GEOINFO
Brazilian Symposium on GeoInformatics
November 29ᵗʰ to December 2ⁿᵈ, 2015 Campos do Jordão, SP, Brazil

GEOGRAPHICAL PRIORITIZATION OF SOCIAL NETWORK MESSAGES IN NEAR REAL-TIME USING SENSOR DATA STREAMS: AN APPLICATION TO FLOODS

Priorização Geográfica de Mensagens de Redes Sociais Próximo do Tempo-Real Usando Streaming de Dados de Sensores: Uma Aplicação para Inundações

Luiz Fernando Ferreira Gomes de Assis[1,2], João Porto de Albuquerque[2,3,4], Benjamin Herfort[3], Enrico Steiger[3] & Flávio Eduardo Aoki Horita[2]

[1]National Institute for Space Research – INPE
Image Processing Division
Av. dos Astronautas, 1758, Jd. da Granja, São José dos Campos - SP, 12227-010, Brazil
luizffga@dpi.inpe.br

[2]University of São Paulo – USP
Institute of Mathematics and Computer Science (ICMC)
Av. Trabalhador Sancarlense, 400, Centro, São Carlos- SP, 13566-590, Brazil
horita@icmc.usp.br

[3]Heidelberg University
GIScience Research Group
Grabengasse 1, 69117, Heidelberg, Germany
enrico.steiger@geog.uni-heidelberg.de, herfort@stud.uni-heidelberg.de

[4]University of Warwick
Centre for Interdisciplinary Methodologies
Gibbet Hill Road, Coventry CV4 7AL, United Kingdom
j.porto@warwick.ac.uk

ABSTRACT

Social networks have been used to overcome the problem of incomplete official data, and to provide a more detailed description of a disaster. However, the filtering of relevant messages on-the-fly remains challenging due to the large amount of misleading, outdated or inaccurate information. This paper presents an approach for the automated geographic prioritization of social network messages for flood risk management based on sensor data streams. It was evaluated using data from Twitter and a monitoring agency of Brazil. The results revealed that the proposed approach is useful for identifying valuable flood-related messages in near real-time.

Keywords: Social Network Message, Sensor Data Stream, Floods.

RESUMO

As redes sociais têm sido utilizadas para superar o problema de dados oficiais incompletos, e fornecer uma descrição mais detalhada de um desastre. No entanto, a filtragem de mensagens relevantes dinamicamente continua sendo um

desafio devido à grande quantidade de informações enganosas, desatualizadas ou imprecisas. Este artigo apresenta uma abordagem para a priorização geográfica automatizada de mensagens de redes sociais para a gestão dos riscos de inundação com base em streaming de dados de sensor. A proposta foi avaliada utilizando dados do Twitter e de uma agência de monitoramento do Brasil. Os resultados revelaram que a abordagem proposta foi capaz de identificar mensagens valiosas relacionadas às inundações de forma próxima ao tempo real.

Palavras chaves: Mensagens de Redes Sociais, Streaming de Dados de Sensores, Inundações.

1. INTRODUCTION

The growing number of natural disasters, such as floods, has been leading for better preparation from vulnerable communities. In this sense, in-situ and mobile sensors are providing historical and updated information through the monitoring of environmental variables (e.g. temperature of the water or the volume of rainfall). Although these data are useful for supporting decision making, further information is required for estimating the overall situation at an affected area (HORITA *et al.* 2015). Social networks like Twitter, Facebook, and Instagram, can overcome this issue either by providing information from areas which are not covered by sensors or complementing semantically the data provided by them (STARBIRD & STAMBERGER 2010, VIEWEG *et al.* 2010, ZIELINSKI *et al.* 2013, HORITA *et al.* 2015).

Despite the fact that the combination of sensor data streams and social network messages might provide better information for supporting decision-making in critical situations like floods (MOONEY & CORCORAN 2011), it raises some challenges. On the one hand, the huge volume of messages shared through social network makes difficult the identification of relevant messages, i.e. decision-makers do not want to analyze thousands of messages, they need the most valuable in order to make faster their decision-making (VIEWEG *et al.* 2014). On the other hand, the near real-time integration of sensor data and social network messages raises issues regarding the combination of distinct data streams (e.g. per second or per minute) and different data formats (e.g. numbers and texts) (DOLIF *et al.* 2013).

In this context, this paper aims to present an approach for supporting flood risk management by means of the near real-time combination of social network messages and sensor data streams. It extends our previous works (ASSIS *et al.* 2015a, ASSIS *et al.* 2015b, ALBUQUERQUE *et al.*

2015) by adopting a workflow analysis which structures and defines an automated near real-time prioritization of social network messages based on the sensor data stream. Furthermore, it describes the formal representation of the problem statement and makes an evaluation of the approach through case studies. In summary, the main contributions of this work are described below:
• To define an approach to combine a sensor data stream with social network messages, aiming at identifying high- value messages in near real-time for flood risk management;
• To learn lessons from the application of the proposed approach in a real-world flood scenario.

The remainder of this paper is structured as follows. Section 2 examines the related works. Section 3 sketches the background of the basic concepts and introduces the approach and methodology employed in this work. Section 4 describes the evaluation of the approach and their results are analyzed in Section 5. Finally, Section 6 draws conclusions and recommends some future works.

2. RELATED WORKS

Several applications have attempted to combine authoritative and non-authoritative data to improve the limitations of each other. Existing approaches integrate authoritative and non-authoritative data to provide location-based eventful visualization, statistical analysis and graphing capabilities in near real-time (WAN *et al.* 2014, SCHNEBELE *et al.* 2014). The combination of information provided by a collaborative platform (esp. Ushahidi) and sensor data collected via a wireless sensor network have been built for decision-making in flood risk management (HORITA *et al.* 2015).

Several other studies aim at analyzing the large amount of information provided by social networks (EDIGER *et al.* 2010, GAO *et al.* 2011), exploring e.g., relations between spatial information from both social network messages

and knowledge about flood phenomena (ALBUQUERQUE *et al.* 2015). An algorithm for monitoring social network messages (esp. tweets) and detecting upcoming events is another eminent approach (SAKAKI *et al.* 2010). This algorithm classifies tweets using their keywords, number of words, and context. There are also systems for processing and analyzing social network messages in near real-time (SONG & KIM 2013). The results of its application for monitoring Korean presidential elections showed that social network could support the detection and prediction in the changing of communities' behavior. Finally, examining earliest social network messages that have produced a trend with the aim of identifying and creating a classification schema allows a categorization of messages, and thus a discovery of potential trends in near real-time (ZUBIAGA *et al.* 2015).

Although some studies have been done in the field, none of them tackles the combined use of social network messages and data collected from sensor streams in near real-time. In this manner, the processing of different data flows and data formats still pose challenges for the use of sensor data as an alternative to support the filtering and extraction of high-value social network messages in near real-time.

3. PROBLEM STATEMENT AND AP-PROACH

In this section, we first present the research design of prioritizing social network messages based on sensor data streams in a flood situation. After that, we describe two workflows of how combining sensor data and social network messages requires dealing with important decisions in several phases.

3.1 Prioritization of Location-based Social Network Messages

This section presents the research design of prioritizing social network messages based on sensor data streams in a flood situation. Firstly, we introduces the problem statement, which is followed by the related research question and our working hypothesis. Then, we provides two definition that are important for the understanding of the overall work.

3.1.1 Problem Statement

Due to the high volume of social network messages, the process of extracting relevant messages has been becoming more complex. This is because most of these messages are shared from several platforms (e.g. Flickr and Twitter) in distinct formats (e.g. photos and texts) with different data flows (e.g. per hour or per second).

3.1.2 Research Question

Is a sensor data stream able to support the near real-time identification of relevant social network messages in flood risk management?

3.1.3 Hypothesis

Given a set of catchments C, sensor data stream S and location-based social network messages M, we assume that the n-messages $M = \{m_1,...,m_n\}$ nearest to the m-flooded areas $F_{tr} = \{f_1,...,f_m\}$ available at a time t_r tend to be more flood-related than the more distant (p-n)-messages $M = \{m_{n+1},...,m_p\}$, where n, m, p and r ε N, and t is a timestamp. F is defined here as a time series of flooded areas. $F = \{F_{t1}, ..., F_{tr}\}$.

3.1.4 Definition 1

This paper uses a set of georeferenced catchments $C = \{c_1,...,c_n\}$ R^2. Each c_j denotes a two-dimensional Euclidean space that can either contain sensors or not. A sensor data stream S $= \{s_1,...,s_m\}$ contains a set of continuous sensor data $s_k = (i, v, t, p, c)$. Each sensor data has an id $s_k.i$, a water level value $s_k.v$ at a timestamp $s_k.t$, a geographic position $s_k.p = (x; y)$ and a s.c catchment. In case a sensor data s_k contains $s_k.v$ equal to "high" at a timestamp $s_k.t$, and a $s_k.p$ within a catchment c_j, this catchment c_j become a flooded area f_p 2 $F_{sk.t}$ C. The f_p is available until that $s_k.v$ and any other sensor value contained in the catchment c_j are not "high" anymore at a subsequent timestamp to $s_k.t$.

$$F = \{F_{t1},..., F_{tr}\} \qquad (1)$$

$$F_{tr} = \{f_0, ..., f_p\} \mid \mathsf{Z} \; s_k \; \varepsilon \; S \text{ and } f_p \; \varepsilon \; F_{tr}, \qquad (2)$$

where,

$s_k.v$ = "high" and $s_k.c = f_p$ and $s_k.t = t_r$

Location-based social network users U = {u₁,...,u_q} produce georeferenced messages represented by M = {m₁,...,m_n}. Each message m_i = [u, t, v, g] contains a value text $m_i.v$, as well as is produced by an user $m_i.u$ at a timestamp $m_i.t$ in a geographic location $m_i.g$. If there is a flooded area f_p, for each new incoming message m_i, a distance d is computed by the nearest neighbour (Euclidean) distance of the $m_i.g$ position to every element of the set F_{tr}) = ff_p$g^n_{p=0}$ at a timestamp t_r.

3.1.5 Definition 2

In general, the cartesian minimum distance between two points p (message location) and p' (the nearest point contained in a flooded area) in a Euclidean space R^n is given by:

$$d(p, p') = \sqrt{(\Sigma n_{i,j=1} |p_i-p_j|^{2)}}, \text{ where } i,j \varepsilon N \quad (3)$$

3.2 Sensor Data Stream and Social Network Message Workflows

Given the extent of the flood phenomena, spatiotemporal characteristics of both sensor data stream and social network can be combined and more explored. Social networks messages can be used to complement sensor data stream with semantic information while sensor data stream can add reliability to the social network messages. For this reason, this approach is designed to suit an on-the-fly prioritization for different levels of data availability that are required to ensure an effective flood risk management.

The proposed workflow is performed in a pipelined way that contains both a sensor data stream S and a social network messages stream M. In the sensor data stream part (see Figure 1), there are three possible kinds of data availability. The first alternative is to have highly qualitative information about the extent of the flood phenomena, which can be provided by Unmanned Aerial Vehicles (UAVs). Although they provide the best degree of accuracy for detecting events in near real-time, they are rarely gathered.

If no direct sensor data is able to show the existence of a flooded area f_p, thus it is analyzed if they can either provide a flood hazard area or not. Local data about the affected areas e.g., maps of risks can potentially provide this kind of information. In the last stage of the verification, if neither the flooded area f_p nor the flood hazard area are initially available, a digital elevation model (provided by a user) is used to calculate a flood hazard area. After calculating this flood hazard area, they are used as an input (along with threshold data) to calculate the flooded area f_p. The flooded area f_p in this part is used to calculate the prioritization-based distance $P(m_j)$ when a new social network message m_j arrives (see Equation 4). In case more than one flooded area is available, the nearest flooded area distance is assign to the message prioritization.

$$p(m_j) = min(d(m_j; f_P)); \quad (4)$$

where,

$$m_j \varepsilon \ M; f_p \ \varepsilon \ F \ and \ j; p \ \varepsilon \ N:$$
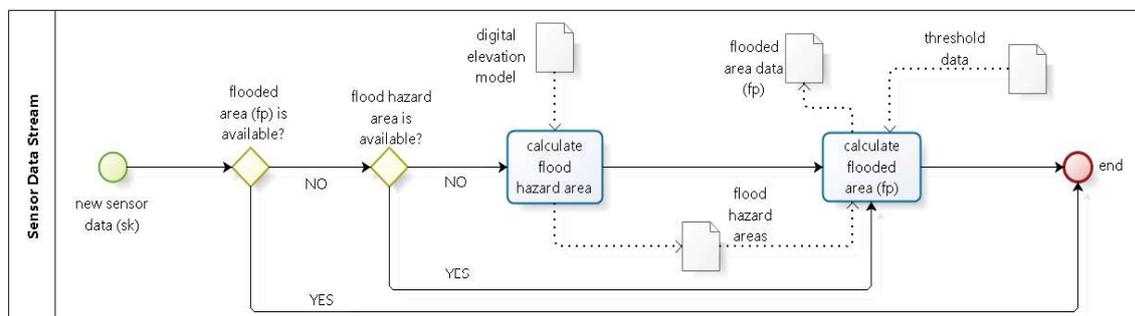
Fig. 1 - Sensor Data Stream Workflow.

Social network messages m_j and sensor data s_k should gather simultaneously so that even delayed flood-related messages can be acquired. In the social network message stream part (see Figure 2), for each new incoming social network message m_j, prioritization-based distance is computed according to the nearest existing flooded area available produced by the sensor data stream part of the workflow. After calculating this prioritization-based distance, the messages are filtered to find messages that are likely to refer to a flood event.
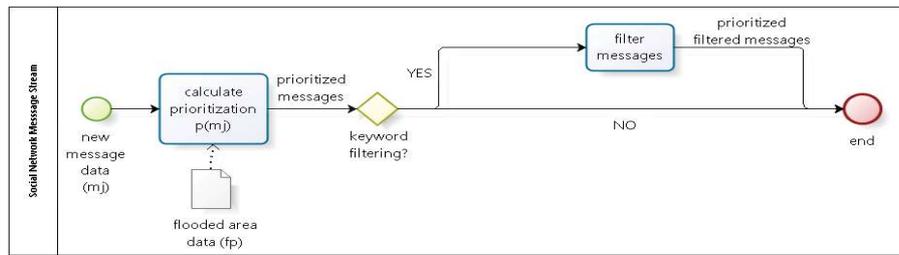
Fig. 2 - Social Network Message Workflow.

At first, our aim is to store all the messages since the specific keywords for floods might change during a flood. In this way, the filtering process can be easily adjusted without losing any flood-related messages.

4. CASE STUDIES AND EXPERIMENTAL SETUP

The approach was evaluated by means of an application case study of floods in Brazil since it is currently taking measures to cope with and alleviate flood situations. This set of measurements includes activities that involve pre-flood planning, managing emergency situations and post-flood recovery (AHMAD & SIMONOVIC 2006). Updated knowledge of river conditions plays an important role in supporting decision-making since several technical factors can prevent this kind of information from being obtained. Countries such as Brazil where there are flash floods caused by heavy rain or the overflow of streams and narrow gullies needs this kind of management to mitigate the damage to the local infrastructure. This means that emergency agencies have to cope with the risk of human casualties and the extent of flood damage in their decision-making.

In this application case study, we considered as the data source, authoritative static data (shapefiles), authoritative dynamic data (sensor data streams), and social network messages (Twitter).

4.1 Case Study: Flash Floods in Brazil

The analysis is confined to São Paulo as a single region within Brazil so that it is easier to compare and link the results of the case study. The shapefile of the State of São Paulo was produced using geotools1 operations and geo-referenced data sets provided by HydroSHEDS2.

It contains 315 small catchments. The sensor data streams were obtained from the Brazilian National Center for Monitoring and Early Warning of Natural Disasters (Cemaden). This agency operates by continuously installing new stations and providing their data through a Rest API.

In the State of São Paulo, Cemaden provided data from 465 stations. Each station and measurement of Cemaden are combined in the same file. The provided measurements from all the stations regarded the last four hours, considering an offset. This is the difference between the distance of the installation position and the real water level. As soon as it starts raining, the stations measure the rainfall every 10 minutes, otherwise, they measure every 60 minutes. The floods in Brazil are represented at their most extreme by flash floods.

The catchments, stations, and flooded areas are depicted in Figure 3, while a density map of the prioritized tweets is shown in Figure 4.

4.2 Experimental Setup

In this subsection, we describe in more details the runtime environment of the experiments.

4.2.1 Runtime Environment:

The implementation to retrieve Cemaden data was based on a simple Java toolkit for JSON[3], while tweets were retrieved using a Java library for Twitter API called Twitter4j [4]. Both of them were implemented in a pipelined fashion as a data stream. The experiments were run on a server with 2GHz AMD Opteron(tm) Processor 4171 HE and 3.4 GB RAM memory running Ubuntu 12.04.5 LTS (64 bit). Twitter Streaming API and Cemaden Rest API were used to retrieve data in the period from April to May 2015.
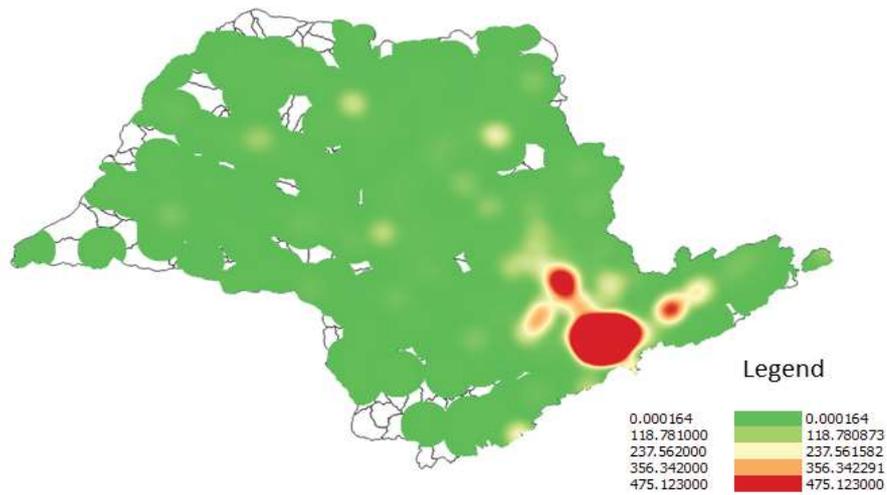
Fig. 3 - São Paulo - An analysis of Brazilian Stations and Catchments.
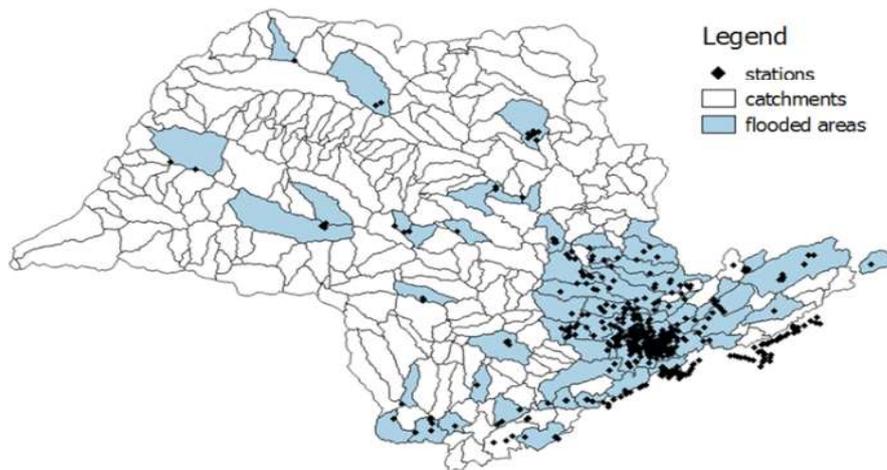
Fig. 4 - São Paulo - An analysis of Prioritized Tweets during periods of floods

Table 1: Keyword-based filtering description of the location-based social network messages

# all the tweets	# prioritized tweets	# prioritized flood-related tweets	# prioritized non flood related tweets
1,136,583 (100%)	68,195 (6%)	403 (0,04%)	67,792 (5.96%)

Table 2: Sensor measurement description of Cemaden stations

# catchments (flooded areas)	# stations (high values)	# measurements high values)	# all the measurements
59 (18.7%)	311 (66.9%)	284,663 (100%)	1,030 (0.0037%)

5. RESULTS AND ANALYSIS

In our case study, only 6% (68,195) of the tweets were prioritized mainly due to the flash floods. Such application case study helped to represent the scenarios when a flood occurs since a large number of tweets were posted at critical moments. At first, all the flood-related tweets (403) were filtered by making a selection of the prioritized tweets (1,136,583) based on specific keywords and their synonyms. This "keyword selection" was based on the Brazilian words in the dictionary for "flood", taking into account differences of case sensitive letters without spelling mistakes. The Brazilian keywords were "enchente", "inundação" and "alagamento".

Fig. 5 - Examples of flood-related tweets containing images that can help in flood risk management

We also decided to calculate the time that our approach takes to prioritize all the tweets. The latency of the tweets was considered to be acceptable for the prioritization of social network messages for floods. This latency was the difference in time between the arrival from the Streaming API and the storage at the database. The latency was only calculated for the prioritized tweets. The processing average time per minute of all the prioritized tweets was less than one second.

Tweets containing relevant information presented not only flood-related text but also georeferenced images that can help in flood risk management. As can be seen, exemplary prioritized tweets containing relevant messages are shown in Table 3 and Figure 5.

In addition, a statistical hypothesis test was conducted to check whether flood-related tweets are closer to hazard areas than those that are non-flood-related during floods. The Mann-Whitney U-test was employed for the samples of prioritized flood-related and non-flood related tweets. The test returns the p-value of a two-sided Wilcoxon rank sum test, which tests the null hypothesis that the distance of independent samples with different lengths from continuous distributions of flood-related and non-flood related tweets are with equal medians, against the alternative that they are not. The p-value of $7.2940e-016$ indicate a rejection of the null hypothesis of equal medians at a 5% significance level. That means, the sample of tweets which contain flood-related keywords are not equally nearer to the hazard areas to the ones that not contain flood-related keywords. The median distance of the sample of non- flood-related tweets was 10,905 meters away from those areas, while the sample of flood-related tweets was 3,027 meters away. Figure 6 shows the two distribution of non flood-related and flood-related tweets based on their prioritization (distance in km to flooded areas).

Table 3: Examples of prioritized Tweets containing flood-related keywords without images (On Topic, Relevant)

Flood-Related Prioritized Tweets	Translation
"tá td alagado aq na marquês"	"everything is flooded here in the Marquês" (name of a place or avenue name)
"Ta alagado aqui"	"It is flooded here" (the georeference of the tweet can supplement this information)
"Nações alagada"	The "Nações (Avenue Name) is flooded"
"Alagamento na av dos Tajuras (at @AG2 Nurun in São Paulo, SP)" https://t.co/r7ECeAtiFB"	"There is a flood in Tajuras avenue" (the georeference of the tweet can suplement this information)
"@VCnoSPTV chuva de 30 minutos e alagamento na região do Brás, pra variar http://t.co/wWVqIKtz3z"	"@VCnoSPTV (Twitter account of a TV program) it has been raining for 30 minutes and the region of Bras is flooded, as always http://t.co/wWVqIKtz3z"
"5min de chuva e rua já fica alagada"	"After 5 minutes of rain, the street is already flooded" (the georeference of the tweet can supplement this information)
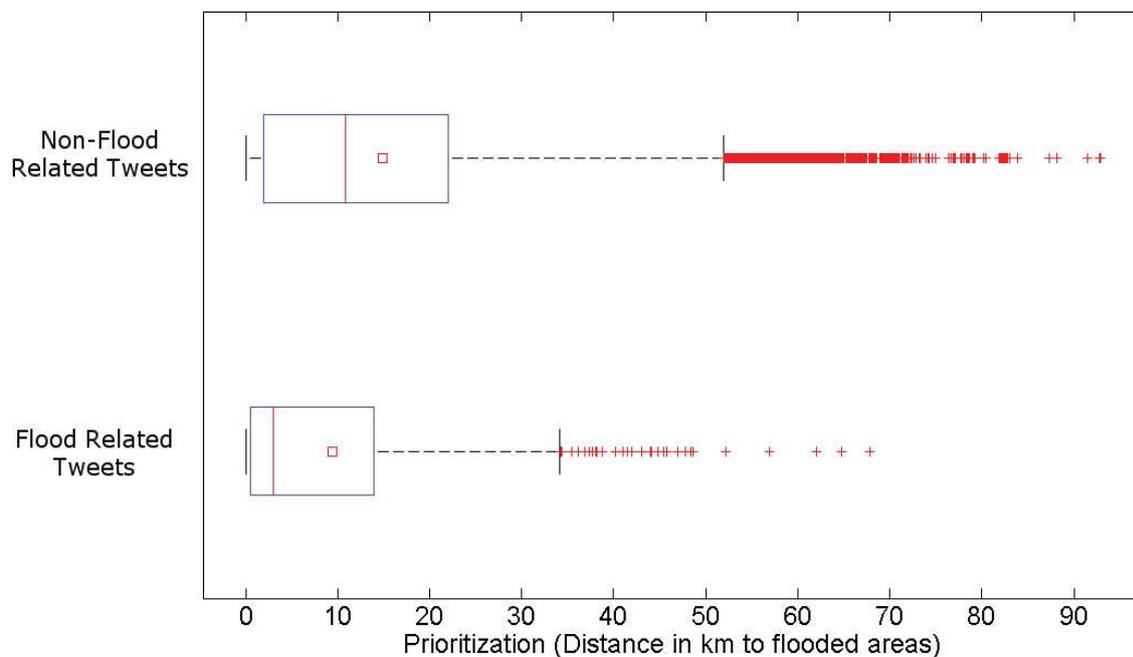
Fig. 6 - Median, Average, and Outliers of flood-related and non flood-related tweets

6. CONCLUSION AND DISCUSSIONS

This paper presents an approach for supporting flood risk management by means of a near real-time prioritization of social network messages based on sensor data streams. A case study was used for evaluating the approach. The results confirmed that the geographical relations are useful for prioritizing social network messages related to floods. They showed that there are about 3,6 times more flood-related social network messages near to flood-affected areas than non-flood-related messages. Although our approach was evaluated in a specific context of floods and using Twitter messages, it could be adapted to other types of disasters (e.g. earthquakes and landslides) and social networks (e.g. Instagram and Flickr), i.e. considering images or videos instead of only texts messages.

Our approach processed the messages during a flood event at an average time of less than one second. Given a large number of messages (peaks should be treated as critical periods since more messages tend to be posted), such processing time to prioritize does not significantly change. This work has shown that social network messages and sensor data streams can complement each other. Sensor data streams are accurate, dynamic, heterogeneous and continuous, although they are scarce and hard to implement and maintain. On the other hand,

social network messages can enhance semantic sensor data, but their large number is not easy to handle since they can be misleading, outdated or inaccurate. Despite the lack of user experience and knowledge, social networks have been used in crisis management revealing their remarkable and positive features.

Although most of the existing approaches are still insufficient for near real-time decision-making since they fail to take note of the fact that data in disasters should be analyzed on-the-fly and automatically. Our approach searches for georeferenced social network messages using a grid 5x5 bounding box based on the catchments dimension. Although most of the messages are not flood-related (and do not contain any important keywords such as "floods" or "inundation"), they were stored in the database after first being filtered because a keyword search is arbitrary, especially for near real-time event detection.

All the social network messages located within a flooded area were prioritized with zero meters "0 m" as distance, which is the main value-based prioritization. Some of the prioritized messages have images embedded in them, which were really useful when they were geolocated because they could show the exact situation of a particular place and sometimes helped more than simply by the words. During

our analysis, a few of the total amount of available messages were both georeferenced and considered to be flood-related.

Furthermore, heavy rains might affect the connection infrastructure (e.g. cell phone services or wi-fi), which in turn may reflect on the unavailability of information sharing. Although this issue is important when dealing with social network messages, there are several types of disasters that do not compromise infrastructure. Furthermore, past examples have shown that the information produced after the re-establishment of infrastructure is still valuable for the coordination of relief efforts.

Future work lines should take account of using our approach for prioritization of social network messages as one step of further filtering and classifying the quality of crowdsourcing. Besides that, this approach can serve as a basis to improve machine learning models to include the consideration of geographical relations.

ACKNOWLEDGMENTS

REFERENCES

AHMAD, S. & SIMONOVIC, S. P. (2006). An Intelligent Decision Support System for Management of Floods. **Water Resources Management**, 20(3):391–410.

ALBUQUERQUE, J. P.; HERFORT, B.; BRENNING, A. & ZIPF, A. (2015). A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. **International Journal of Geographical Information Science**, 29(4):667–689.

ASSIS, L. F. F. G.; HERFORT, B.; STEIGER, E.; HORITA, F. E. A. & ALBUQUERQUE, J. P. (2015b). Geographical prioritization of social network messages in near real-time using sensor data streams: an application to floods. **XVI Brazilian Symposium on Geoinformatics (GEOINFO)**, pp. 26-37. Campos do Jordão, SP, Brazil.

ASSIS, L. F. F. G.; HERFORT, B.; STEIGER, E.; HORITA, F. E. A. & ALBUQUERQUE, J. P. (2015a). A geographic approach for on-the-fly prioritization of social-media messages towards improving flood risk management. **4th Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)**, pages 1–12.

DOLIF, G.; ENGELBRECHT, A.; JATOBÁ, A.; DA SILVA, A. J. D.; GOMES, J. O.; BORGES, M. R.; NOBRE, C. A. & DE CARVALHO, P. V. R. (2013). Resilience and brittleness in the alerta rio system: a field study about the decision-making of forecasters. **Natural Hazards**, 65(3):1831–1847.

EDIGER, D.; JIANG, K.; RIEDY, J.; BADER, D.; CORLEY, C.; FARBER, R. & REYNOLDS, W. (2010). Massive social network analysis: Mining twitter for social good. **39th International Conference on Parallel Processing (ICPP)**, pages 583–593.

GAO, H.; BARBIER, G. & GOOLSBY, R. (2011). Harnessing the crowdsourcing power of social media for disaster relief. **IEEE Intelligent Systems**, 26(3):10–14.

HORITA, F. E.; ALBUQUERQUE, J. P.; DEGROSSI, L. C.; MENDIONDO, E. M. & UEYAMA, J. (2015). Development of a spatial decision support system for flood risk management in Brazil that combines volunteered geographic information with wireless sensor networks. **Computers & Geosciences**, 80:84–94.

MOONEY, P. & CORCORAN, P. (2011). Can Volunteered Geographic Information be a participant in eEnvironment and SDI? **In Environmental Software Systems. Frameworks of eEnvironment**, pages 115–122.

SAKAKI, T.; OKAZAKI, M. & MATSUO, Y. (2010). Earthquake shakes twitter users: Real-time event detection by social sensors. **19th International Conference on World Wide Web**, pages 851–860.

SCHNEBELE, E.; CERVONE, G., KUMAR, S. & WATERS, N. (2014). Real time estimation of the Calgary floods using limited remote

sensing data. **Water**, 6(2):381–398.

SONG, M. & KIM, M. C. (2013). Rtˆ2m: Real-time twitter trend mining system. In Proceedings of the 2013 **International Conference on Social Intelligence and Technology**, pages 64–71.

STARBIRD, K. & STAMBERGER, J. (2010). Tweak the tweet: Leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting. **17th International Conference on Information Systems for Crisis Response and Management (ISCRAM)**, pages 8-9, Seattle, USA.

VIEWEG, S.; CASTILLO, C. & IMRAN, M. (2014). Integrating social media communications into the rapid assessment of sudden onset disasters. **Social Informatics**, 8851:444–461.

VIEWEG, S.; HUGHES, A. L.; Starbird, K. & PALEN, L. (2010). Microblogging during two natural hazards events: what twitter may contribute to situational awareness.

Special Interest Group on Computer-Human Interaction Conference on Human Factors in Computing Systems, pages 1079–1088. ACM.

WAN, Z.; HONG, Y.; KHAN, S.; GOURLEY, J.; FLAMIG, Z.; KIRSCHBAUM, D. & TANG, G. (2014). A cloud-based global flood disaster community cyber-infrastructure: development and demonstration. **Environmental Modelling & Software**, 58:86–94.

ZIELINSKI, A.; MIDDLETON, S. E.; TOKARCHUK, L. & WANG, X. (2013). Social media text mining and network analysis for decision support in natural crisis management. **10th International Conference on Information Systems for Crisis Response and Management (ISCRAM)**, pages 840–845.

ZUBIAGA, A.; SPINA, D.; MARTÍNEZ, R. & FRESNO, V. (2015). Real-time classification of twitter trends. **Journal of the Association for Information Science and Technology**, 66(3):462–473.