Impromptu Crisis Mapping to Prioritize Emergency Response

Marco Avvenuti, University of Pisa

Stefano Cresci, Fabio Del Vigna, and Maurizio Tesconi, Italian National Research Council

To visualize post-emergency damage, a crisis-mapping system uses readily available semantic annotators, a machine-learning classifier to analyze relevant tweets, and interactive maps to rank extracted situational information. The system was validated against data from two recent disasters in Italy.

After a disaster, information about damage to the stricken areas is vital to the task of prioritizing responder interventions. Social media (SM) platforms can be invaluable assistants in providing this information.[1,2] Hurricane Sandy in Central and North America and the Emilia earthquake in Italy, both in 2012, and the Himalayan earthquake in Nepal in 2015 are examples of disasters that were the subject of shared information through such platforms. Ubiquitous and real-time data sharing through widespread mobile devices provided critical actionable and time-sensitive information to first responders. This information meets the needs of those helping as well as those affected, and has sparked the interest of emergency responders in envisioning innovative approaches that can merge the data collected from traditional physical sensors, such as traffic cams, with that from social sensors—crowdsourced information from social networks.[3]

Crowdsourced data is often unstructured, heterogeneous, and fragmented over a large number of messages and must be mined and aggregated to provide contextual information that emergency responders can use.[4] Crisis mapping increases situational awareness by enabling the real-time gathering and visualization of data contributed by many individuals. Crisis maps can also support resource allocation and prioritization during emergencies, when key resources are overwhelmed by the sudden increase in demand.[5] During recent disasters, civil protection agencies developed and maintained live Web-based crisis maps to help visualize and track stricken locations, assess damage, and coordinate rescue efforts.[6]

However, tools for crisis maps cannot rely on geospatial metadata to geolocate SM messages containing crisis-specific keywords. Indeed, statistics show that only 4 percent or less of SM messages carry GPS coordinates,[7] which is not enough for a meaningful crisis map. Geoparsing emergency reports—which involves

 0018-9162/16/$33.00 © 2016 IEEE

extracting mentions of known locations in the report text—can use preloaded names to help overcome this limitation but can also result in an extreme amount of data to load and manage, which tends to restrict the area that can be monitored.

To overcome these limitations, we created a general and flexible SM-based crisis-mapping system that can use Twitter messages to create a situational description on the fly without any prior knowledge of the affected area's location or the extent of damage. To better support resource prioritization during emergency response, the system also ranks identified stricken areas according to the estimated amount of damage they suffered, thus aiming to support the utilitarian goal of doing the greatest good for the greatest number.[5]

Our system works solely with tweets related to unfolding emergencies. It exploits linguistic features and a machine-learning classifier to detect mentions of damage to infrastructures or injuries, geoparses messages by relying on readily available online semantic-annotation tools and collaborative knowledge bases, and produces Web-based interactive crisis maps.

To validate our system, we tested its damage detection and geoparsing components on published datasets and assessed the accuracy of our crisis maps relative to authoritative data related to disasters that resulted in severe damage to several Italian regions: an earthquake in 2012 and a flood in 2013. Validation against these datasets, as well as against data from another earthquake, yielded an accuracy of 0.97 in detecting the most damaged areas. We believe such accuracy is sufficient for our system to support a practical crisis-mapping pipeline.

CRISIS-MAPPING CHALLENGES

The possibility of exploiting SM data for crisis mapping was first envisioned in 2010.[8,9] Since then, interest has grown in all areas related to crisis mapping—from data acquisition and management to analysis and visualization.[10] Current popular crisis-mapping platforms include Ushahidi, ESRI ArcGIS, and CrisisCommons.[11] These platforms combine automatic data acquisition, data fusion, and visualization with user participation. As such, they are hybrid crowdsensing systems: users can voluntarily load data onto the system, or the system can be configured to automatically perform data acquisition as the need arises.

CRISIS MAPPING INCREASES SITUATIONAL AWARENESS AND ENABLES FIRST RESPONDERS TO ADDRESS THE MOST DAMAGED AREAS.

Geoparsing issues

Early tools for producing crisis maps from crowdsourced data usually crawled SM for crisis-specific keywords, and geolocated messages according to metadata on GPS latitude and longitude coordinates. Geoparsing was seen as a way to increase geospatial information by looking up a number of preloaded resources, such as the Geonames and GEOnet Names global gazetteers, that contained all the possible matches between a set of toponyms (names of places) and their geographic coordinates.[6] The approach required an offline phase, during which the system would try to match its preloaded resources with information about an area with limited monitoring. The larger the monitored area, the higher the amount of data to load and manage, which created scalability issues. More importantly, incidents occurring outside the monitored area could not be mapped, which limited geoparsing's usefulness.

Another issue related to geoparsing is toponymic polysemy—the challenge of toponyms with multiple meanings, some of which might not refer to a location or might refer to multiple locations. Washington, for example, might mean the first US president or the US state or capital.

Information extraction for crisis mapping

Recent reports describe novel solutions to address geoparsing problems as well as issues in extracting situational awareness from microtexts.[12] Some crisis-mapping systems perform geoparsing by preloading geographic databases for areas at risk[6] and then generate crisis maps by comparing the volume of SM messages that mention specific locations with a statistical baseline.

Other researchers are experimenting with heuristics,[13] open source software that recognizes named entities,[14] and

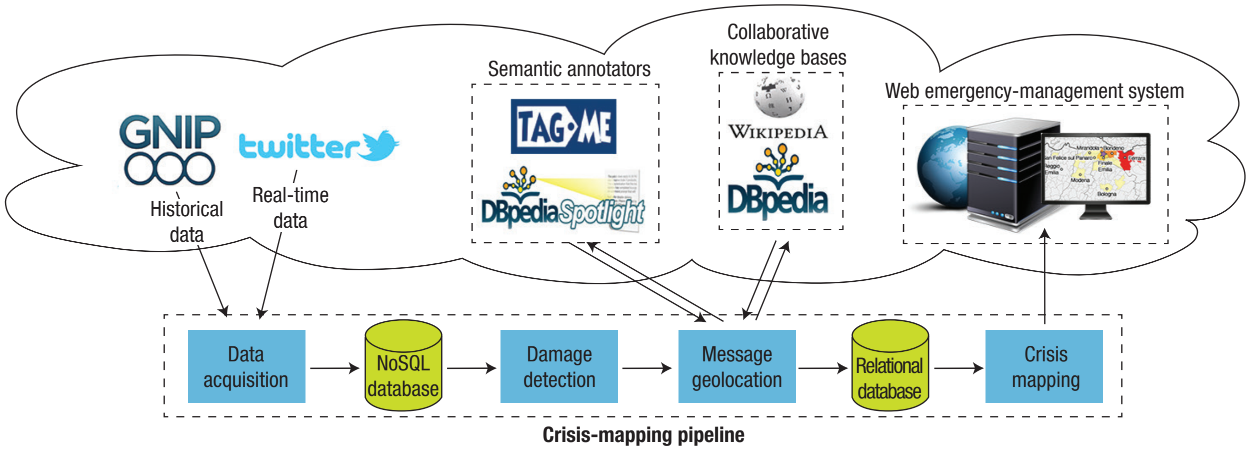

FIGURE 1. Architectural overview of our crisis-mapping system, which has four main components. Once the system acquires tweets, it stores them in a NoSQL database and then parses them for mentions of damage and location information, and finally stores them in a relational database. Using the combination of damage and location information, it then creates a crisis map, coloring areas with the most damage.

machine-learning techniques.[15] One effort applies natural-language processing techniques to detect messages carrying relevant information for situational awareness during emergencies.[16] Another research group developed a technique to extract information nuggets from tweets—self-contained information relevant to disaster response.[17]

All of these proposed solutions provide fully automatic knowledge extraction. Another solution adopts a hybrid approach, exploiting both human and machine computation to classify messages.[18] A recent survey presents an extensive review of current literature in the broad field of SM emergency management.[19]

Although these linguistic-analysis techniques are suitable in extracting relevant information from disaster-related messages, none has been used in an actual crisis-mapping task. In contrast, we evaluated our crisis-mapping system against data in two case studies of actual emergencies—an earthquake and a flood—both of which struck large areas of Italy, causing widespread damage and several deaths.

SYSTEM COMPONENTS

As Figure 1 shows, our crisis-mapping system has four main components: data acquisition, damage detection, message geolocation, and crisis mapping.

Data acquisition

The data-acquisition component exploits Twitter's streaming API (https://dev.twitter.com/streaming/overview) to enable real-time message acquisition, while data related to past disasters can be bought from Twitter resellers (https://gnip.com/historical). Both Twitter's streaming API and the resellers' historical data APIs provide access to a global set of tweets that are filtered by search keywords. The system exploits a specific set of search keywords for each disaster type, which ensures that it collects only the most relevant tweets. A flood might have an entirely different set of keywords than an earthquake, for example. In seconds, the data-acquisition component captures globally produced tweets that match the specified keywords and stores them in a NoSQL database that supports rapid storage and throughput of social media.

Damage detection

The damage-detection component draws tweets from the database and analyzes their text, with the twofold goal of discarding irrelevant tweets and labeling the relevant ones according to the presence or lack thereof of damage mentions. In our system, "damage" refers both to damage to buildings and other structures and to injuries, casualties, and missing people. In other words, damage encompasses all harmful consequences of an emergency on infrastructures and communities.

Damage detection is a multiclass classification task, which means that filtering and labeling operations are carried out in a single step by a multiclass machine-learning classifier, as opposed to a binary one. The classifier labels tweets according to the three classes:

- *damage*—tweets related to the disaster that convey damage information;
- *no damage*—tweets related to the disaster that do not convey information relevant to damage assessment; and
- *not relevant*—tweets not related to the disaster.

The machine-learning classifier performs a multilevel linguistic analysis and operates on texts that are morphosyntactically tagged and dependency-parsed by the DeSR parser. DeSR is a shift-reduce dependency parser for

TABLE 1. Dataset characteristics.

Dataset	Disaster type	Year	Users	Tweets			
				Damage	No damage	Not relevant	Total
L'Aquila	Earthquake	2009	563	312	480	270	1,062
Emilia	Earthquake	2012	2,761	507	2,141	522	3,170
Sardegna	Flood	2013	597	717	194	65	976

TABLE 2. Results of the damage-detection task on the datasets.

Dataset	Accuracy	Damage			No damage			Not relevant		
		Precision	Recall	F-M (stdev)	Precision	Recall	F-M (stdev)	Precision	Recall	F-M (stdev)
L'Aquila	0.83	0.92	0.87	0.89 (0.025)	0.81	0.87	0.84 (0.032)	0.77	0.71	0.73 (0.078)
Emilia	0.82	0.91	0.88	0.90 (0.039)	0.85	0.89	0.87 (0.016)	0.54	0.46	0.49 (0.060)
Sardegna	0.78	0.86	0.93	0.89 (0.019)	0.50	0.46	0.47 (0.099)	0.31	0.14	0.29 (0.113)

F-M (stdev): F-measure (standard deviation)

the Italian language that uses a multilayer perceptron as the learning algorithm.[20] DeSR is highly efficient, processing up to 200 sentences per second.

How it works. Given a set of features and a training corpus, the classifier creates a statistical model using the feature statistics extracted from the training corpus. It then employs the model in classifying new tweets from the data-acquisition component. We implemented the damage-detection component as a linear support vector machine (SVM) classifier using LIBSVM as the machine-learning algorithm.

We focused on a wide set of features organized into five main categories:

- raw and lexical text features (including token count, *n*-grams and *n*-gram repetition, hashtag number, and punctuation);
- morphosyntactic features, such as part-of-speech *n*-grams;
- syntactic features, which cover lexical and type dependencies;
- lexical expansion features; and
- sentiment-analysis features, including emoticons analysis, polarity *n*-grams, and polarity modifiers.

These categories largely mirror the levels of linguistic analysis automatically carried out on the text being evaluated (tokenization, lemmatization, morphosyntactic tagging, and dependency parsing).[20] The first three categories are related to the linguistic analysis of tweets. The last two are external lexical resources. Lexical-expansion features are frequently used to overcome the problem of the lexical sparsity in tweets, which typically have few words. Sentiment-polarity features are used to infer the polarity of text. Recent work has demonstrated that these features—which include but are not limited to punctuation and emoticons—actually contribute to damage assessment because post-emergency tweets and other text messages typically reflect the eyewitness's emotional state.[21]

Evaluation results. We evaluated the damage-detection component on three datasets related to different disasters that struck Italy in recent years. Table 1 shows statistics on the total collected data per disaster. To facilitate comparison with other work, we also included data from the L'Aquila earthquake, which occurred in 2009. Table 2 shows the results of our evaluation, carried out with a 10-fold cross-validation process against well-known learning-evaluation metrics, including precision, recall, and F-measure, which looks at both precision and recall to measure test accuracy. Results in Table 2's Accuracy column show that the system achieved good global accuracy, from 0.78 for Sardegna to 0.83 for L'Aquila. The scores obtained in recognizing the damage class are particularly important, and the F-measure score (fifth column in Table 2) for this class was always 0.89 or higher, which is suitable for practical application.

Message geolocation

The low number of tweets natively carrying GPS geospatial metadata[7] requires geoparsing techniques to increase the number of geolocated tweets and avoid sparse crisis maps. The message-geolocation component

TABLE 3. Results of benchmarking the TagMe geoparsing technique on two datasets.

Dataset	Precision	Recall	Accuracy	F-measure
Milan blackout	0.98	0.94	0.92	0.96
Christchurch earthquake	0.97	0.88	0.94	0.92

TABLE 4. Results of geoparsing a random sample of 1,900 tweets.

Annotator	Precision	Recall	Accuracy	F-measure
TagMe	0.88	0.80	0.86	0.84
DBpedia Spotlight	0.85	0.51	0.74	0.64

builds on readily available semantic-annotation tools and collaborative knowledge bases to disambiguate toponyms with possible multiple meanings.

How it works. Semantic annotation augments a plaintext message (such as a tweet) with pertinent references to resources in knowledge bases like Wikipedia and DBpedia. The annotated text is richer because mentions of entities are linked to the corresponding entity in the knowledge base. The message-geolocation component extends semantic annotation by checking whether the linked knowledge-base entities are actually places or locations.

Semantic annotation also alleviates geoparsing errors from toponymic polysemy. For plaintext terms that can link to multiple knowledge-base entities, semantic annotators automatically perform a disambiguating operation, returning only the most likely reference to a knowledge-base entity for every annotated term.

We used the TagMe[22] and DBpedia Spotlight[23] annotators to implement and validate the geoparsing technique used in the message-geolocation component. Tweets that made it through the damage-detection component were annotated through queries to the two annotators' APIs. Because the annotator can return multiple annotations for a single tweet, the message-geolocation component sorts returned annotations according to their confidence score. Thus, annotations that are more likely to be correct are processed first. It then geolocates a tweet using the coordinates of the first annotation that correspond to a place or location. Geographic information about annotations is fetched through a Wikipedia crawler or through SPARQL queries to DBpedia.

Evaluation results. To allow for a comparison with results reported in the literature, we benchmarked our geoparsing technique with the TagMe annotator against well-known datasets from the 2013 Milan blackout and the 2011 earthquake in Christchurch, New Zealand. The results in Table 3 are comparable to those in earlier work.[6,14]—an F-measure of 0.96 and 0.92 (fourth column).

We then used implementations of both TagMe and DBpedia Spotlight to geocode all the tweets of our datasets. Following an approach adopted by Stuart E. Middleton, Lee Middleton, and Stefano Modafferi,[6] we manually annotated a random sample of 1,900 tweets to validate the geoparsing operation. Table 4 shows the results for this sample. Our geoparsing technique achieved better results on the benchmark datasets (shown in Table 3) than on our sample datasets, which are related to emergencies in rural areas: the F-measures for the Milan blackout and Christchurch earthquake were 0.96 and 0.92, respectively, while the highest F-measure for our sample was 0.84 (TagMe). These results are evidence that crisis mapping for a rural and sparsely populated area is more difficult than it is for a highly populated metropolitan area.

However, if we report the number of tweets natively carrying GPS geospatial metadata and the number of geolocated tweets using TagMe and DBpedia Spotlight implementations, the results are quite different. As Table 5 shows, the average for all tweets in the Sardegna dataset jumps from a low of 4.6 percent with GPS to 25.7 percent with DBpedia Spotlight and 34.7 percent with TagMe for the same dataset. The weighted average percentages of geolocated tweets across the three datasets (using the number of tweets per dataset from Table 1) are 4.6 with GPS, 27.3 with DBpedia Spotlight, and 39.0 with TagMe—a considerable increase.

The weighted average percentages of geolocated tweets improve even more for tweets in the damage class (again using the number of tweets per dataset from Table 1 along with the percentages in the bottom half of Table 5): 3.4 percent with GPS, 28.1 percent with DBpedia Spotlight, and 48.6 percent with TagMe. The implication is that tweets reporting damage also report location information more often than tweets that do not report damage—a finding that further motivates the combination of damage detection and message geolocation in a crisis-mapping system.

Crisis mapping

Given a set of tweets with damage and geolocation information, the crisis-mapping component uses choropleth mapping to represent the

TABLE 5. Contribution of the proposed geoparsing technique on the number of geolocated tweets.

Dataset	GPS		DBpedia Spotlight		TagMe	
	No. of geolocated tweets	Percentage of geolocated tweets	No. of geolocated tweets	Percentage of geolocated tweets	No. of geolocated tweets	Percentage of geolocated tweets
All tweets in the datasets						
L'Aquila	0	0	285	26.8	522	49.1
Emilia	198	6.2	888	28.0	1,169	36.9
Sardegna	45	4.6	251	25.7	339	34.7
Tweets only in the damage class						
L'Aquila	0	0	91	29.2	180	57.7
Emilia	23	4.5	139	27.4	252	49.7
Sardegna	26	3.6	208	29.0	252	35.1

geographical distribution of a statistical variable and provide a clear picture of the unfolding emergency.

How it works. In choropleth mapping, subareas of a map are filled with different shades of a color, in proportion to the measurement of the variable being displayed. The technique is usually applied to depict the spatial distribution of demographic features such as population, land use, and crime diffusion, but we use it to show the spatial distribution of damage after a disaster or other emergency. The ability to apply different shades to different areas is a clear advantage over the on–off maps used in existing crisis-mapping systems. This gives responders an at-a-glance look at areas with high damage, which fits well with the need for rapid prioritization in the early stages of emergency response.

Evaluation results. To evaluate our crisis-mapping component and the whole pipeline of our system, we created crisis maps using data from the 2012 Emilia earthquake and 2013 Sardegna flood. Both emergencies affected large parts of Italy, causing widespread damage and several deaths.

Figure 2 shows the choropleth crisis maps generated by our system for the Emilia earthquake. Despite geolocating tweets in all of northern Italy, the system correctly identified the areas with the most damage. This can be highlighted by comparing the crisis map generated by our system, shown in Figure 2a, against a map derived from authoritative data about economic loss, shown in Figure 2b.

Figure 3 shows similar crisis maps for the Sardegna flood.

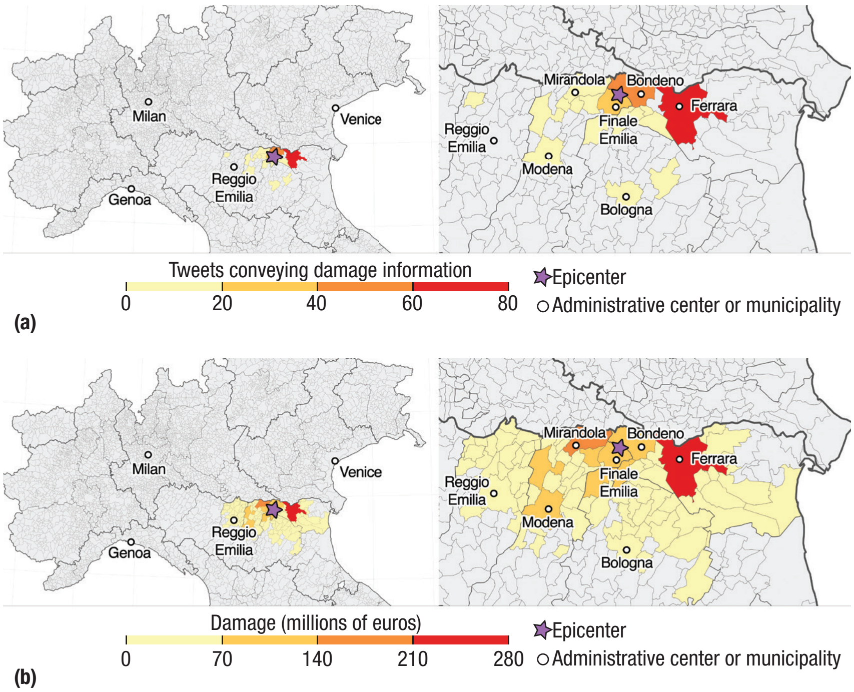

FIGURE 2. Crisis maps for the 2012 Emilia earthquake in Italy. (a) Map of damage generated by our system on the basis of tweets mentioning damage with overview (left) and details (right) and (b) map of economic loss generated from authoritative data. In (a), the system assigns a color to a municipality according to the number of damage tweets (tweets with damage information) geolocated in that municipality. Areas in which the system did not geolocate any damage tweets are gray, while areas with the most damage tweets are in orange and red. Economic loss data courtesy of the Emilia Romagna regional district (www.openricostruzione.it).

QUANTITATIVE EVALUATION OF CRISIS MAPS

The first assessment of crisis-mapping accuracy is a rough on–off comparison with authoritative maps, not accounting for damage intensity. As others have done,[6] we conducted

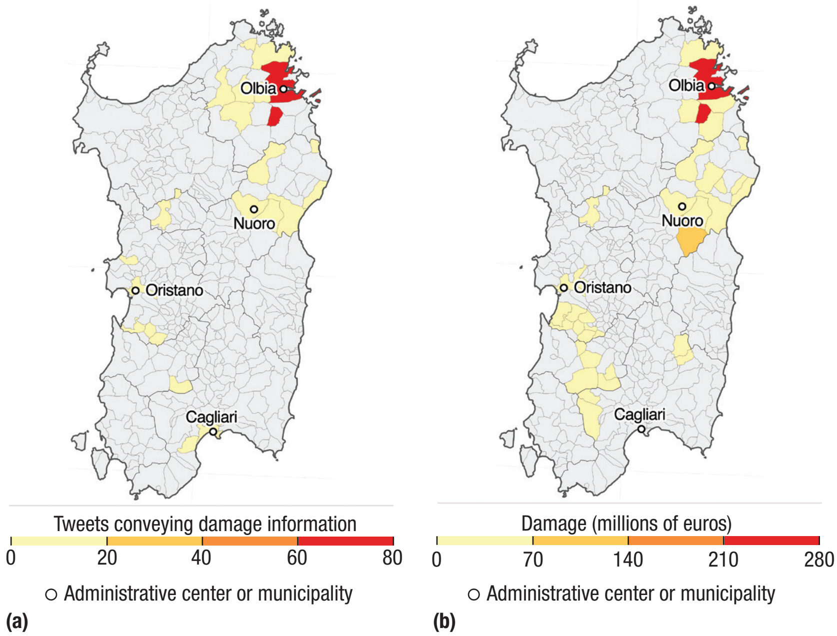

FIGURE 3. Crisis maps for the 2013 Sardegna flood in Italy. (a) Map of most damaged areas (red) and (b) map of areas with the highest economic loss. Economic loss data courtesy of the Civil Protection Agency of Sardegna regional district as reported in "Final Report on the Requirements Survey, Sardegna Regional District," Feb. 2014; www.regione.sardegna.it/documenti/1_231_20140403083152.pdf (in Italian).

a quantitative evaluation of our crisis maps as a classification task, in which the system's goal was simply to detect damaged municipalities. We used well-known metrics for machine-learning evaluation to compare crisis maps generated by our system with those generated from official data.

Our evaluation metrics included precision, recall, specificity, accuracy, the F-measure, and the Matthews correlation coefficient (MCC). The MCC is essentially an unbiased version of the F-measure with a range of values from 1 (total agreement) to –1 (total disagreement) indicating the degree to which the predicted class agrees with the real class. The class comparison checks whether a municipality with associated economic loss as reported in official data (Figures 2b and 3b) also appears as a damaged area in our crisis maps (Figures 2a and 3a).

Once we complete the identification of all damaged areas, we perform a deeper analysis by ascertaining which of the identified municipalities had the most damage.

Identifying all damaged areas

Table 6 reports the results of this comparison for the Emilia earthquake. We first consider all the municipalities of the affected region, and then repeat the comparison by considering only municipalities that suffered significant damage. For example, from data about economic loss, we identified Ferrara as having the most damage. We then looked at municipalities that suffered more than 10 percent of the damage that Ferrara had.

As Figure 2 clearly shows, our crisis-mapping system accurately identified the areas where damage actually occurred. However, the low recall values in Table 6 (first row) indicate that the system did not identify all damaged municipalities. However, removing municipalities that suffered the least damage raises the recall metric to a more acceptable value of 0.813. This pattern is an indication that most identification errors occurred for municipalities with relatively low damage, not those requiring immediate attention. We observed the same pattern in a comparison of Sardegna flood data. In Table 7, the recall metric improves from 0.410 for all affected municipalities to 0.643 for municipalities that suffered significant damage.

Overall, the results of evaluating our system's ability to detect damaged areas are comparable to those reported by Middleton, Middleton, and Modafferi,[6] which we consider to be state of the art. However, our system was able to pinpoint damage in specific areas within regions that were both rural and sparsely populated. In contrast, the other system has a fine resolution only for an emergency affecting a densely and uniformly populated area such as Manhattan, New York; the authors present results with a coarse resolution for a disaster striking a wide area, such as the state of Oklahoma. When considering only municipalities that suffered significant damage, our results were better

TABLE 6. Results of evaluating our system's ability to detect areas damaged by the Emilia earthquake.

Task	Evaluation metrics					
	Precision	Recall	Specificity	Accuracy	F-measure	MCC*
Detect all damaged areas	0.895	0.202	0.992	0.797	0.330	0.365
Detect areas that suffered significant damage	0.867	0.813	0.992	0.982	0.839	0.830

*MCC: Matthews correlation coefficient

TABLE 7. Results of evaluating our system's ability to detect areas damaged by the Sardegna flood.

Task	Evaluation metrics					
	Precision	Recall	Specificity	Accuracy	F-measure	MCC
Detect all damaged areas	0.640	0.410	0.973	0.915	0.500	0.470
Detect areas that suffered significant damage	0.500	0.643	0.973	0.960	0.563	0.545

than those reported for the other system (accuracy of 0.982 and F-measure of 0.839 for the Emilia earthquake case study).

Identifying areas with the most damage

Our system's ability to rank municipalities according to the number of tweets conveying damage information is an unprecedented feature. To evaluate it, we used typical performance metrics of ranking systems, such as search engines, to compare the ranking of damaged municipalities based on tweets with a ranking derived from authoritative sources.

In our evaluation, we viewed our system as a basic search engine that returns a list of areas and must then answer a single complex query: which areas suffered the most damage? Search engines' results are evaluated on their ability to order retrieved documents, and evaluation metrics generally include the normalized discounted cumulative gain (nDCG) and Spearman's rho correlation coefficient.

The nDCG metric compares the order of documents returned against the ideal document order and assesses ranking quality over a 0 to 1 range, with 1 representing the ideal ranking. Spearman's rho correlation coefficient assesses how well the relationship between two variables can be described using a monotonic function. Being a correlation coefficient, it ranges from –1 to 1, with values of 0 indicating no correlation.

The results of assessing our system's ability to detect the most stricken areas against official post-event data on economic losses in the affected municipalities showed considerable agreement between tweet-derived rankings and those based on the authoritative data. The nDCG metric was 0.894 for Emilia and 0.765 for Sardegna, and the Spearman's rho correlation coefficient was 0.596 for Emilia and 0.521 for Sardegna. A simple test for statistical significance resulted in a confidence score of more than 99 percent—further evidence of our system's ability to detect the most damaged areas accurately.

We have demonstrated the feasibility of generating accurate impromptu crisis maps that give responders an overview of the various damage levels in areas affected by a disaster. The results of evaluating our system's ability to detect the most damaged areas are promising, particularly taking into account the rural nature of the areas studied. The number of potential tweeters in a monitored area is always a consideration, as the performances of any SM-based system can be impaired by the lack of data from a low message rate.

Our evaluations also raised issues that require further investigation.

ABOUT THE AUTHORS

MARCO AVVENUTI is a professor of computer systems in the Department of Information Engineering at the University of Pisa. His research interests include human-centric sensing and social media analysis. Avvenuti received a PhD in information engineering from the University of Pisa. He is a member of the IEEE Computer Society. Contact him at marco.avvenuti@unipi.it.

STEFANO CRESCI is a PhD student in the Department of Information Engineering at the University of Pisa and a Research Fellow at the Institute of Informatics and Telematics of the Italian National Research Council (IIT-CNR). His research interests include social media mining and knowledge discovery. Cresci received an MSc in computer engineering and an MSc in big data analytics and social mining from the University of Pisa. Contact him at stefano.cresci@iit.cnr.it.

FABIO DEL VIGNA is a PhD student in the Department of Information Engineering at the University of Pisa and an associate researcher at IIT-CNR. His research interests include social media mining and anomaly detection. Del Vigna received an MSc in computer engineering from the University of Pisa. Contact him at fabio.delvigna@iit.cnr.it.

MAURIZIO TESCONI is a computer science researcher at IIT-CNR. His research interests include social Web mining, social network analysis, and visual analytics within the context of open source intelligence. Tesconi received a PhD in information engineering from the University of Pisa. He is a member of the permanent team of the European Laboratory on Big Data Analytics and Social Mining. Contact him at maurizio.tesconi@iit.cnr.it

Among them is the need to conduct a deeper linguistic analysis aimed at identifying the object that suffered the damage (a building, bridge, or person, for example) and the severity of that damage. The damage-detection component could then output tuples of <object, degree of damage>, thus enabling a more specific prioritized intervention.

Another area of concern is geoparsing validation, particularly the recall metric. A fine-grained assessment of geoparsing results with semantic annotators is needed, as is a comparison with other geoparsing approaches. With recent innovations of semantic annotators, it might be possible to provide more implementations of our geoparsing technique. For example, we could simultaneously exploit multiple annotators in an ensemble or voting system.

We would like to extend our analysis pipeline to languages other than Italian and assess the resulting system performance. Also, tweets contain more than just text, and other content, such as images and URLs, might also be useful. We could enhance our system to include the online analysis of images and content of linked webpages to improve damage detection and geoparsing.

Finally, given its modest requirements, our system could be easily generalized for noncrisis scenarios, such as event monitoring and nowcasting (near-future forecasting of complex processes), that require time-sensitive spatiotemporal analyses on big data streams.

ACKNOWLEDGMENTS

This research was supported in part by the PRA 2016 project "Analysis of Sensory Data: from Traditional Sensors to Social Sensors," funded by the University of Pisa, and by the "SoS—Social Sensing" project (http://socialsensing.it/en) funded by Registro.it, the Italian (.it) domain registration authority.

REFERENCES

1. B. Shneiderman, P. Pirolli, and J. Preece, "Cyberinfrastructure for Social Action on National Priorities," *Computer*, vol. 43, no. 11, 2010, pp. 20–21.
2. H. Gao, G. Barbier, and R. Goolsby, "Harnessing the Crowdsourcing Power of Social Media for Disaster Relief," *IEEE Intelligent Systems*, vol. 26, no. 3, 2011, pp. 10–14.
3. P. Meier, "New Information Technologies and Their Impact on the Humanitarian Sector," *Int'l Rev. of the Red Cross*, vol. 93, no. 884, 2011, pp. 1239–1263.
4. M. Avvenuti et al., "A Framework for Detecting Unfolding Emergencies Using Humans as Sensors," *SpringerPlus*, vol. 5, Jan. 2016; http://link.springer.com/article/10.1186%2Fs40064-016-1674-y.

5. L. Besaleva and A.C. Weaver, “Applications of Social Networks and Crowdsourcing for Disaster Management Improvement,” *Proc. IEEE Int’l Conf. Social Computing* (SocialCom 13), 2013, pp. 213–219.
6. S.E. Middleton, L. Middleton, and S. Modafferi, “Real-Time Crisis Mapping of Natural Disasters Using Social Media,” *IEEE Intelligent Systems*, vol. 29, Feb. 2014, pp. 9–17.
7. Z. Cheng, J. Caverlee, and K. Lee, “You Are Where You Tweet: A Content-Based Approach to Geolocating Twitter Users,” *Proc. 19th ACM Int’l Conf. Information and Knowledge Management* (CIKM 10), 2010, pp. 759–768.
8. R. Goolsby, “Social Media as Crisis Platform: The Future of Community Maps/Crisis Maps,” *ACM Trans. Intelligent Systems and Technology*, vol. 1, no. 1, 2010, article no. 7.
9. P. Meier, “Crisis Mapping in Action: How Open Source Software and Global Volunteer Networks Are Changing the World, One Map at a Time,” *J. Map & Geography Libraries*, vol. 8, no. 2, 2012, pp. 89–100.
10. M. Avvenuti et al., “Pulling Information from Social Media in the Aftermath of Unpredictable Disasters,” *Proc. 2nd Int’l Conf. Information and Communication Technologies for Disaster Management* (ICT-DM 15), 2015, pp. 258–264.
11. J. Bauduy, “Mapping a Crisis, One Text Message at a Time,” *Social Education*, vol. 74, no. 3, 2010, pp. 142–143.
12. S. Cresci et al., “Crisis Mapping During Natural Disasters via Text Analysis of Social Media Messages,” *Proc. Web Information Systems Eng. Conf.* (WISE 15), 2015, pp. 250–258.
13. J. Gelernter and N. Mushegian, “Geoparsing Messages from Microtext,” *Trans. GIS*, vol. 15, no. 6, 2011, pp. 753–773.
14. J. Gelernter and S. Balaji, “An Algorithm for Local Geoparsing of Microtext,” *GeoInformatica*, vol. 17, no. 4, 2013, pp. 635–667.
15. M.G. de Oliveira et al., “Producing Volunteered Geographic Information from Social Media for LBSN Improvement,” *J. Information and Data Management*, vol. 6, no. 1, 2015, pp. 81–91.
16. S. Verma et al., “Natural Language Processing to the rescue? Extracting ‘Situational Awareness’ Tweets during Mass Emergency,” *Proc. 5th AAAI Int’l Conf. Web and Social Media* (ICWSM 11), 2011; www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2834/3282.
17. M. Imran et al., “Extracting Information Nuggets from Disaster-Related Messages in Social Media,” *Proc. 10th Int’l Information Systems for Crisis Response and Management Conf.* (ISCRAM 13), 2013, pp. 791–801.
18. S. Vieweg and A. Hodges, “Rethinking Context: Leveraging Human and Machine Computation in Disaster Response,” *Computer*, vol. 47, no. 4, 2014, pp. 22–27.
19. M. Imran et al., “Processing Social Media Messages in Mass Emergency: A Survey,” *ACM Computing Surveys*, vol. 47, no. 7, 2015, article no. 67.
20. S. Cresci et al., “A Linguistically-Driven Approach to Cross-Event Damage Assessment of Natural Disasters from Social Media Messages,” *Proc 24th ACM Int’l Conf. World Wide Web, Companion Volume*, 2015, pp. 1195–1200.
21. M. Avvenuti et al., “EARS (Earthquake Alert and Report System): A Real-Time Decision Support System for Earthquake Crisis Management,” *Proc. 20th ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining* (KDD 14), 2014, pp. 1749–1758.
22. P. Ferragina and U. Scaiella, “Fast and Accurate Annotation of Short Texts with Wikipedia Pages,” *IEEE Software*, vol. 29, no. 1, 2012, pp. 70–75.
23. P.N. Mendes et al., “DBpedia Spotlight: Shedding Light on the Web of Documents,” *Proc. 7th ACM Int’l Conf. Semantic Systems* (I-Semantics 11), 2011; http://blog.semantic-web.at/wp-content/uploads/2011/09/p1_mendes.pdf.