# Extracting Situational Awareness from Microblogs during Disaster Events

Anirban Sen*, Koustav Rudra† and Saptarshi Ghosh*‡
*Department of Computer Science and Technology
Indian Institute of Engineering Science and Technology Shibpur, Howrah - 711103, India
†Department of Computer Science and Engineering
Indian Institute of Technology Kharagpur, Kharagpur – 721302, India
‡Max Planck Institute for Software Systems, Kaiserslautern–Saarbruecken, Germany

***Abstract*—Microblogging sites such as Twitter and Weibo are increasingly being used to enhance situational awareness during various natural and man-made disaster events such as floods, earthquakes, and bomb blasts. During any such event, thousands of microblogs (tweets) are posted in short intervals of time. Typically, only a small fraction of these tweets contribute to situational awareness, while the majority merely reflect the sentiment or opinion of people. Real-time extraction of tweets that contribute to situational awareness is especially important for relief operations when time is critical. However, automatically differentiating such tweets from those that reflect opinion / sentiment is a non-trivial challenge, mainly because of the very small size of tweets and the informal way in which tweets are written (frequent use of emoticons, abbreviations, and so on). This study applies Natural Language Processing (NLP) techniques to address this challenge. We extract low-level syntactic features from the text of tweets, such as the presence of specific types of words and parts-of-speech, to develop a classifier to distinguish between tweets which contribute to situational awareness and tweets which do not. Experiments over tweets related to four diverse disaster events show that the proposed features identify situational awareness tweets with significantly higher accuracy than classifiers based on standard bag-of-words models.**

## I. Introduction

Microblogging sites such as Twitter and Weibo have become important sources of real-time information on the Web. In particular, microblogging sites are serving as useful sources of information during disaster events [1]–[3], including natural disasters (e.g., hurricanes, earthquakes) as well as man-made disasters (e.g., riots and bomb-blasts). Recent research [4]–[7] has shown the value of microblogging sites in enhancing *situational awareness* [8] during disasters, i.e., for gaining a high-level understanding of the situation and thus informing decision-making processes (e.g., coordinating disaster relief efforts).

However, among the thousands of microblogs (commonly called 'tweets') posted during a disaster event, only a small fraction contribute to situational awareness (SA), while the majority of the tweets merely reflect the opinion of the masses (e.g., sympathizing with the victims affected by the disaster). While humans are the best judges of what information contributes to SA [1], [2], tweets are posted so rapidly during large-scale disasters that it is infeasible for humans to even comprehend the tweet stream in real-time, leave alone identifying tweets which contribute to SA. Thus, use of human judgement to identify SA is feasible only during post-hoc analyses, and not during (or immediately after) the actual disaster, the time when it is most critical to gain SA. Hence, it is necessary to develop *automated* mechanisms to identify tweets which contribute to SA.[1] In this scenario, Natural Language Processing (NLP) based techniques can be used to extract the tweets which contribute to SA [5], [9], [10]. The principal challenge for NLP-based techniques is the informal way in which tweets are written – because of the size restriction (at most 140 characters), tweets often contain abbreviations, colloquial languages, and so on. Hence NLP techniques developed for formal English text often do not work as well for tweets [1], [11].

The objective of this study is to develop a NLP-based classification scheme that can distinguish and extract SA tweets from tweet streams posted during such events. There have been prior NLP-based efforts to extract particular types of situational information from tweet streams during the 2011 Japan earthquake, such as, problem reports and aid messages [10], and safety information of individuals [9]. Since these studies focused on a particular event, they could use event-specific resources, e.g., common last names of individuals and location names in Japan. However, development of mechanisms to identify SA tweets in general, i.e., irrespective of a specific type of information or a specific disaster event, still remains a challenge. To our knowledge, only one prior study [5] has used NLP techniques to distinguish between SA and non-SA tweets in general. [5] used the unigrams and parts-of-speech tags contained in the tweets as features for classifiers, to differentiate between SA and non-SA tweets across several dimensions, such as their subjective / objective nature, formal / informal register, and their personal / impersonal nature.

The present study also uses NLP techniques to develop classifiers for distinguishing SA and non-SA tweets. However, rather than estimating abstract characteristics such as subjectivity, linguistic register and personal / impersonal tone of tweets (as attempted in [5]), we rely on low-level syntactic features which can be directly extracted from the tweets. The

[1] Tweets which contribute to situational awareness are henceforth referred to as *SA tweets*, and tweets which do not as *non-SA tweets*.

motivation for our choice is explained in Section II, and the proposed features are described in Section IV. Experiments with tweets posted during four recent and diverse disaster events (described in Section III) demonstrate that this approach can distinguish between SA and non-SA tweets with much higher accuracy than classifiers trained on standard bag-of-words features (detailed in Section V).

## II. Background and Motivation

This section briefly describes the challenges in differentiating SA tweets (which contribute to situational awareness) from non-SA tweets, and the recent advances in NLP which enables us to address the challenges.

It was empirically observed by Verma *et al.* [5] that SA tweets are likely to be written in a more objective, impersonal, and formal linguistic style compared to non-SA tweets; hence, the identification of these characteristics should be useful for distinguishing SA tweets from non-SA ones. Prior NLP research has already developed mechanisms to measure characteristics such as the subjectivity [12] and formal / informal nature of English text [13], [14]. However, since tweets often contain incomplete sentences, abbreviations, and emoticons, and are seldom written in a grammatically coherent way (primarily because their size is restricted to 140 characters), tools designed for normal English text may not perform well in case of tweets [11], [15]. Hence, Verma *et al.* [5] used a bag-of-words model (i.e., the unigrams and parts-of-speech contained in the tweets were used as features) to develop *domain-specific* subjectivity and linguistic register classifiers.

However, this methodology has an obvious disadvantage – the bag-of-words classifiers are heavily dependent upon the vocabulary of the specific event on which the classifier is trained, and they perform well only when the test event is very similar to the training event (in-domain classification). As Verma *et al.* [5] themselves reported, classification performance is significantly degraded in the *more practical* cross-domain classification, i.e., when the classifier is trained on tweets about past events, and then used to classify tweets related to newly occurring events. This practical limitation of bag-of-words classifiers motivated us to develop better mechanisms for SA vs. non-SA tweet classification.

The present study adopts a similar approach to [5] and extracts linguistic features from the tweets to train a SA vs. non-SA classifier. However, there are important differences between the features considered in the present study and those in [5]. Rather than attempting to estimate abstract characteristics such as subjectivity, linguistic register and personal / impersonal tone of tweets (as done in [5]), we focus on extracting *lower-level syntactic features* which can be directly measured from the tweet text (e.g., presence of some specific parts-of-speech in the tweet). The use of syntactic features looks promising since [16] recently demonstrated that it is possible to use syntactic features to characterize the linguistic style of tweets. Also, the present work builds upon recent advancements in NLP of microblogs – such as, POS taggers [17] and subjectivity lexicons [18] specifically developed for tweets – to extract simple syntactic features which can abstract more complex notions such as subjectivity and formal / informal register of tweets.

## III. Dataset

This section describes our dataset of tweets related to disasters, and how we establish a 'gold standard' (through human judgement) for whether the tweets contribute to SA.

### A. Disaster events

We considered the following four disaster events for the present study:

1) **SHshoot** – an assailant killed 26 people, including 20 children, at the Sandy Hook elementary school in Connecticut, USA [19]
2) **TBopha** – a strong cyclone code-named Typhoon Bopha / Pablo hit Philippines [20]
3) **HBlast** – two bomb blasts in the city of Hyderabad, India [21]
4) **UFlood** – devastating floods and landslides in the Uttaranchal state of India [22].

Note that the events are widely varied, including both man-made and natural disasters in various regions of the world. Hence, the vocabulary / linguistic style in the tweets related to the various events can be expected to be diverse as well.

We collected relevant tweets posted during each event through the Twitter API [23] using keyword-based matching; for instance, the keywords 'Sandy Hook', 'school' and 'shooting' were used to identify tweets related to the SHshoot event.

Specifically for this study, we consider 500 randomly selected English language tweets related to each event. Following the approach of [11], we considered a tweet to be in English if at least half of the words in the tweet appear in the `wamerican-small` English dictionary. Also, we ignored duplicate tweets [24], such as copies of the same tweet posted or retweeted by different users. However, we did *not* attempt to remove tweets which gave the same semantic information using different linguistic styles, to verify if the proposed scheme can identify SA-tweets irrespective of the specific linguistic style.

### B. Establishing gold standard

The 500 tweets for each event were annotated by three human volunteers, who are regular users of Twitter and have good knowledge of English. Before the annotation task, the volunteers were shown examples of SA and non-SA tweets identified in prior studies [1], [2], [5]. The three volunteers then independently annotated each tweet as SA or non-SA based on whether it contributes to situational awareness. There was unanimous agreement for 76% of the tweets; for the rest, we considered the majority verdict as the Gold Standard.

Out of the 500 tweets for each of the events, 168, 180, 130 and 192 tweets were judged to be SA tweets for the SHshoot, TBopha, HBlast and UFlood datasets, respectively. For each event, we then randomly selected an equal number of tweets from among the ones which were annotated as non-SA. Thus, our dataset consists of 1040 tweets in total, of which half are annotated as SA tweets and the other half as non-SA tweets.

| Type | Event | Tweet text |
|---|---|---|
| **SA tweets (contributing to situational awareness)** | | |
| Situational information | SHshoot | police entering connecticut school where shooting reported; evacuations underway |
| | UFlood | Army commandos locate 1,000 stranded in mountains btw Rambara-Gaurikund. Many critical. IAF to airdrop rations and meds asap #uttarakhand |
| | TBohpa | intense rains and thunderstorms in malaybalay, bukidnon as #pabloph nears. no power in area. |
| Help relief operations | HBlast | #HyderabadBlasts Dilshuknagar Hospitals: Sigma 40-67120218; Good Life 49640328 |
| | UFlood | #Uttarakhand Flash Flood Helpline numbers are: 0135-2710335, 2710233 |
| | TBopha | ndrrmc urges residents near cagayan de oro river to move to higher ground amid onslaught of typhoon pablo |
| **Non-SA tweets** | | |
| Sentiment | SHshoot | My prayers go out to the victims of the shooting at #SandyHook Elementary :( |
| | UFlood | I salute the armed forces & all those who are courageously dedicating themselves to relief work across Uttarakhand. |
| Opinion | SHshoot | Sandy Hook shooting is a clear indication that the US should revise their gun laws to ensure no more mass shootings. |
| Event analysis | UFlood | #Deforestation in #Uttarakhand aggravated #flood impacts. Map showing how much forestland diverted http://t.co/A4m06IvCDg |
| | HBlast | #HyderabadBlasts: Police suspect one of the bombs may have been kept on a motorcycle; the other in a tiffin box. |
| Charities | SHshoot | r.i.p to all of the connecticut shooting victims. for every rt this gets, we will donate $2 to the school and victims |
| | TBopha | want to help typhoon #pabloph victims? you can also donate to @philredcross via sms |

**TABLE I: Examples of various types of SA tweets (which contribute to situational awareness) and non-SA tweets.**

### C. *Types of SA and non-SA tweets*

During the above annotation process, the volunteers observed the different types of information contained in SA and non-SA tweets, some examples of which are shown in Table I. The following broad classes of tweets were observed.[2]

**Types of SA tweets:** Tweets which can be used to gauge the current situation and to inform decision-making processes, are primarily of the following two types: (i) *Situational information* – updates such as the number of casualties, and the current situation in various regions affected by the disaster, and (ii) *Helping relief operations* – information that can immediately help relief operations, e.g., phone numbers of nearby hospitals.

**Types of non-SA tweets:** Non-SA tweets are generally of the following types: (i) *Sentiment tweets* – sympathizing with the victims, or praising / criticizing the relief operations, (ii) *Opinion* – opinion / suggestions on issues such as how relief operations can be improved, how similar tragedies can be prevented in future, (iii) *Event analysis* – post-analysis of how and why the disaster occurred, findings from police investigation in case of man-made emergencies, and (iv) *Organizing charities* – tweets related to charities being organized to help the victims.

This understanding of the contents of SA and non-SA tweets helped us to select linguistic features for distinguishing between SA and non-SA tweets, as discussed in the next section.

[2]These observations agree with those in [2] which studied the nature of posts in Sina Weibo (a Twitter-like microblogging system in China) after an earthquake in China.

## IV. Features for classification

This section describes the linguistic features used for distinguishing SA and non-SA tweets, which rely on the presence of particular types of words and parts-of-speech (POS) tags in the tweets. The POS are identified using a probabilistic tokenizer and POS tagger designed explicitly for tweets [17], which can also identify emoticons and exclamations apart from standard parts-of-speech.

We propose to use the following low-level syntactic features to classify between SA and non-SA tweets.

**(F1) F-measure:** The formality measure proposed by Heylighen *et al.* [13] is defined as:
*F = (noun freq + adjective freq + preposition freq + article freq – pronoun freq – verb freq – adverb freq – interjection freq + 100) / 2*
where the frequencies are expressed as the percentages of words of a certain category, among the total number of words in a document (tweet). The $F$-value varies between 0 and 100, and is higher for documents which use more formal language. Since SA tweets are usually written more formally [5], we expect SA tweets to have higher F-measure than non-SA tweets.

**(F2) Use of numerals:** Compared to non-SA tweets, SA tweets are more likely to contain numerical information, such as the number of casualties, and contact numbers of helplines / hospitals (see Table I). Hence, we consider the count of numerals present in a tweet as a feature.

**(F3) Use of subjective words:** Non-SA tweets which convey

| Feature of tweets | HBlast | | TBopha | | UFlood | | SHshoot | |
|---|---|---|---|---|---|---|---|---|
| | SA | non-SA | SA | non-SA | SA | non-SA | SA | non-SA |
| mean F-measure ($F \in [0, 100]$) | 75.16 | 66.46 | 75.68 | 63.80 | 72.46 | 71.90 | 72.35 | 70.04 |
| mean fraction of subjective words | 0.039 | 0.085 | 0.052 | 0.114 | 0.064 | 0.129 | 0.059 | 0.073 |
| mean count of numerals | 1.30 | 0.39 | 1.81 | 0.18 | 1.16 | 0.44 | 1.23 | 0.13 |
| mean count of personal pronouns | 0.069 | 0.390 | 0.066 | 0.491 | 0.125 | 0.312 | 0.143 | 0.371 |
| contains emoticons/exclamations | 4.62% | 10.00% | 3.87% | 27.22% | 5.73% | 11.71% | 5.36% | 11.90% |
| contains intensifiers | 1.54% | 4.62% | 3.31% | 7.78% | 4.69% | 8.78% | 6.55% | 6.55% |
| contains modal verbs | 0.0% | 6.92% | 0.55% | 3.33% | 1.04% | 2.93% | 2.38% | 2.98% |
| contains '?' | 0.77% | 11.54% | 2.76% | 10.00% | 1.04% | 7.80% | 1.79% | 14.29% |

**TABLE II: Comparison of SA and non-SA tweets with respect to proposed features. For real-valued features (top 4), the mean value per SA / non-SA tweet is stated. For binary features (bottom 4), the percentage of SA / non-SA tweets which contain the feature is stated.**

the sentiment of users are likely to contain a higher fraction of 'subjective' words, than SA tweets which are more objective in nature. We consider a set of strongly subjective words from a recently developed English subjectivity lexicon for tweets [18], and compute the fraction of words in a tweet which are strongly subjective.[3]

**(F4) Use of personal pronouns:** SA tweets are usually written in an impersonal manner, whereas non-SA tweets are often written from a personal standpoint [5]. To identify whether a tweet is written from a personal standpoint, we check the number of commonly used personal pronouns in the first-person (e.g., *I, me, myself, we*) and the second-person (e.g., *you, yours*), that are present in the tweet.

**(F5) Presence of emoticons and exclamations:** Tweets which convey sentiment often contain emoticons and exclamatory words (e.g., :(, 'omg!', 'oh no!'). We expect non-SA tweets to contain emoticons and exclamations much more frequently than SA tweets.

**(F6) Presence of intensifiers:** Intensifiers are adverbs that are used to boost meaning [25], and are mostly used in informal non-SA tweets to convey stronger sentiments, as in "this situation is *so* weird", "*very* sad to know", "*really* awful news". We check whether a tweet contains any of the 25 most commonly used intensifiers in English [26].

**(F7) Presence of modal verbs:** Modal verbs (such as 'could', 'might', 'must', 'would', 'should') are primarily found in non-SA tweets which reflect the opinion of the users on issues such as how similar tragedies could be prevented (see Table I). We check whether a tweet contains any of the commonly used modal verbs in English (or their shortened forms, such as 'cud' and 'shud').

**(F8) Presence of question marks:** We find that a significant fraction of non-SA tweets contain question marks as part of expressing sentiment (shock, disgust), as in "*another school shooting? how can someone kill eighteen children?*" or "*what kind of sick mentality is this?? #blast*". On the other hand, SA tweets usually state some fact and are far less likely to contain queries.

**Feature Analysis:** Before we turn to classifying tweets, we analyze the tweets with respect to the above features, in order to verify whether the features look promising for the SA vs. non-SA classification task.

Table II states, for the four datasets, the mean values of the real-valued features (F1–F4) per SA tweet and non-SA tweet, and the percentage of SA and non-SA tweets which contain a certain binary feature (F5–F8). Recall that the low-level syntactic features are intended to serve as proxies for more complex notions such as subjectivity, linguistic register, and personal / impersonal tone of tweets, which are more challenging to identify. The values in Table II indicate that SA tweets are indeed written in a more formal, more impersonal and less subjective style, as compared to non-SA tweets. For instance, across all the four datasets, SA tweets have higher F-score and contain significantly more numerals on average, as compared to non-SA tweets. Conversely, non-SA tweets are much more likely to contain emoticons, intensifiers, modal verbs, query marks, subjective words, and personal pronouns. These observations show that the chosen features consistently distinguish between SA and non-SA tweets across diverse disaster events.

## V. SA VS. NON-SA CLASSIFICATION RESULTS

We now use the features described in the previous section to classify between SA and non-SA tweets. We compare classifiers trained on two sets of features – (i) the proposed set of features, and (ii) a bag-of-words model (BOW) where each distinct word (unigram) in the set of tweets is considered as a feature. For the BOW classifier, following the approach of [5], we pre-process the tweets by replacing URLs and @user-mentions with unique symbols, case-folding, and removing standard English stopwords.[4]

For both cases, we use a Support Vector Machine (SVM) classifier; specifically, we use the LIBSVM package [27] with default settings, i.e., a Radial Basis Function kernel. We match

[3] Specifically, [18] assigns to each term, a likelihood of the term to be used in a subjective context; we consider those terms which have subjectivity likelihood more than 0.5.

[4] We verified that the classification accuracies of the BOW classifier are actually improved by these pre-processing steps.

| Train set | Test set | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | HBlast | | TBopha | | SHshoot | | UFlood | |
| | BOW | Proposed | BOW | Proposed | BOW | Proposed | BOW | Proposed |
| HBlast | 67.308% | **68.077%** | 47.922% | **76.731%** | 52.083% | **67.560%** | 51.515% | **60.101%** |
| TBopha | 50.000% | **65.385%** | 50.969% | **83.103%** | 50.000% | **72.024%** | 48.485% | **59.091%** |
| SHshoot | 50.385% | **65.385%** | 50.139% | **81.995%** | 66.369% | **77.976%** | 51.515% | **63.384%** |
| UFlood | 50.000% | **63.846%** | 49.862% | **71.191%** | 50.000% | **72.917%** | 51.515% | **58.333%** |

**TABLE III: Classification accuracies of SVM using (i) bag-of-words features (BOW), (ii) proposed features. Random baseline accuracy is 50%. Diagonal entries are for in-domain classification, while the non-diagonal entries are for cross-domain classification.**

the predictions of the classifier with the human annotations of the tweets (as described in Section III) to measure the accuracies.

We evaluate the performance of the classifiers in two types of classification tasks – (i) *in-domain classification*, where tweets related to the same event are used to train and test the classifier via standard 10-fold cross validation, and (ii) *cross-domain classification*, where we train the classifier on one dataset (event) and test on another. As stated earlier, the cross-domain performances are more important, since in reality, SA / non-SA classifiers would be trained over past events and then deployed to classify tweets on newly occurring events.

**In-domain classification:** The classification accuracies (percentage of tweets correctly classified) considering the two feature-sets are shown in the *diagonal entries* of Table III. It is expected that the BOW model would perform relatively well in in-domain classification, since the training event and test event share a common vocabulary. However, the performances of the SVM classifier with the proposed set of features are significantly better than those of the BOW classifier.

**Cross-domain classification:** The *non-diagonal entries* in Table III show the cross-domain classification accuracies with the two sets of features. In each case, the event on the left is the training event (on which the classifier is trained), and the event at the top of the column is the test event (on which prediction is done).

The performances of the BOW model are significantly inferior for cross-domain classification, sometimes close to (or even lesser than) the random baseline performance of 50%. This is because the training and testing datasets (related to two different disaster events) have very different vocabularies. On the other hand, the classifier based on the proposed linguistic features significantly out-perform the BOW classifier in all cases. This implies that the selected low-level features can robustly distinguish between SA and non-SA tweets irrespective of the vocabulary / linguistic style related to specific events. Thus, classifiers can be trained over these features extracted from past disasters, and then deployed to classify tweets posted during future events (which is the realistic application of SA vs. non-SA classifiers).

**Analysis of misclassified tweets:** Finally, we attempt to analyze the limitations of the proposed classification scheme by observing the tweets which were *misclassified*. Since it is a more critical error to misclassify SA tweets as non-SA (than to misclassify non-SA tweets as SA), we focus on the SA tweets (according to human judges) which were misclassified by the proposed classifier. Table IV shows some examples of such tweets.

| |
|---|
| oh no !! unconfirmed reports that the incident in #newtown #ct may be a school shooting. police on the way |
| first image: kids walking out, crying as they evacuate sandy hook elem. in #newtown. <URL> |
| pls contact on 919454624822 to track 2000 people near gaurikund ... he is with my aunt |
| these supplies needed, please help. trucks departing regularly for relief camps. <URL> |
| heavy rain since 12am in cdeo :( whoa ! can now feel typhoon pablo. stay safe #mindanao |

**TABLE IV: Examples of misclassified SA tweets. Most of these contain multiple fragments, some of which convey situational information while some other fragments are more conversational in nature.**

We observe that a large majority of the misclassified SA tweets contain *multiple* fragments – while some of the fragments contain SA information (and are written relatively formally), the other fragments convey personal sentiments written subjectively. Such mixing of SA and personal sentiments in the same tweet brings down the overall formality / objectivity of the tweet – for instance, the F-measure values for most of such 'mixed' tweets are less than 60.0, as compared to the overall mean F-measure of 74.04 for SA tweets – and this probably leads to misclassification. In future, we plan to investigate whether, for the tweets which contain multiple fragments, it is better to analyze each fragment individually for formality and subjectivity.

## VI. CONCLUDING DISCUSSION

In this study, we developed a NLP-based classifier to identify microblogs (tweets) which contribute to situational awareness during disaster events. We demonstrated that low-level syntactic features can be used to abstract more complex notions of subjectiveness, formal / informal linguistic register, and personal / impersonal style of tweets, and hence can be used to develop streamlined classification models which outperform much heavier bag-of-words models. The classifier developed in this work can function as an important building block for filtering out particular types of tweets for more advanced tasks like summarization and rumor detection.

As future work, we plan to improve the SA vs. non-SA classifier by incorporating more features. It can be noted that use of some Twitter-specific features – such as, whether a tweet has been retweeted (propagated) by different users – may help to increase the classification accuracies, since SA tweets are usually more heavily retweeted by the masses. However, it takes some time before a tweet gets sufficiently retweeted, and we observe that several tweets containing critical SA information are not retweeted at the time when they are first obtained (though they get retweeted later). Hence, we decided *not* to use retweets as a feature, so as not to bias the classifier towards more popular tweets (which might potentially delay the identification of SA tweets). Also, given that many of the SA tweets contain names of geographical locations in the region affected by the disaster, we would like to explore the utility of event-specific geographical lexicons [10] in identifying situational awareness tweets.

## ACKNOWLEDGEMENTS

This research was partially supported by grants from the Information Technology Research Academy (ITRA), under the Department of Electronics and Information Technology (DeitY), Government of India (Ref. No.: ITRA/15 (58) /Mobile/DISARM/05). Additionally, K. Rudra was supported by a fellowship from Tata Consultancy Services, and S. Ghosh was supported by a fellowship from the Alexander von Humboldt Foundation.

## REFERENCES

[1] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen, "Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM, 2010, pp. 1079–1088.

[2] Y. Qu, C. Huang, P. Zhang, and J. Zhang, "Microblogging After a Major Disaster in China: A Case Study of the 2010 Yushu Earthquake," in *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW)*. ACM, 2011, pp. 25–34.

[3] J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power, "Using Social Media to Enhance Emergency Situation Awareness," *IEEE Intelligent Systems*, vol. 27, no. 6, pp. 52–59, 2012.

[4] M. A. Cameron, R. Power, B. Robinson, and J. Yin, "Emergency Situation Awareness from Twitter for Crisis Management," in *Proceedings of the International Conference Companion on World Wide Web (WWW)*. ACM, 2012, pp. 695–698.

[5] S. Verma, S. Vieweg, W. J. Corvey, L. Palen, J. H. Martin, M. Palmer, A. Schram, and K. M. Anderson, "Natural Language Processing to the Rescue? Extracting "Situational Awareness" Tweets During Mass Emergency," in *Proceedings of International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2011.

[6] S. E. Vieweg, "Situational Awareness in Mass Emergency: A Behavioral and Linguistic Analysis of Microblogged Communications," Ph.D. dissertation, University of Colorado Boulder, USA, 2012.

[7] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: real-time event detection by social sensors," in *Proceedings of International Conference on World Wide Web (WWW)*. ACM Press, 2010, pp. 851–860.

[8] N. B. Sarter and D. D. Woods, "Situation awareness: a critical but ill-defined phenomenon," *The International Journal of Aviation Psychology*, vol. 1, no. 1, pp. 45–57, 1991.

[9] G. Neubig, Y. Matsubayashi, M. Hagiwara, and K. Murakami, "Safety information mining – what can nlp do in a disaster –," in *Proceedings of International Joint Conference on Natural Language Processing (IJCNLP)*, 2011, pp. 965–973.

[10] I. Varga, M. Sano, K. Torisawa, C. Hashimoto, K. Ohtake, T. Kawai, J.-H. Oh, and S. D. Saeger, "Aid is out there: Looking for help from tweets during a large scale disaster." in *Proceedings of Annual Meeting of the Association for Computational Linguistics*. The Association for Computer Linguistics, 2013, pp. 1619–1629.

[11] A. Hannak, E. Anderson, L. F. Barrett, S. Lehmann, A. Mislove, and M. Riedewald, "Tweetin' in the Rain: Exploring societal-scale effects of weather on mood," in *Proceedings of International AAAI Conference on Weblogs and Social Media (ICWSM)*. The AAAI Press, 2012.

[12] E. Riloff and J. Wiebe, "Learning Extraction Patterns for Subjective Expressions," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 2003, pp. 105–112.

[13] F. Heylighen and J.-M. Dewaele, "Formality of Language: definition, measurement and behavioral determinants," Center "Leo Apostel", Free University of Brussels, Tech. Rep., 1999.

[14] A. Mosquera and P. Moreda, "Enhancing the discovery of informality levels in Web 2.0 texts," in *Proceedings of the Language Technology Conference (LTC)*, 2011.

[15] 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*, 2009.

[16] Y. Hu, K. Talamadupula, and S. Kambhampati, "Dude, srsly?: The Surprisingly Formal Nature of Twitter's Language," in *Proceedings of International AAAI Conference on Weblogs and Social Media (ICWSM)*. The AAAI Press, 2013.

[17] O. Owoputi, B. O'Connor, C. Dyer, K. Gimpel, N. Schneider, and N. A. Smith, "Improved part-of-speech tagging for online conversational text with word clusters," in *Proceedings of NAACL-HLT*, 2013, pp. 380–390.

[18] S. Volkova, T. Wilson, and D. Yarowsky, "Exploring Sentiment in Social Media: Bootstrapping Subjectivity Clues from Multilingual Twitter Streams," in *Proc. ACL (Vol2: Short Papers)*, 2013.

[19] "Sandy Hook Elementary School shooting – Wikipedia," 2012, http://en.wikipedia.org/wiki/Sandy_Hook_Elementary_School_shooting.

[20] "Typhoon Bopha – Wikipedia," 2012, http://en.wikipedia.org/wiki/Typhoon_Bopha.

[21] "2013 Hyderabad blasts – Wikipedia," 2013, http://en.wikipedia.org/wiki/2013_Hyderabad_blasts.

[22] "2013 North India floods – Wikipedia," 2013, http://en.wikipedia.org/wiki/2013_North_India_floods.

[23] "GET statuses/sample — Twitter Developers," 8 2013. [Online]. Available: https://dev.twitter.com/docs/api/1/get/statuses/sample

[24] K. Tao, F. Abel, C. Hauff, G.-J. Houben, and U. Gadiraju, "Groundhog Day: Near-duplicate Detection on Twitter," in *Proceedings of the International Conference on World Wide Web (WWW)*, 2013, pp. 1273–1284.

[25] R. Ito and S. Tagliamonte, "Well weird, right dodgy, very strange, really cool: Layering and recycling in english intensifiers," *Language in Society*, vol. 32, no. 2, pp. 257–279, 2003.

[26] R. Quirk, S. Greenbaum, G. Leech, J. Svartvik, and D. Crystal, *A comprehensive grammar of the English language*. Cambridge Univ Press, 1985, vol. 397.

[27] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.