



HAL
open science

A Comparative Study of Microblogs Features Effectiveness for the Identification of Prominent Microblog Users During Unexpected Disasters

Imen Bizid, Nibal Nayef, Oumayma Naoui, Patrice Boursier, Sami Faiz

► To cite this version:

Imen Bizid, Nibal Nayef, Oumayma Naoui, Patrice Boursier, Sami Faiz. A Comparative Study of Microblogs Features Effectiveness for the Identification of Prominent Microblog Users During Unexpected Disasters. Second International Conference on Information Systems for Crisis Response And Management - Mediterranean Countries, Oct 2015, Tunis, Tunisia. pp.15-26, 10.1007/978-3-319-24399-3_2 . hal-01287166

HAL Id: hal-01287166

<https://hal.science/hal-01287166>

Submitted on 15 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Comparative Study of Microblogs Features Effectiveness for the Identification of Prominent Microblog Users during Unexpected Disaster

Imen Bizid^{1,2}, Nibal Nayef¹, Oumayma Naoui², Patrice Boursier^{1,3}, and Sami Faiz²

¹ L3i Laboratory, University of La Rochelle, La Rochelle, France
{imen.bizid, patrice.boursier, nibal.nayef}@univ-lr.fr

² LTSIRS Laboratory, Tunis, Tunisia
sami.faiz@insat.rnu.tn

³ IUMW, Kuala Lumpur, Malaysia
patrice@iumw.edu.my

Abstract. This paper presents a learning-based approach for the selection of relevant feature categories in the context of information retrieval from microblogs during unexpected disasters. Our information retrieval strategy consists of identifying prominent microblog users who are susceptible to share relevant and exclusive information in a disaster case. To identify these users, we evaluate the effectiveness of the state-of-the-art features characterizing microblog users for the identification of prominent users in a specific context. We experimented with a different sets of feature categories to determine those that discriminate prominent users sets from non-prominent ones interacting in Twitter during the 2014 Herault floods that occurred in France. The achieved results show that on- and off-topical user activities features are the most representative features for identifying prominent users in a disaster context. We also note that SVM outperforms the ANN learning algorithm for this classification context especially when it is trained with additional spatial features.

Key words: Effectiveness of feature categories, prominent microblog users, disaster management

1 Introduction

The climate change unleashes a multitude of unexpected disaster characteristics and effects that have never been perceived in our planet. Heat waves in summer, winter without snow, climate disruption, floods in some regions of Europe while other neighboring regions suffer from terrible droughts. Climate change manifests itself in diverse unexpected forms. Such phenomena still turn into disasters, causing irreversible damages in many places of our planet.

The challenges of managing such disasters are related especially to situation awareness and real-time information collection. The need for emergency teams to go on the disaster affected zones, risking their lives, in order to collect information

about what is taking place diminishes greatly. People from surrounding areas can provide nearly real-time observations about disaster scenes by interacting in microblogs. Citizens in the affected zones can share information about what they are experiencing; watching or hearing during a disaster. These microblogging platforms represent a rich source of information fundamental to have an accurate insight into what is happening on the ground in order to efficiently manage these unexpected disasters.

Although these microblogs such as Twitter provide many specificities (e.g. number of favorites, number of retweets of a tweet, etc.) reflecting other microblog users feedback regarding the shared information, it is still challenging to retrieve relevant and exclusive information from the huge amount of shared data. These microblogs specificities remain inaccurate as they refer mostly to the information shared by popular users independently of the relevance and freshness of their content. Therefore, it is more rational to associate the relevance and the quality of the shared information with user's prominence during the disaster.

By tracking prominent microblog users who are sharing relevant and exclusive information during an unexpected event, emergency first responders can have a real-time global view of what is happening in the threatened or affected areas. The identification of these key users have been widely explored in the context of influencers and domain experts identification. However, it has never been explored in the context of prominent users identification during unexpected disasters.

Prominent users, in the context of this paper, refer to microblog users who are susceptible to share exclusive and relevant information during a given unexpected event. Finding such users depends generally on the effectiveness of the selected categories of features describing these microblog users according to the specific context.

In this paper, we aim at evaluating the effectiveness of both the state-of-the-art and our prior proposed features describing microblog users for the identification of prominent microblog users in the context of unexpected disasters. This study focuses on the selection of the most descriptive categories of features that may lead us to the identification of these microblogs key users during disasters.

The rest of this paper is organized as follows. Section 2 reviews related works for identifying prominent microblog users. Section 3 describes the different categories of features evaluated in this paper. Experimental results are presented in Section 4. The experiments are discussed in Section 5. Finally, we conclude this paper with direction to future work in Section 6.

2 Related Work

Current Information retrieval systems in microblogs for disaster management are mostly based on the content analysis of microbogs posts. Tweak the tweet system [1] provided a hashtag based syntax to make text mining of the huge amount of information shared in microblogs during disasters easily processed. Imran et al. [2] proposed a classification model for disaster-related information extraction

by analyzing tweets text content. MicroFilters system [3] extracted the valuable disaster-related-images shared in microblogs based on image analysis techniques. These systems have yielded promising results for the identification and classification of disaster-related-content. However, they are computationally expensive, on the one hand, and they are still sensitive to redundant and outdated information on the other hand. Moreover, it is more logical to identify prominent users that may share relevant and exclusive information during the disaster and track them in order to access in real time to their shared disaster-related-information.

To the best of our knowledge, the issue of prominent users' identification has never been studied in the context of disasters. However, it has been widely explored in the different contexts defining key users as influential users in the network or as domain experts who are active and popular in a specific topic or domain [4, 5, 6].

Existing approaches for the identification of social media influencers are based on standard centrality measures such as eigenvector centrality and its variants HITS [7] and PageRank [8]. These adapted measures to microblogs specificities (e.g. number of tweets, mentions, retweets ...) are computationally expensive and sensitive to well-connected users (e.g. celebrities, communication channels...) [6]. Therefore, these approaches could not be used in real time scenarios, on the one hand, and they could not lead us to identify users sharing fresh information during unexpected disasters on the other hand.

Apart from the above research studies, domain experts identification has been explored using supervised and unsupervised learning techniques based on a set of features describing the activities of users regarding only the particular analyzed topics [9, 10, 11]. IA-Rank [12] ranked users based on the features characterizing how the user name is amplified via mentions, replies or retweets by other users. Pal et al. [9] proposed a new identification model using a set of features characterizing microblog users according to the different nature of their activities and their social position in the network. Xianlei et al. [10] proposed a Gradient Boosted Decision Tree to identify domain experts in Sina Microblog based on profile and tweeting behavior features.

These features have yielded promising results in the identification of domain experts. However, they have never been explored in the context of the identification of prominent users during disasters. Hence, in this paper, we evaluate the effectiveness of the different categories of both state-of-the-art features and our prior proposed features [13, 14] in a disaster context.

3 User Modeling Using Microblogs Features

The identification of prominent users problem in the context of disasters can be casted into a binary classification problem. Many supervised learning algorithms can be used to learn the classification model for this purpose. The performance of the used algorithms is potentially associated to the strength of the selected features to model the user behavior in a disaster context. The more the features

are representative for the prominent and non-prominent users behavior during a disaster the more the learned classifiers are efficient.

To learn the classification model, we study a large set of the state-of-the-art features and some new features proposed in our prior work [14, 13]. The listed features may reflect the behavior and the importance of each user interacting about the disaster. We have split these features into five broad categories: profile features (PrF), on-topical features (OnAF), off-topical features (OfAF), spatial features (SpF) and social network structure features (SnF). The rest of this section describes in depth these main categories of user features.

3.1 Profile Features

Profile Features (PrF) characterize the user profile description. This description is registered by the user himself (e.g. location, domains of interest...) or generated automatically by the microblogging service in order to report the user activeness rate in the network (e.g. Number of collected favorites, Number of followers...). Table 1 presents the set of user profile features. These features are extractable from any user profile using Twitter APIs.

Table 1. Profile Features (PrF) from the microblogging platform Twitter.

Name	Features
P1	Certified user [10]
P2	Enabled geolocation [14]
P3	Protected [10]
P4	Number of produced tweets [10]
P5	Number of collected favorites [14]
P6	Creation date of the Twitter account [14]
P7	Number of followers [10]
P8	Number of followees [10]

PrF give a general representation of each user independently of his activeness rate during the disaster. At the first sight, we can note that P2 and P1 may be descriptive for prominent users during disasters. P2 refers to the user information precision during the disaster where the geographic information is important. P1 can be used as a strong proof or indicator to evaluate the veracity of the information shared by each user. P7 and P8 are generally used in order to detect celebrities and domain experts. These features are evaluated in order to study if there is a correlation between the user’s popularity in the network and their prominence during unexpected disasters. Moreover, P4 and P5 which refer to the user activeness in the network are studied in order to evaluate if users who are generally active in the network may be prominent during unexpected disasters or not and vice versa.

3.2 User Activity Features

Microblog users can express what they are seeing, hearing and experiencing during a disaster using different nature of tweets:

User's own produced tweets are original tweets shared by the profile owner. These tweets are expressed by a simple content which do not include any retweet or mention symbols.

Mention tweets are tweets destined specially to particular users to make them aware about a particular information. These tweets include the @ symbol followed by the name of users to whom the tweet content is destined.

Repeated tweets are original tweets posted by someone else and rebroadcasted by the user in order to share it with his followers. These tweets are informally called retweets and can be identified by the RT@username label that is automatically inserted at the beginning of the tweet.

All these three TYPES of tweets can refer to valuable contents that are indispensable to manage unexpected disasters. Hence, we need to analyze any nature of tweets shared by users interacting during a disaster in order to identify the prominent ones.

Moreover, in order to differentiate prominent users activities from non-prominent ones, we analyze both the user's on-topic tweets related to the disaster and the off-topic ones. The categorization of on and off-topic user activities was proposed in our prior work [14] under the assumption that users affected by the disaster would be interested only by the disaster news and would neglect any other off-topical information diffused in the network.

Thus, we divide the different user activities features during the disaster into two categories On-topic Activities Features (OnAF) and Off-topic ones (OfAF). These feature categories are measured respectively according to the user on-topic and off-topic activities:

On-topic an activity is considered on-topic when it contains a subset of a list of keywords and hashtags which are defined to describe the unexpected disaster under consideration

Off-topic an off-topic activity refers to any activity that was not recorded as an on-topic one

Additionally, in this paper we assume that tweets referring to the disaster and including at least one keyword reflecting non-serious or non-valuable contents (e.g. advertising or joke words and symbols such as sale, rent, pub, lol and so on), it will be directly recorded as an off-topic one. Thus, users who share non valuable contents could be penalized.

Our rationale behind the extraction of on-topic and off-topic activities is based on penalizing users who are toggling among several topics, and who may share outdated information. Using this strategy, users are evaluated based on their impact on the analyzed disaster, and on the strength of their attachment to that disaster. For example, Top news outlets sharing news about several topics are penalized as they do not focus mainly on the analyzed disaster.

Table 2 presents both the state-of-the-art features characterizing user activities during the disaster and our prior proposed features.

Table 2. On-topic User Activities Features (OnAF) and Off-topic User Activities Features (OfAF) in Twitter.

Name	Features	On	Off
Original tweets			
T1	Number of original tweets [9, 10, 14]	+	+
T2	Number of links shared [15, 14]	+	+
T3	Number of keyword and hashtags [9, 14]	+	-
T4	Number of favorites of original tweets [14]	+	+
Retweets			
T5	Number of retweets of other’s tweets [16, 10, 14]	+	+
T6	Number of unique users retweeted by the user [14]	+	+
T7	Number of retweets of author’s tweets [14]	+	+
T8	Number of unique users who retweeted author’s tweets [16, 14]	+	+
Mentions			
T9	Number of mentions of other users by the author [17, 9, 14]	+	+
T10	Number of unique users mentioned by the author [17, 9, 14]	+	+
T11	Number of mentions by others of the author [17, 9, 14]	+	+
T12	Number of unique users mentioning the author [17, 9, 14]	+	+

On- and Off-topical user activity features will be studied separately in this paper in order to estimate the effect of each category of features in the identification of prominent users in the context of unexpected disasters.

3.3 Spatial Features

Spatial Features (SpF) characterize microblog users according to their assigned location and geolocation regarding the threatened disaster zone. Such features may be essential to determine which are the users geolocated in the disaster zone and who may play the role of sensors to diffuse information about what is really happening on the ground. The following spatial features described in Table 3 are studied in the context of disasters :

Table 3. Spatial Features (SPF) characterizing the geographic position of microblog users regarding the disaster.

Name	Features
S1	User Location [13]
S2	Shared geo-coordinates [13]

S1 indicates if the location indicated in the user profile has been stricken by the disaster or not.

S2 measures the inclusion rate of the geo-coordinates related to the user shared tweets are included in the territory threatened by the disaster or not.

3.4 Network Structure Features

Many works have explored the microblogs structure features in order to identify mostly influential and popular users. However, all of these works have used mainly time consuming algorithms that are not feasible in real time and unsuitable for the disaster context. Moreover, the used profile features referring to the number of user’s followers and followees may promote popular users who are toggling between several topic and who are sharing outdated information and neglect real prominent users having a small number of connections in the network. Thus, in this category of features, we focus only on the user followers and followees who are interacting about the disaster. Table 4 presents the network structure features studied in this paper.

Table 4. Network Structure Features (SnF) characterizing the social position of microblog users.

Name	Features
NS1	Number of user’s topical followers [13, 9]
NS2	Number of user’s topical followees[13, 9]
NS3	Number of user’s topical followers adjusted by the total number of followers [13, 14]
NS4	Number of user’s topical followees adjusted by the total number of his followees [13, 14]

4 Experiments and Results

4.1 Dataset

To conduct experimental performance evaluation on real data, we collected most of the tweets shared during the floods that have occurred from 29th to 30th September 2014 in the Hérault area, situated in the south of France. The flooded area witnessed record-shattering 252mm of rainfall in just three hours, causing important damages estimated between 500 and 600 million Euros. Data collection was processed using our multi-agent System called MASIR [18]. At the lowest level, the system detects the different users who have shared at least one on-topic tweet (i.e. talking about the floods) during the analyzed period. On-topic tweets are detected using the hashtags and keywords: #Hérault, #Herault, #intempéries, #crues, #flooding, #Montpellier, #Alert, #Inondations, #RedAlert, which were employed by Twitter users to share information about the disaster.

The system then crawls all the on-topic and off-topic tweets shared by the detected users from 29th September at 00:00AM to 1st October at 00:00AM. We collected 60195 tweets composed of on and off-topic tweets shared by 3332 users during the two days of the disaster.

Ground Truth: For the purposes of training and evaluation, users tracked in our dataset have to be assigned labels indicating whether they are prominent users or not. We have asked some volunteers from our laboratory to classify the tracked users according to the relevance and exclusivity of their tweets during the event. The number of users to be labeled varied from 66 to 200 users per volunteer.

These volunteers labeled each user as one of the two classes: C1 for prominent users, or C2 for non-prominent ones. The labeling is based on the subjective study of the tweets' content of each user (varying from 1 to 82 tweets per user). According to this study, 90 users have been classified as C1, and 3242 users have been classified as C2.

4.2 Evaluation Tools and Metrics

We describe in this section the methodologies used to evaluate the effectiveness of the different categories of features for identifying prominent microblog users in a disaster context.

Features categories selection approaches

We studied the effectiveness of the different feature categories in a disaster context with two learning algorithms. Support Vector Machine (SVM) [19] and Artificial Neural Networks (ANN) [20] are used for this study. Using these algorithms, we tested all the combination of feature categories that may represent prominent microblog users interacting during a disaster.

SVM is used in order to learn a linear SVM model for the identification of microblog prominent users in a disaster case based on different feature categories.

ANN is used in order to train and test a multi-layer ANN classification model by experimenting several combination of feature categories representative to microblog users.

In order to deal with the unbalanced data classification problem, we gave a more important weight to the class C1 of prominent users ($W_1 = 10$) than the class C2 of non prominent users ($W_2 = 1$). These parameters were set experimentally in the training phase of SVM.

As there is no parameters to tune the class weights using ANN, we have duplicated the dataset of prominent users by 30 in order to balance the two datasets of prominent and non prominent users in the training phase of ANN.

Experimental setup and Evaluation metrics

For experimental set-up, we randomly sampled 60% of both prominent and non-prominent labeled users datasets as training data for learning the classification and ranking models based on different feature categories, and the remaining 40% as test data to test the efficiency of the learned model.

Table 5. Training and test datasets description

	Training Dataset (60%)	Test Dataset (40%)
Number of Prominent users	54	36
Number of Non-prominent users	1945	1297

Through the different experiments conducted to evaluate the efficiency of the different feature categories for the classification of microblog users, we have used standard precision, recall and F1-score (i.e. F-measure) evaluation metrics.

$$\text{Precision} = \frac{\# \text{Correctly classified prominent users}}{\# \text{Users classified as prominent users}}$$

$$\text{Recall} = \frac{\# \text{Correctly classified prominent users}}{\# \text{Ground truth prominent users}}$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4.3 Feature Categories Effectiveness

In order to select the most representative feature categories for microblog prominent users in a disaster, we evaluate the effectiveness of each category of the state-of-the-art features separately. Table 6 reports the experimental results evaluating the effectiveness of each category of features using two different learning algorithms.

Table 6. Effectiveness of each feature categories for prominent users identification in terms of Precision, Recall and F1-score evaluation metrics.

Feature Categories	N ^o of features	SVM			ANN		
		Precision	Recall	F1	Precision	Recall	F1
OnAF*	12	0.43	0.86	0.57	0.29	0.80	0.42
OfAF	11	0	0	0	0.04	0.33	0.07
PrF	8	0	0	0	0.01	0.33	0.03
SnF	4	0.05	0.02	0.03	0.09	0.61	0.15
SpF	2	0	0	0	0	0	0

According to the test results returned by both the learned SVM and ANN models, the category of features characterizing the on-topical user activity in microblogs during the disaster (OnAF) is the most representative category for prominent users in a disaster context. The remaining categories of features have yielded poor results and failed to identify microblog prominent users. However, these categories may yield improvement in terms of precision and recall if they are combined with other categories. Therefore, we study the effectiveness of these categories when they are combined with the OnAF feature category. Table 7 reports the results of the different pairs of categories used to learn the ANN and SVM models for the prominent user identification.

Table 7. Effectiveness of each pair of feature categories (OnAf, An additional Feature Category) for prominent users identification in terms of Precision, Recall and F1-score evaluation metrics.

Feature Categories	N ^o of features	SVM			ANN		
		Precision	Recall	F1	Precision	Recall	F1
OnAF+OfAF*	23	0.47	0.75	0.58	0.43	0.80	0.56
OnAF+PrAF	20	0.42	0.86	0.56	0.36	0.86	0.51
OnAF+SnF	16	0.40	0.86	0.55	0.24	0.66	0.35
OnAF+SpF	14	0.43	0.86	0.57	0.39	0.88	0.54

We note that combining the two categories of features OnAF and OfAF improves the identification results using both ANN and SVM. Thus, these two categories are more efficient for the identification of microblog prominent users in the context of disasters when they are combined. Regarding the other combinations, we observe that they have a negative effect when they are combined with the OnAF category. Thus, we select the most efficient combination (OnAF+OfAF) in order to enrich it with the remaining categories of features. The identification of prominent users results using the new categories of features combination are reported in Table 8.

Table 8. Effectiveness of 3 combined feature categories (OnAf, OfAF, An additional Feature Category) for prominent users identification in terms of Precision, Recall and F1-score evaluation metrics.

Feature Categories	N ^o of features	SVM			ANN		
		Precision	Recall	F1	Precision	Recall	F1
OnAF+OfAF+SpF	25	0.48	0.75	0.60	0.41	0.80	0.54
OnAF+OfAF+PrAf	31	0.43	0.72	0.54	0.32	0.75	0.45
OnAF+OfAF+SnF	27	0.45	0.75	0.56	0.36	0.77	0.50

According to these results, there is no significant enhancements when adding a third category of features to OnAF and OfAF. Only the spatial category of features slightly improves the identification results obtained by learning the SVM model. However, training an ANN classification model based on these same categories, decreases the identification performance compared to the previous resulted ANN learned based on OnAF and OfAF categories.

We experiment in Table 9 OnAF OfAF and SpF categories with additional categories in order to evaluate the effectiveness improvement rate using our two learning approaches.

We notice that all the trained combinations reported in Table 8 have failed to improve prominent users identification effectiveness.

Table 9. Effectiveness of 4 combined feature categories (OnAf, OfAF, SpF, An additional Feature Category) for prominent users identification in terms of Precision, Recall and F1-score evaluation metrics.

Feature Categories	N ^o of features	SVM			ANN		
		Precision	Recall	F1	Precision	Recall	F1
OnAF+OfAF+SpF+SnF	29	0.45	0.75	0.56	0.38	0.80	0.52
OnAF+OfAF+SpF+PrAf	33	0.42	0.72	0.53	0.30	0.77	0.43

5 Discussion

The obtained results in this study have led us to conclude on the importance of using the on-topical and off-topical activities features categories and the spatial features category to learn an efficient classification SVM model identifying prominent microblog users during real-time disasters. On- and off-topical features are extremely useful in disaster management scenarios where prominent users focus mainly in sharing disaster-related information. Thus, using off-topical activity features, users toggling between different topics will be penalized. In addition, referring to the on-topical activities features, users focusing potentially on the unexpected disaster will be promoted. Such a property has shown that users faced by a disaster would share mainly on-topical information and neglect the other topics-related information. Moreover, we showed that users geolocated in the disaster area are more susceptible to share fresh and relevant information about what is happening around them than the others. These results have been validated using the Hérault Floods database. An open access to Twitter would be necessary to evaluate these features using more microblogs disasters databases which could not be afforded.

6 Conclusion and Future Work

In this paper, we analyzed the effectiveness of different state-of-the-art categories of features and the additional categories of features proposed in our prior work in the context of microblog prominent users identification during unexpected disasters. We tested different combinations that may lead to an efficient classification model using two different learning algorithms ANN and SVM. We found that on-topic activities category of features and the off-topic one are almost the only categories that can differentiate microblog prominent users from non prominent one in a disaster context. Moreover, we noted that the SVM algorithm is more suited to identify prominent users especially when it is learned based on the spatial, on- and off-topic categories of features.

For future work, we aim to analyze the effectiveness of each feature characterizing prominent users independently of their category using different feature selection algorithms. Moreover, we wish to propose additional engineered features which are more representative for active microblog users in the context of disaster management.

References

1. Starbird, K., Stamberger, J.: Tweak the tweet: Leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting. In: 7th international ISCRAM Conference. (2010)
2. Imran, M., Castillo, C., Lucas, J., Meier, P., Rogstadius, J. In: Coordinating human and machine intelligence to classify microblog communications in crises. The Pennsylvania State University (2014) 712–721
3. Ilyas, A.: Microfilters: Harnessing twitter for disaster management. In: GHTC. (2014) 417–424
4. Weng, J., Lim, E.P., Jiang, J., He, Q.: Twiterrank: Finding topic-sensitive influential twitterers. In: WSDM, ACM (2010) 261–270
5. Zhang, M., Sun, C., Liu, W.: Identifying influential users of micro-blogging services: A dynamic action-based network approach. In Seddon, P.B., Gregor, S., eds.: PACIS. (2011) 223
6. Silva, A., Guimaraes, S., Meira, Jr., W., Zaki, M.: Profilerank: Finding relevant content and influential users based on information diffusion. In: SNAKDD, ACM (2013) 2–9
7. Agichtein, E., Castillo, C., Donato, D., Gionis, A., Mishne, G.: Finding high-quality content in social media. In: WSDM. (2008) 183–194
8. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: WWW. (2010) 591–600
9. Pal, A., Counts, S.: Identifying topical authorities in microblogs. In: WSDM, New York, NY, USA, ACM (2011) 45–54
10. Xianlei, S., Chunhong, Z., Yang, J.: Finding domain experts in microblogs. In: WEBIST. (2014)
11. Ghosh, S., Sharma, N., Benevenuto, F., Ganguly, N., Gummadi, K.: Cognos: Crowdsourcing search for topic experts in microblogs. In: SIGIR. (2012) 575–590
12. Cappelletti, R., Sastry, N.: Iarank: Ranking users on twitter in near real-time, based on their information amplification potential. In: SOCIALINFORMATICS. (2012) 70–77
13. Bizid, I., Boursier, P., Morcos, J., Faiz, S.: A classification model for the identification of prominent microblogs users during a disaster. In: ISCRAM. (2015)
14. Bizid, I., Nayef, N., Boursier, P., Faiz, S.: Prominent users detection during specific events by learning on- and off-topic features of user activities. In: ASONAM. (2015) –accepted to appear–
15. Java, A., Pranam, Finin, T., Oates, T.: Modeling the spread of influence on the blogosphere. In: www. (2006)
16. Boyd, D., Golder, S., Lotan, G.: Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In: 43rd Hawaii International Conference on System Sciences (HICSS). (Jan 2010) 1–10
17. Honey, C., Herring, S.: Beyond microblogging: Conversation and collaboration via twitter. In: 42nd Hawaii International Conference on System Sciences. (2009) 1–10
18. Bizid, I., Boursier, P.G., Morcos, J., Faiz, S.: Masir : A multi-agent system for real-time information retrieval from microblogs during unexpected events. In: 9th International KES Conference on Agents and Multi-Agent Systems: Technologies and Applications. KES-AMSTA-15 (2015) 3–13
19. Osuna, E., Freund, R., Girosi, F.: Support vector machines: Training and applications. Technical report, Cambridge, MA, USA (1997)
20. Zhang, G.P.: Neural networks for classification: a survey. In: and Cybernetics - Part C: Applications and Reviews. (2000)