitter text data. On the one hand, tweets can be categorized by using machine-learning algorithms, e.g. Naïve Bayes [2,3] or logistic regression [4]. Moreover, [5] compared the performance of Decision Trees, Naïve Bayes, and Random Forests, and [6] compared Support Vector Machine (SVM), Naïve Bayes, and Random Forests; in both cases, Random Forests provided the best results. On the other hand, tweets can be categorized by using neural networks. For example, [7] used BERT (Bidirectional Encoder Representations from Transformers). [8] compared Convolutional Neural Networks with the SVM and Random Forests and demonstrated that results were similar in performance. However, a manual analysis of the errors revealed that neural networks were better at capturing semantic characteristics relevant for the task of detecting flood-related messages.

It should be noted that the performance of supervised classifiers, grounded on machine-learning or neural-network models, is limited by the size and coverage of the training dataset. Moreover, since any event of interest has its own characteristics, the model should be trained with a different dataset for every different emergency situation (e.g. earthquakes, floods or wildfires, among others), so that the performance of the system is not affected. This requirement conflicts with the development of a multi-domain system like ours, which is intended to classify new micro-texts on the ground of dynamically created categories of social problems. For all of these reasons, our solution was aimed at dealing with flood-event detection from a knowledge-based approach.

3. Methodology

3.1. Collecting Data

Micro-texts are collected by scratching the content of Twitter feeds based on user-defined settings, such as a list of Twitter hashtags. The acquisition of tweets is performed through the Twitter API with a RESTful web service by setting specific keywords. As messages are stored in an Elasticsearch database, duplicate tweets can be filtered out by checking the MD5 hash generated for each micro-text.

3.2. Processing Natural Language

As we adopt a symbolic approach to problem detection, the system is provided with a knowledge base consisting of a number of datasets, e.g. CATEGORIES, SENTIMENTS, NEGATION and MODIFIERS. CATEGORIES is used to store the significant features related to a topic, in the form of stems together with their part of speech (POS). SENTIMENTS holds the stems of words associated with positive or negative polarity. NEGATION and MODIFIERS compose the main source of knowledge for valence shifters, i.e. words and phrases that affect the values of the topic and sentiment attributes of some of the ngrams in the micro-text.

In the first stage, each micro-text is split into sentences, and then each sentence is tokenized and POS-tagged. At this point, a tweet is represented as the vector $T_m = (w_{m1}, w_{m2}, \dots w_{mp})$, where w_{mn} represents an object for every word that occurs in the tweet and p is the total number of words. Each w_{mn} is defined with attributes such as the position in the micro-text, the word form, the stem, the POS, the topic and the sentiment. The next stage consists in detecting significant stems with respect to the topic (i.e category) and the sentiment. On the one hand, the weight 1 was assigned to the attribute topic of every w_{mn} in T_m whose stem and POS was found as a lexical feature f_{ij} in a category C_i, which was stored in CATEGORIES. On the other hand, the values p or n (i.e. positive or negative) were assigned to the attribute sentiment of every w_{mn} in T_m according to the polarity of the stem in SENTIMENTS. Finally, valence shifters are applied to neighbouring words within the micro-text. Negation cues make all the ngrams in the scope be no longer significant for topic and sentiment, so the values of their attributes are re-computed to 0. By contrast, intensifiers and diminishers change the degree of polarity of the ngrams involved by multiplying the values of the above attributes by 3 or 0.5, respectively. Whereas negation cues are applied to all the words within the scope, modifiers act only on the first polar expression that is found in the scope. The scope of valence shifters is three words, where the direction of this scope is determined by the information included in NEGATION and MODIFIERS.

3.3. Detecting Problems

We aim to determine the salience of user-generated text data by analysing two dimensions of messages. On the one hand, the text dimension helps us assess the relevance of the message, i.e. if the message contributes to situational awarenesss for managing a problem related to an in-progress event. On the other hand, the network dimension helps us assess the magnitude of the problem, i.e. we focus on the range of individuals concerned with the problem and the extent of their reactions. In this context, the most salient tweets for a given problem are detected by means of the Problem-Impact Index (PII), which combines the language-aware approach of the Problem-Perception Index (PPI) with the language-agnostic approach of the Tweet-Impact Index (TII). The remainder of this section provides a detailed account of the measures employed to obtain these scores.

3.3.1 Computing the Problem-Perception Index

The PPI is calculated not only to measure how reliable we can feel that a given tweet deals with a problem about a given hazard but also to set alert thresholds from which the severity of the problem could be rated. This measure consists of two components, i.e. Category (C_i) and Sentiment (S), as shown in Eq. (1). PPI(T_m) outcomes normalized values.

$$PPI(T_m) = \sqrt{C_i(T_m) * S(T_m)} \tag{1}$$

The computation of the PPI involves two steps. On the one hand, we calculate the Category score using cosine similarity as a measure of semantic distance. In our case, we deal with binary values for topic relatedness and the number of topic-related stems in T_m is equal to or less than the number of relevant features in C_i. Therefore, the relatedness function between T_m and C_i can be reduced to Eq. (2), where w is the number of words (unigrams) in T_m that correspond to a category feature of C_i and f is the number of all the features that serve to describe C_i.

$$C_i(T_m) = \frac{w}{\sqrt{w} * \sqrt{f}} \tag{2}$$

Indeed, C_i is regarded as the function that computes the Category score for a specific tweet with respect to a specific topic of interest. Therefore, a tweet is linked to a given category if the Category score is greater than 0.

On the other hand, we calculate the Sentiment score of given tweet with a measure originally used to assess political positions in texts. Particularly, [9] proposed the logit scale to locate party positions (i.e. left or right) on a continuous scale from the sentences of political texts that were previously coded into these two categories. Indeed, this scaling procedure allows the system to convert the counts of sentiment-coded stems in T_m into a point on the sentiment dimension by means of Eq. (3), where p and n refer to the total value of positively and negatively marked ngrams in T_m, respectively, and α is a user-adjustable parameter ranging from 0 to 1.

$$\begin{gathered} S(T_m)' = \log(p + 0.5) - \log(n + 0.5) \\ S(T_m) = \begin{matrix} 1 - \frac{1}{\log(|S(T_m)'|+2)} * \alpha & , \text{if } S(T_m)' < 0 \\ 0 & , \text{if } S(T_m)' \geq 0 \end{matrix} \end{gathered} \tag{3}$$

3.3.2. Computing the Tweet-Impact Index

Three types of measures have been devised to discover influential users in Twitter [10]:

(a) activity measures, where "users are *active* when their participation in the social network is constant and frequenct in a period of time", (b) popularity measures, where "a user is *popular* when he is recognized by many other users on the network", and (c) influence measures, where "a user is *influential* whether his actions in the network are capable to affect the actions of many other users in the network". In our case, activity and popularity measures are not pertinent, since a tweeter who is not very active or popular can post a high-impact message. Therefore, our research focuses on influence measures. In this regard, the inventory of influence measures in [11] was rather inspiring. However, since we are concerned with searching for influential messages instead of influential users, we adapted their measures for our purposes.

In this context, we devised the TII measure, which consists of three components, i.e. Retweet Impact (RTI), Reply Impact (RPI) and Information Diffusion (ID). The TII measure is computed with Eq. (4), where q is the number of unique users who retweeted T_m, r is the number of unique users who replied T_m, a is the number of unique users who posted original tweets (i.e. neither retweets nor replies) on C_i after T_m, b is the number of unique users who posted original tweets on C_i before T_m, and β is a user-adjustable parameter where $\alpha + \beta = 1$. TII(T_m) outcomes normalized values.

$$
\begin{aligned}
&TII(T_m)' = RTI * RPI * ID \\
&RTI = \log(q+2) \\
&RPI = \log(r+2) \\
&ID' = \log(a+1) - \log(b+1) \\
&ID = \begin{matrix} ID' & \text{if } ID' > 0 \\ 1 & \text{if } ID' \leq 1 \end{matrix} \\
&TII(T_m) = \left(1 - \frac{1}{\log(TII(T_m)'+1)}\right) * \beta
\end{aligned}
\tag{4}
$$

To gain a better understanding of this measure, an explanation of the notions "time frame" (*TF*) and "time slice" (*TS*) is required. Suppose that τ represents the stream of tweets, which are posted along a succession of *TFs*. In turn, each *TF* consists of a series of *TSs* of the same length, which can be seconds, minutes or hours. In other words, $\tau = \{TF_1, TF_2, \ldots, TF_k\}$ and $TF_m = \{TS_1, TS_2, \ldots, TS_n\}$, where k and n represent the number of *TFs* and the number of *TSs*, respectively. In this context, we use TS_r to refer to the time slice that becomes the focus of interest, e.g. the *TS* in which the tweet under analysis was posted. It is noteworthy that the temporal unit of *TF* and *TS* should be determined in accordance with the task in mind. For example, in the case of first responders, who must rapidly identify and understand high-impact events, *TF* and *TS* will be shorter than in the case of a system tailored for journalists.

Suppose that P contains all the original tweets whose PPI is greater than 0, being $P \subseteq T$, then it can be asserted that, for example, P_{TS_r} represents the set of all the original tweets whose PPI is greater than 0 that were posted in the current *TS*, or P_{TS_1,TS_3} represents the set of all the original tweets whose PPI is greater than 0 that where posted during TS_1, TS_2 and TS_3. Therefore, back to the ID formula in Equation (4), if T_m is posted in P_{TS_r} , then a can be formally described as $|P_{TS_1,TS_{r-1}}|$ and b as $|P_{TS_{r+1},TS_n}|$. It should also be noted that RTI, RPI and ID take into consideration only tweets that were posted in the same *TF* in which T_m was posted.

Unlike the PPI, which takes the form of a static score, the TII provides a dynamic score for T_m, which becomes static only when T_m pertains to a past *TF*.

3.3.3. Computing the Problem-Impact Index

Finally, the PII measure assesses the impact of T_m on the basis of the PPI and the TII, which can be computed in parallel through Eq. (5).

$$PII = \sqrt{PPI(T_m) * TII(T_m)} \tag{5}$$

The strength of the PII is that the PPI and the TII are complementary. On the one hand, the PPI is derived from semantic information regarding the author's attitude towards the topic of interest, which can serve to detect significant messages that, however, could not be able to generate massive activity on social media. On the other hand, the TII yields insight into data traffic on social networks, which can amplify the signal of the most influential tweets. To conclude, Figure 1 illustrates the whole process of problem detection.

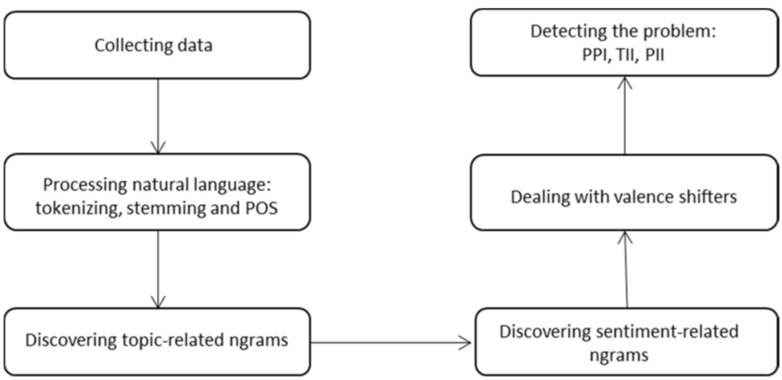

Figure 1. Knowledge-based system for problem detection.

4. Evaluation

We evaluated the research with a corpus of 8,036 tweets posted during a slow-moving storm system, officially known as a "high-level isolated depression" (Depresión Aislada en Niveles Altos, DANA), that affected about 30,000 people almost all over Spain in September 2019. In this experiment, five representative words of the event (i.e. DANA, *desbordamiento* [overflowing], *deslizamiento* [landslide], *inundación* [flood], and *lluvia* [rain]) were used to semi-automatically determine the significant features of this type of category (e.g. *aguacero* [downpour], *anegar* [flood] or *diluvio* [deluge], among many others).[2] Then, tweets dated from 11 to 29 September 2019 were retrieved by setting the significant features of the Flood category as specific keywords through the Twitter API. Figure 2 shows the averaged PPI, TII and PII scores derived for each *TF* (one day). The value of α and β in Eq. (3) and Eq. (4), respectively, was 0.5.

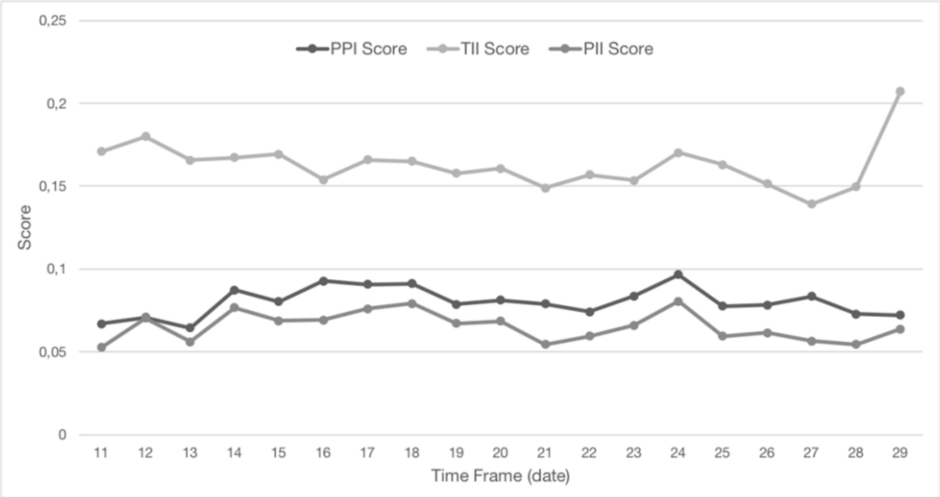

Figure 2. Averaged PPI, TII and PII scores on a time-frame basis.

To contextualize the results, we employed two supplementary information sources: meteorological reports, which give a scientific account of the occurrence of the event, and news articles, which provide insights into the situation of the event. On the one hand, the State Meteorological Agency (AEMET) reported the most relevant facts during the period under study:[3]

- 12 Sept: 300mm of rainfall in 24 hours in East and Southeast Spain (i.e. the provinces of Valencia, Alicante, Murcia, Albacete and Almería)
- 13 Sept: 200mm of rainfall in 24 hours in the provinces of Alicante and Murcia
- 14-15 Sept: the storm is moving Northwest and North Spain
- 16-17 Sept: a new DANA is moving Southwest Spain
- 18 Sept: torrential rain and severe storms in large parts of the country
- 23 Sept: Hurricane Humberto brings heavy precipation and strong winds on the Northern coast of Spain

---

[2] The WordNet-based process of lexical expansion was described in [15].

[3] The information was obtained from the 9-15 September report (https://aemetblog.es/2019/09/20/informe-operativo-de-la-semana-del-9-al-15-de-septiembre-de-2019/), the 16-22 September report (https://aemetblog.es/2019/09/23/informe-operativo-semanal-semana-del-16-al-22-de-septiembre-de-2019/), and the 23-29 September report (https://aemetblog.es/2019/10/16/informe-operativo-semanal-del-23-al-29-de-septiembre-de-2019/).

On the other hand, news agencies (e.g. Agencia EFE) reported the adverse effects of the floods:[4]

- 12 Sept: almost 300 people evacuated in Murcia
- 13 Sept: Emergency Response Plan activated in Almería; River Segura overflows; railway services suspended in Murcia, Albacete, Valencia and Alicante; the Government of Murcia strongly recommends not using the car; five people dead
- 14 Sept: President Sánchez visits flood-stricken areas in Alicante and Murcia; overflowing rivers cause the isolation of several populations and many road and railway blockages
- 15 Sept: 1,500 people evacuated from a campsite

We can conclude that the peak areas of PII shown in Figure 2 correspond to (a) the first day of the DANA (i.e. 12), (b) the day after the critical point of the storm (i.e 14), (c) the arrival of a new DANA (i.e. 17 and 18), and (d) the effects of Hurricane Humberto (i.e. 24).

Moreover, we performed a qualitative analysis to determine if our model is able to select the messages that contribute to understanding the crisis situation on the ground, thus creating situational awareness. In this regard, for example, researchers such as [12] and [13], among others, employed a test dataset where instances had been categorized by crowdsourcing workers on the basis of informativeness (i.e. related and informative, related but not informative, not related, and not applicable). As demonstrated by [14], informativeness proves to be a rather subjective category. For this reason, we chose to classify a sample of the tweets with respect to five categories that are aimed at providing citizens, emergency responders and governmental agencies with actionable information about what is happening in the affected communities during the event:[5]

- Mitigation (i.e. tweets reporting information about actions that can prevent the disaster or reduce the effects of the disaster)

(6) La falta de limpieza en los cauces es la clave en la tragedia de la gota fría de estos días.

- Preparedness (i.e. tweets reporting information about preparation, emergency plans, staying home and keeping safe, stocking up goods, evacuation, advice for behaviour during the disaster, or monitoring and tracking the disaster)

(7) La gente de Fulgencio me cuenta que están esperando a que llegue el agua ya a la zona, que ya ha anegado Dolores.

- Impact (i.e. tweets reporting information about closing businesses, disaster-caused deaths, problems with internet and utility services, infrastructure damage, things or people affected, or commuting problems)

---

[4] https://www.efe.com/efe/espana/

[5] These categories serve to reflect the main stages in which disaster management is traditionally modelled. Moreover, as suggested by [16], we include Impact, which is crucial for disaster response.

(8) Cientos de hectáreas continúan anegadas en la zona 0 de la riada del río Segura

- Response (i.e. tweets reporting information about disaster response and recovery organizations, staying in a shelter, getting free meals, emergency power, or rescues of disaster victims)

(9) Cada gota suma. Han empezado a llegar camiones de @CREAndalucia con toneladas de provisiones de agua para abastecer a las familias afectadas por la #DANA en la provincia

- Recovery (i.e. tweets reporting information about reopening businesses, removing debris, getting back to work, school or home, return of internet and utility services, fund raising and donation, repairing or rebuilding infrastructure, relief actions, or restoration of transportation services)

(10) El gobierno de la Generalitat aprueba unos míseros 23.500.000 euros de ayuda para los damnificados de las riadas de la pasada semana

In particular, the 100 most-significant tweets in our test dataset, i.e. those with the highest PII score, were manually annotated with the above categories, resulting in the following distribution: 39% Impact, 25% Other, 13% Recovery, 11% Preparedness, 10% Mitigation, and 2% Response. It should be noted that the category Other covers disaster-related tweets that are generally regarded as relevant in other studies but that were irrelevant in this experiment because they do not provide meaningful data to make a decision or solve a problem in the context of this particular event, as shown in Ex. (11) and Ex. (12).

(11) Un edil de la CUP se ríe de los policías que combaten la gota fría y les amenaza: "Mirad debajo del coche"

(12) En verdad tenemos lo que nos merecemos por estar cargándonos el planeta así que no se de que coño nos quejamos

Therefore, we conclude that precision in the 100 top-ranked tweets is 0.75.

5. Conclusions and Future Work

Social sensing is a two-way communication channel between organizations and individuals, since not only governmental agencies can deliver official information to citizens but emergency managers can also gain insight by monitoring their posts. In this context, this research demonstrated that inspecting user-generated text data allows learning what people are thinking and doing with respect to a given disaster (e.g. flood events), thus providing actionable information to be used in disaster-risk reduction and response. Future research is mainly aimed at applying a multilingual, multidomain and multimodal approach to our model of problem detection.

Acknowledgments

Financial support for this research has been provided by the Spanish Ministry of Science, Innovation and Universities [grant number RTC 2017-6389-5].

References

[1] Doran D, Severin K, Gokhale S, Dagnino A. Social media enabled human sensing for smart cities. AI Communications. 2016;29:57–75.
[2] Spielhofer T, Greenlaw R, Markham D, Hahne A. Data mining Twitter during the UK floods: Investigating the potential use of social media in emergency management. In: 2016 3rd International Conference on Information and Communication Technologies for Disaster Management (ICT-DM). 2016. p. 1–6.
[3] Arthur R, Boulton CA, Shotton H, Williams HTP. Social sensing of floods in the UK. PLoS ONE. 2018;13(1):1–18.
[4] Barker JLP, Macleod CJA. Development of a national-scale real-time Twitter data mining pipeline for social geodata on the potential impacts of flooding on communities. Environmental Modelling & Software. 2019;115:213–27.
[5] Nair MR, Ramya GR, Sivakumar PB. Usage and analysis of Twitter during 2015 Chennai flood towards disaster management. Procedia Computer Science. 2017;115:350–8.
[6] Moumtzidou A, Andreadis S, Gialampoukidis I, Karakostas A, Vrochidis S, Kompatsiaris I. Flood Relevance Estimation from Visual and Textual Content in Social Media Streams. In: Companion Proceedings of the The Web Conference 2018. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee; 2018. p. 1621–1627. (WWW '18).
[7] de Bruijn JA, de Moel H, Jongman B, de Ruiter MC, Wagemaker J, Aerts JCJH. A global database of historic and real-time flood events based on social media. Scientific Data. 2019;6(1):311.
[8] Lorini V, Castillo C, Dottori F, Kalas M, Nappo D, Salamon P. Integrating Social Media into a Pan-European Flood Awareness System: A Multilingual Approach. In: Proceedings of 16th ISCRAM Conference. Valencia; 2019.
[9] Lowe W, Benoit K, Mikhaylov S, Laver M. Scaling Policy Preferences from Coded Political Texts. Legislative Studies Quarterly. 2011 Feb 1;36(1):123–55.
[10] Riquelme F, González-Cantergiani P. Measuring User Influence on Twitter. Information Processing and Management. 2016;52(5):949–975.
[11] Pal A, Counts S. Identifying Topical Authorities in Microblogs. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. New York, NY, USA: Association for Computing Machinery; 2011. p. 45–54. (WSDM '11).
[12] Olteanu A, Vieweg S, Castillo C. What to Expect When the Unexpected Happens: Social Media Communications Across Crises. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. New York, NY, USA: Association for Computing Machinery; 2015. p. 994–1009. (CSCW '15).
[13] Zhang Y, Szabo C, Sheng QZ, Zhang WE, Qin Y. Identifying Domains and Concepts in Short Texts via Partial Taxonomy and Unlabeled Data BT - Advanced Information Systems Engineering. In: Dubois E, Pohl K, editors. Cham: Springer International Publishing; 2017. p. 127–43.
[14] Periñán-Pascual C, Arcas-Túnez F. The Analysis of Tweets to Detect Natural Hazards. In: Chatzigiannakis I, Tobe Y, Novais P, Amft O, editors. Intelligent Environments 2018. Amsterdam: IOS Press; 2018. p. 87–96.
[15] Periñán-Pascual C, Arcas-Túnez F. Detecting environmentally-related problems on Twitter. Biosystems Engineering. 2019;177:31–48.
[16] Huang Q, Xiao Y. Geographic Situational Awareness: Mining Tweets for Disaster Preparedness, Emergency Response, Impact, and Recovery. International Journal of Geo-Information. 2015 Aug 28;4:1549–68.