

Towards a Data-driven Approach to Identify Crisis-Related Topics in Social Media Streams

Muhammad Imran
Qatar Computing Research Institute
Doha, Qatar
mimran@qf.org.qa

Carlos Castillo
Qatar Computing Research Institute
Doha, Qatar
chato@acm.org

ABSTRACT

While categorizing any type of user-generated content online is a challenging problem, categorizing social media messages during a crisis situation adds an additional layer of complexity, due to the volume and variability of information, and to the fact that these messages must be classified as soon as they arrive. Current approaches involve the use of automatic classification, human classification, or a mixture of both. In these types of approaches, there are several reasons to keep the number of information categories small and updated, which we examine in this article. This means at the onset of a crisis an expert must select a handful of information categories into which information will be categorized. The next step, as the crisis unfolds, is to dynamically change the initial set as new information is posted online. In this paper, we propose an effective way to dynamically extract emerging, potentially interesting, new categories from social media data.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous;
D.2.8 [Software Engineering]: Metrics—*complexity measures, performance measures*

Keywords

stream classification, text classification, information types, Social media content analysis.

1. INTRODUCTION

In disaster situations, people use social media to gather and disseminate information for a variety of purposes [4]. Previous work includes numerous attempts to break down this information into a set of categories/topics. For instance, [17] describes 28 information categories, while [3] describes 13, and [9] describes 10. There is some overlap between these categorizations, and in general we can say that at this point a large repertoire of crisis-relevant information categories is

known. However, we also know that the information contained in the messages, and the proportion of messages that will be posted in different categories can vary significantly across crises, to the extreme that in some cases a category of information that is very common in one disaster can be almost completely absent in another.

1.1 Information Variability

Changes in the proportion of information types posted in social media in different information categories can be observed even in crises associated with the same cause (e.g. the same type of natural hazard) and happening in the same location. For instance, in [17] two datasets regarding the Red River Floods in North America in 2009 and 2010 are analyzed by categorizing messages into 28 classes. The top information categories posted on Twitter in each case are shown in Table 1. While the top categories in both cases share many elements in common, information is spread into a larger number of categories in 2009 than in 2010, and the “Historical” category of information, which is among the most posted about in 2009, is not among the top ones in 2010. A similar observation can be made in the more recent events of Typhoon Pablo in 2012, and Typhoon Haiyan in 2013, both of which took place in the Philippines [13]. The results of this study, which involved a different taxonomy having only 6 classes, are also included in Table 1. We observe some common elements between 2012 and 2013 (e.g. messages about infrastructure and utilities are the smaller class in both cases). However, the order in which those elements appear in both events is different.

The two cases we describe (the floods in the United States, and the typhoons in the Philippines) constitute to a certain extent a best-case scenario for information classification. They are recurring crises in which, even if we consider that social media practices evolve over time, we can to some extent anticipate the information classes that will be most prevalent. In general, the classification of messages can be quite challenging, especially in the setting of stream classification, which we explain next.

1.2 Stream Classification

To be useful and actionable for emergency managers during a crisis or disaster, information must be delivered to them in a timely fashion. In the case of social media data (or SMS data, which can be processed in a similar way), this timeliness is achieved by using a stream processing paradigm, in which data items are processed as soon as they arrive. Stream processing is different from batch processing, in which an archive with the information to be analyzed pre-exists.

Table 1: Largest information categories shared on Twitter in two regions that experienced similar crises in two consecutive years

Red River Floods 2009	Red River Floods 2010	Typhoon Pablo 2012	Typhoon Haiyan 2013
1. Status - Hazard	1. Status - Hazard	1. Other relevant information	1. Donations and volunteering
2. Preparation	2. Advice - Information Space	2. Caution and advice	2. Sympathy and support
3. Advice - Information Space	3. Preparation	3. Sympathy and support	3. Other relevant information
4. Response - Formal	4. Response - Formal	4. Affected individuals	4. Affected individuals
5. Historical	5. Response - Community	5. Donations and volunteering	5. Caution and advice
6. Status - Infrastructure	6. Status - Infrastructure	6. Infrastructure and utilities	6. Infrastructure and utilities
+ 22 categories	+ 18 categories		

The stream processing from social media or other sources during a crisis is usually done using some combination of human labeling and automatic labeling. Human labeling cannot scale to the data volumes typical of large-scale crises, and is usually done on a sample of the input data. Automatic labeling can be done using a supervised classification approach, and to get best results, event-specific training data should be used [8]. A hybrid approach involves the combination of a stream of event-specific training data, provided by humans, which is used to train and re-train an automatic classification system [10].

In all cases of stream processing, message classification (and message filtering, which is a special case of message classification where the two classes are “accept message,” and “reject message”) requires at the onset a definition of a taxonomy or taxonomies into which the messages will be classified. Such a crisis taxonomy usually includes a set of categories (e.g. shelter, infrastructure damage, food, etc.) and a brief description for each category.

1.3 Problem statement

In all cases, while a large number of categories can be used when doing batch analysis of pre-existing data, there are multiple reasons to keep the number of categories small when doing stream processing (to classify a stream of new and unseen data). Automatic classifiers work better with fewer categories, as training data is scarce and expensive, and fewer categories mean there is more training data for each category.

Human coders are also affected by having a large number of categories, as this makes the task more cognitively taxing and thus slower and less accurate. This may increase dropout when annotators are volunteers, as is usually the case in disaster situations. A large number of categories mean some degree of training or preparation is necessary, reducing the potential pool of coders that can work in this task. Additionally, once the messages have been classified, they might be easier to consume if they are not spread into a very large number of categories. In practice, the number of categories used seem to follow a “7 plus/minus 2” rule [12], indicating that the people who design the crowdsourcing tasks assumes that the annotators cannot hold in working memory the definitions of a large number of categories while performing an annotation.

Hence, categories for this sort of classification task need to be defined carefully. A mismatch between the information categories selected for an annotation task, and the actual data being posted in social media or elsewhere, can manifest itself in many ways. First, a category can be in practice empty, wasting space in the list of categories by preventing another, more active category, from being shown to coders. Second, a category can grow too large and heterogeneous, and thus include information whose heterogeneity makes the interpretation of the labeled data difficult.

Thus, the problem can be stated as follows: how can we classify items arriving as a data stream into a small number of categories, if we cannot anticipate exactly which will be the most frequent categories?

Paper organization. Next section reports on related work. In section 3, we present the proposed solution and formal definitions of various concepts used in the paper. Experiments details including dataset description and final results are presented in the section 4. Finally, we concluded the paper and provide details of future work in section 5.

2. RELATED WORK

Crisis situations, particularly those with no prior warning, require rapid assessment of the available information to make timely decisions. Information posted at the time of crisis on social media platforms such as Twitter, can be very useful in disaster response, if processed timely and effectively [5, 16]. Many approaches based on human annotation, supervised learning, and unsupervised learning techniques have been proposed to process social media data (for a survey see e.g. [6]). However, in this paper, we employ unsupervised learning techniques to improve crowdsourcing-based and supervised learning-based systems by finding latent categories in social media data streams.

2.1 Supervised learning approaches

In the domain of supervised classification of online streams, systems learn to classify unseen documents using a set of predefined classes/categories with a handful of training examples [1, 7]. Over time, researchers who study disasters, learned about suitable categories that are useful for disaster response. However, crisis data on social media exhibits diverse information categories [13, 15], and some categories may be unanticipated, thus making it difficult for even a reasonably well trained system to identify useful informa-

tion. In [18], authors tried to solve the problem of concept/category drift, which is a related problem to ours. They presented a framework based on active learning strategies to deal with emerging/drifted concepts in the context of data streams by acquiring fresh labels to update already trained models. However, their approach cannot be applied to cases where completely new concepts/categories emerge. For this purpose, in order to facilitate the process of choosing right set of categories at right time, in this paper we analyze the use of approaches that can aid supervised learning systems without using any additional training data.

2.2 Unsupervised approaches

Under the umbrella of unsupervised approaches, the topic modeling approach provides a way to analyze text documents to discover abstract latent topics in them. For instance, Latent Dirichlet allocation (LDA) is a generative probabilistic topic modeling technique that represents documents as a mixture of topics where each topic is formed of words with certain probability [2]. In the field of crisis informatics, [11] were among the first to study the applications of topic modeling in disaster-related twitter data, but not many other papers have applied this technique to this domain since then. However, the topic modeling technique is applied in other related domains. For instance, [14] identified health-related topics on Twitter using the LDA topic modeling technique. In this paper, we employ the LDA topic modeling approach to generate candidate categories.

3. PROPOSED SOLUTION

A general solution to this problem is to dynamically change the set of active categories being used for labeling as the crisis progresses. However, the method we discuss in this paper takes a more specific approach, combining top-down, and bottom-up aspects. The top-down aspect is the initial set of information categories proposed by experts, the bottom-up aspect is the discovery of relevant and prevalent categories in the “miscellaneous” category. The discovery of new categories in the “miscellaneous” category can take many forms. We can envision a very general setting in which an automatic system modifies the list of classification categories, or assists users in the creation, merging, splitting, or deletion of categories as a crisis progresses.

As a first step in that direction, in this paper we propose a simple yet effective method:

1. An expert defines the types of information (i.e. categories) that are of interest to him/her.
2. Out of these information types, a small initial set of categories is selected. These types are known to be present in social media in similar disasters and/or in the same region.
3. Messages are categorized by human coders and/or automatic means into these categories plus an extra “Miscellaneous” or “None of the above” category.
4. Categories of interest that are frequent in the data, but not in the initial set of categories, are identified inside the “Miscellaneous” category, and added to the set of categories for further labeling.

Steps 3 and 4 can be repeated until no new categories of interest are identified, or until the number of categories

grows too large to continue expanding. The main aspects of this method that need to be specified are: (i) how to select and generate candidate categories; and (ii) how to select among those categories, one that might be relevant. The next sections expand these aspects in detail.

3.1 Candidate Generation

In principle, a candidate is any subset of messages in the “Miscellaneous” category. These subsets are not necessarily disjoint. The LDA (Latent Dirichlet Allocation [2]) method is a popular method used for finding latent topics in a collection of documents. Concretely, we propose to apply LDA to the “Miscellaneous” category, and consider how each of the topics discovered in this category by LDA is a potential candidate for a new category to be added to the active set of categories.

The input to LDA is a set D containing n documents, in this case all the messages in the “Miscellaneous” category, plus a number m indicating how many topics need to be created from the data. The output of LDA is a $n \times m$ matrix in which $cell(i, j)$ indicates the extent to which document i corresponds to topic j according to the LDA algorithm.

3.2 Evaluating a Candidate

After a number of candidates have been generated, they need to be sorted according to some criteria, to reduce the workload of the expert who eventually decides whether to expand the list of categories or not.

We propose the following criteria to decide what constitutes a good candidate category: *volume*, *novelty*, *intra-similarity*, *inter-similarity*, and *cohesiveness*.

- **Volume.** A candidate category should include a relatively large number of messages, or at least not be too small in comparison with the existing categories. Adding a category for which few messages exist would not change the annotations except by making the annotation task more difficult for annotators (by having one more category to think about).
- **Novelty.** A candidate category must not overlap or be too similar to the existing categories. Overlapping categories are a source of confusion for annotators, reduce the effectiveness of annotated data for training automatic classifiers, and wastes a valuable “slot” in the list of categories that could be used by a category that is actually new.
- **Cohesiveness (based on intra- and inter-similarity).** A candidate category should be cohesive, in the sense of describing a set of messages that are strongly related to each other. A cohesive category is easier to describe than a more disperse or amorphous category (i.e. a category containing messages that are not clearly related to each other).

Each of these aspects can be quantified in practice. Let D be the set of messages collected with respect to a crisis, with a similarity metric $dist(a, b)$ for two documents $a, b \in D$ indicating the textual distance between them (e.g. cosine similarity, or Jaccard coefficient).

Let T a set of topics $T = \{T_1, T_2, \dots, T_n\} \cup \{T_o\}$ where T_o represents the “Miscellaneous” category/topic. Each element $T_i \in T$ is a subset of documents (in this case a set of messages), $T_i \subseteq D$.

The idea is to produce a new set of topics T' so that $T' = \{T_1, T_2, \dots, T_n, T_o^*\} \cup \{T_o \setminus T_o^*\}$. The category T_o^* is chosen from a set of candidate categories C_1 through C_k , which are a set of sub-topics found in T_o using the candidate generation step described above, which is based on LDA.

The volume of a candidate C_i is simply $|C_i|$ (note that $C_i \subseteq D$), the number of documents in it.

The novelty of a candidate C_i is $Z - \minDist(a, b)$, where $a \in C_i, b \in T \setminus T_o$ and $Z = \maxDist(C_i, T \setminus T_o)$ —used as a normalizing factor—is the maximum distance between the candidates and the existing topics. The $\minDist(a, b)$ corresponds to the minimum distance between a document in C_i and one document categorized into one of the existing categories $T_i, i = 1 \dots n$.

The cohesiveness of a candidate topic is defined using metrics borrowed from cluster analysis as $\text{intra}(C_i)/\text{inter}(C_i, T_o \setminus C_i)$. Where $\text{intra}(C_i)$ is the intra-topic distance in category C_i , defined as the average of $\text{dist}(a, b)$ for $a, b \in C_i$, and $\text{inter}(C_i, T_o \setminus C_i)$ is the inter-topic distance of category C_i and the rest of the topics found in T_o , defined as the average of $\text{dist}(a, b)$ for $a \in C_i, b \in T_o \setminus C_i$. A good candidate has a small intra-topic distance (meaning elements inside the topic are similar to each other) and a large inter-topic distance (meaning elements inside the topic are different from elements outside the topic).

In order to test our hypotheses regarding the relationship between these metrics (volume, novelty, and cohesiveness) and the potential usefulness of these categories for emergency response, in the next section we describe an experimental validation with two experts in social media analysis domain during emergencies.

4. EXPERIMENTAL TESTING

We test the proposed solution to seek an answer to the following question: to what extent do the volume, novelty, and cohesiveness of a candidate category match what an expert would consider useful?

In this section we describe our experimental setting, including the dataset, the candidate generation and ranking methods, and the expert evaluations.

4.1 Dataset and Generation of Candidates

We use the CrisisLexT26 dataset [13], which corresponds to social media messages from Twitter posted during 26 different crises that took place in 2012 and 2013. We selected 17 crises occurring in countries with a large English-speaking population, as this is the language that our experts are most familiar with. Table 2 lists the crises and the prevalence of non-other as well as other categories. In the dataset, each crisis corresponds to 1,000 tweets annotated along 3 dimensions: informativeness, information type, and information source.

We use the “Information Type” annotation, which classifies tweets into the following categories:

- A. Affected individuals: deaths, injuries, missing, found, or displaced people, and/or personal updates.
- B. Infrastructure and utilities: buildings, roads, utilities/services that are damaged, interrupted, restored or operational.

Table 2: Datasets details including crisis name, year, % of tweets in other and non-other categories.

Crisis name	Year	% tweets in A-E	% tweets in Z (other)
Colorado wildfires	2012	55%	45%
Philippines floods	2012	93%	7%
Typhoon Pablo	2012	84%	16%
Alberta floods	2013	82%	18%
Australia bushfire	2013	60%	40%
Bohol earthquake	2013	84%	16%
Boston bombings	2013	59%	41%
Colorado floods	2013	76%	24%
Glasgow helicopter crash	2013	74%	26%
LA airport shootings	2013	71%	29%
Manila floods	2013	89%	11%
NY train crash	2013	54%	46%
Queensland floods	2013	70%	30%
Savar building collapse	2013	67%	33%
Singapore haze	2013	64%	36%
Typhoon Yolanda	2013	87%	13%
West Texas explosion	2013	77%	23%

- C. Donations and volunteering: needs, requests, or offers of money, blood, shelter, supplies, and/or services by volunteers or professionals.
- D. Caution and advice: warnings issued or lifted, guidance and tips.
- E. Sympathy and emotional support: thoughts, prayers, gratitude, sadness, etc.
- Z. Other useful information not covered by any of the above categories.

To expand this list, candidate categories are generated by applying LDA over the messages in the “Z. Other useful information” category, setting the number of topics to 5. This number is set heuristically, as we observe that slightly larger values (from 5 to 10) do not yield significantly different results, and smaller values tend to yield very general categories. Given that LDA generates a set of scores for all categories for every document, we consider that a message belongs to a category if this score is larger than 0.06. This value is also set heuristically after experimenting with values in the [0.03, 0.1] range.

4.2 Candidate Annotation

We recruited two experts in the classification of social media messages sent during disasters. Both hold leadership positions in digital humanitarian organizations that specialize in the handling of social media messages during disasters. We first provided context by explaining our research objectives, questions, and dataset.

Next, for each of the 17 crises we presented them 5 candidate categories indicating:

- The name of the crisis event (e.g. “Bohol Earthquake 2013”).
- The top 20 words with which LDA represents the corresponding topic.
- 3 tweets selected at random from those having at least a score of 0.05 for the category.

Topic words (20)	Tweet-1	Tweet-2	Tweet-3
nsw nswfire climat chang fire auspol abbott say link peopl toni abcnew tonyabbottmhr fight minist emerg disast servic condit adambandt	RT @ahleach: "PMOnAir: Adam Bandt is a disgrace. I can't think for a more inappropriate tweet. #nswfires #sydfires #pmlive http://t.co/rS...	RT @graham_dj: Fewer people can now claim disaster payments under Federal Govt changes made yesterday. http://t.co/XV9mVyz3H6 #NSWFIRE @To...	RT @ch150ch: Abbott (who claims exps for runs, swims & bike rides) changes rules so less people to get #nswfires disaster payment http://t....
fire australia bush militari rage blame start australian use caus investig news one record threat nsw eastern exercis year rain	@THEKINGColucci Australian military probes fire link: Australia's military is investigating whether a training exercise usi... @deadmau5	@deadmau5 Australia Fires: Military Investigates Blaze: The military reveals it was conducting an explosives training... @THEKINGColucci	Australian military probes fire link: Australia's military is investigating whether a training exercise using ... http://t.co/Vh1vWNhK5w
fire australia south new wale burn blaze bushfir imag firefight bush across wildfir get right look fear merg could two	Fears Australia blazes could merge: Bushfires in New South Wales could merge to form a massive blaze, the fire... http://t.co/yvS1XJfGuz	RT @NASA: A bad fire day in New South Wales, Australia, yesterday. Look at this image from our Aqua satellite: http://t.co/VOCp0b3gZ6 @NASA...	Australia fires: Firefighters fought scores of bush fires blazing in Australia's New South Wales early Friday. http://t.co/YCfKPuHBM
nswfire nswrfs near live today crew abcnew help updat rfs water call rd ash thetodayshow work thank state media back	RT @ABCCentralWest: RFS 1105 update: Winds are picking up. Some helicopter water bombing has been suspended due to high winds. #nswfires	Thanks to #nswrfs 4 doing an incredible job of keeping the Hume Hwy open today. Back to the #2UE newsroom 2morrow 4 more #nswfires coverage	RT @TripleMSydney: Pull back, pull back, pull back. Say Again over... The chilling footage from the #nswfires frontline @NSWRFS http://t....
sydney nswfire fire bushfir photo smoke blue amaz sydneyfir mountain sky smh road orang us katoomba put danger turn big	Sydney CBD skyline, shrouded in smoke, blue skies reflected in the glass of the buildings. #nswfires #bushfires http://t.co/u1ta5Cn258	RT @VaZHANG: Bushfire smoke from West to East over Sydney and the CBD in the background @smh #nswfires #bushfires #sydney http://t.co/9uN1f...	Amazing and terrifying RT @workmanallice: Amazing photo from a friend in Katoomba on the road to Blackheath. #nswfires http://t.co/17TQT40AF

Figure 1: LDA generated topics using the 2013 Australia bushfire dataset. Each row represents a topic with 20 words and 3 tweets.

Figure 1 shows 5 LDA generated topics, each consists of 20 words and 3 tweets from the 2013 Australia bushfire dataset.

We asked both experts to independently indicate on a scale from 1 to 5 how useful they considered a discovered category. To account for usefulness, they were instructed to look for “categories that do not overlap with the existing ones” and “categories that are well-defined and cohesive.” In total each expert investigated 85 tasks (i.e., 5 candidate categories per crisis) and the annotation process took approximately 3 to 4 hours for each expert.

This is a subjective task, and inter-annotator agreement was relatively low. We measure inter-annotator agreement by first converting the scores into a binary variable, considering as “not useful” all the cases in which an annotator gave a score lower than 3 (on a 1 to 5 scale) and as “useful” the cases in which an annotator gave a score higher than 3. Under this mapping, both experts agreed on 63% of the labels and Cohen’s Kappa measure was 0.26, which is customarily interpreted as a fair level of agreement.

4.3 Results

We evaluated the results by comparing the expert annotations with each of the metrics we derived automatically from each category. This comparison was done by the following procedure for each of the crisis examined and for each of the metrics (volume, novelty, intra-similarity, inter-similarity, and cohesiveness):

1. Averaging the scores of the two experts for each candidate topic.
2. Grouping the topics into those obtaining an average score of 2.5 or less (considered “bad” topics) and those obtaining an average score of 3.5 or more (considered “good” topics). A crisis is not considered for evaluation (next step), If all of its topics receive an average score either below or above 3.0.
3. Comparing the average value of the metric for the “good” topics with the value of the “bad” topics. If the value of the metric is larger for the “good” topics than for the “bad” topics, this is count as a hit, otherwise it is a miss.

Table 3: Evaluation of metrics for evaluating a candidate.

Metric	Hits	Misses
Volume Hit = higher volume	7	5
Novelty Hit = less similar to existing categories	8	4
Intra-similarity Hit = high similarity among in-topic documents	8	4
Inter-similarity Hit = less similar between in-topic and off-topic documents	5	7
Cohesiveness Hit = highly cohesive	8	4

Table 3 summarizes the experimental results obtained by this procedure. In 5 crises, all the topics were either below, or above, the set threshold value described in step 2. For the remaining set, we can observe that novelty, intra-similarity, and cohesiveness are useful in identifying good topics, in the sense that when a topic is “good,” it is twice as likely to have higher values along this metric than when it is “bad.” An exception is the “inter-similarity” measure, for which the result is the opposite as the one anticipated, and even “good” topics are not so separable from the rest of the “Other” category.

5. CONCLUSION

In the domain of online classification of data streams, an emerging challenge is to keep the categories used for classification up-to-date. While online supervised learning methods exist, in the domain of crisis informatics these methods must be aware of the particularities of this domain, in particular, they should be able to take input from human experts.

We have described an approach that has manual, top-down, elements, as well as automatic, bottom-up elements. This approach allows an expert to provide existing informa-

tion categories of interest, and to discover new information categories. Most importantly, these categories must be novel (to help labelers and automatic systems reduce ambiguity in classification), and cohesive (to ensure messages inside each category are related to each other). In these cases, having a large novelty or large intra-similarity or high cohesion means the chances of being a good candidate topic are double of those of being a small candidate topic.

Future work. The method we have presented describes a way of finding sub-topics inside the “Miscellaneous” category, but does not decide when it is correct to do so. We evaluated candidate categories to understand important characteristics that can be used in ranking. However, in future work, we will be working on developing criteria to *rank candidate categories*. We would also like to apply some criteria by which, if no candidate surpasses a certain score, no suggestion is done to extend the “Miscellaneous” category. This would provide a stopping criterion for the iterative refinement process described in this paper. Additionally, this method can be extended into a more general one in which any category can be expanded, and categories can be merged and removed dynamically (but with assistance from an expert) as the crisis progress.

6. REFERENCES

- [1] Z. Ashktorab, C. Brown, M. Nandi, and A. Culotta. Tweedr: Mining twitter to inform disaster response, 2014.
- [2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.
- [3] A. Bruns, J. E. Burgess, K. Crawford, and F. Shaw. #qldfloods and @qpsmedia: Crisis communication on twitter in the 2011 south east queensland floods, 2012.
- [4] J. D. Fraustino, B. Liu, and Y. Jin. *Social media use during disasters: A review of the knowledge base and gaps*. National Consortium for the Study of Terrorism and Responses to Terrorism, 2012.
- [5] A. L. Hughes and L. Palen. Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management*, 6(3):248–260, 2009.
- [6] M. Imran, C. Castillo, F. Diaz, and S. Vieweg. Processing social media messages in mass emergency: A survey. *arXiv preprint arXiv:1407.7071*, 2014.
- [7] M. Imran, C. Castillo, J. Lucas, P. Meier, and S. Vieweg. Aidr: Artificial intelligence for disaster response. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, pages 159–162. International World Wide Web Conferences Steering Committee, 2014.
- [8] M. Imran, C. Castillo, J. Lucas, M. Patrick, and J. Rogstadius. Coordinating human and machine intelligence to classify microblog communications in crises. *Proc. of ISCRAM*, 2014.
- [9] M. Imran, S. M. Elbassuoni, C. Castillo, F. Diaz, and P. Meier. Extracting information nuggets from disaster-related messages in social media. *Proc. of ISCRAM, Baden-Baden, Germany*, 2013.
- [10] M. Imran, I. Lykourentzou, and C. Castillo. Engineering crowdsourced stream processing systems. *arXiv preprint arXiv:1310.5463*, 2013.
- [11] K. Kireyev, L. Palen, and K. Anderson. Applications of topics models to analysis of disaster-related twitter data. In *NIPS Workshop on Applications for Topic Models: Text and Beyond*, volume 1, 2009.
- [12] G. A. Miller. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956.
- [13] A. Olteanu, S. Vieweg, and C. Castillo. What to expect when the unexpected happens: Social media communications across crises. In *In Proc. of 18th ACM Computer Supported Cooperative Work and Social Computing (CSCW’15)*, 2015.
- [14] K. W. Prier, M. S. Smith, C. Giraud-Carrier, and C. L. Hanson. Identifying health-related topics on twitter. In *Social computing, behavioral-cultural modeling and prediction*, pages 18–25. Springer, 2011.
- [15] S. Roy Chowdhury, M. Imran, M. R. Asghar, S. Amer-Yahia, and C. Castillo. Tweet4act: Using incident-specific profiles for classifying crisis-related messages. In *10th International ISCRAM Conference*, 2013.
- [16] K. Starbird, L. Palen, A. L. Hughes, and S. Vieweg. Chatter on the red: what hazards threat reveals about the social life of microblogged information. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 241–250. ACM, 2010.
- [17] S. E. Vieweg. Situational awareness in mass emergency: A behavioral and linguistic analysis of microblogged communications, 2012.
- [18] I. Zliobaite, A. Bifet, G. Holmes, and B. Pfahringer. Moa concept drift active learning strategies for streaming data. In *WAPA*, pages 48–55, 2011.