Calling for Response: Automatically Distinguishing Situation-Aware Tweets During Crises

Xiaodong Ning(✉), Lina Yao(✉), Xianzhi Wang, and Boualem Benatallah(✉)

University of New South Wales, Sydney, NSW, Australia
{z5122770,xiaodong.ning}@student.unsw.edu.au, lina.yao@unsw.edu.au, sandyawang@gmail.com, boualem.benatallah@unsw.edu.au

Abstract. Recent years have witnessed the prevalence and use of social media during crises, such as Twitter, which has been becoming a valuable information source for offering better responses to crisis and emergency situations by the authorities. However, the sheer amount of information of tweets can't be directly used. In such context, distinguishing the most important and informative tweets is crucial to enhance emergency situation awareness. In this paper, we design a convolutional neural network based model to automatically detect crisis-related tweets. We explore the twitter-specific linguistic, sentimental and emotional analysis along with statistical topic modeling to identify a set of quality features. We then incorporate them to into a convolutional neural network model to identify crisis-related tweets. Experiments on real-world Twitter dataset demonstrate the effectiveness of our proposed model.

Keywords: Convolutional neural network · Situational awareness

1 Introduction

With the arrival of the information age, Twitter has become a popular platform for people to post situations, exchange information, seek and offer advice during crises [2]. Such tweets has great significance for both the people affected by crises and those who plan to help the affected people. First, rich and useful situation-aware tweets are an important information source for decision-making agencies like governments to make a reasonable emergency plan for allocating rescuers and relief materials. Second, tweets have the advantage of timeliness, which can reflect the situations and circumstances at real time. According to the expert experience, there is a 72 h 'golden window' for post-crisis relief, and as the time passes, rescue efficiency degrades significantly. Therefore, quick acquisition of tweets about crises can improve the response speed of government, and further reduce the casualties and property damage. For the above reasons, it is essential to design an efficient information extraction technique that can capture the valuable tweets about an crisis as soon as it happens [3].

G. Cong et al. (Eds.): ADMA 2017, LNAI 10604, pp. 195–208, 2017.
https://doi.org/10.1007/978-3-319-69179-4_14

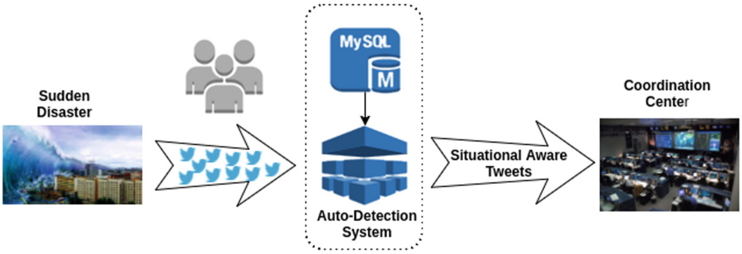


Fig. 1. Overall flow chart of detecting situation-aware tweets

Generally, tweets can be collected by event-related keywords using Twitter's public API during crisis. However, the collected tweets are generally in huge number, a significant fraction of which is noise (or irrelevant information). Consequently, facing with the overwhelming amount of information, it is unfeasible to select the situation-aware tweets manually. Therefore, it is essential to design an automatic, efficient model to detect the situation-aware valuable ones from online tweets. In fact, situation-awareness is a focused field of study aimed at understanding the environment and is critical to the decision-making as response to mass emergencies. Although there are several previous research efforts on situation awareness, most of them [6,7,9] train and validate their models for each individual event. These models do not show the capability of handling cross-crisis task. Some other researchers [1,8] train cross-crisis model from dataset of previous events and validate it for a new event which are partly annotated by humans. For this reason, they require data annotated by humans to train the model before it can start working while annotation can be highly time-consuming whereas time is precious in crises.

Based on the above discussion, we propose an automatic model to discriminate the situation awareness tweets for a *new* crisis event without human-participation.

- We present an automatic approach to capturing the critical crisis-related information conveyed on social media to enhance public responses to crisis situation in the real world.
- We design a set of quality features based on text-based measures such as emotions, linguistics, topics and entities to characterize various aspects of situation-aware tweets.
- We learn an effective one dimensional CNN-based predictive model for detecting crisis-related situation-aware tweets from a *new* crisis event, and deploy the model on a real-world twitter dataset including 6 categories of crisis events. Our model yields an increase of multiple evaluation metrics compared with a series of baseline and the state-of-the-arts methods. It also provides us a thorough understanding of the predictive results.

2 Approach Overview

As we mentioned above, only a fraction of tweets filtered by the crisis related key words and locations are ‘Situation-awareness’. For example, we show two tweets about ‘Boston Bombing’ which are collected by key words in Fig. 2.

Fig. 2. Two tweets about the ‘Boston Bombing’ crisis

We can find that both of the two tweets are collected by the keyword ‘Boston Bomb’ using Twitter’s API, yet the first one is situation-aware to the crisis while the second one is not situation-aware (the first one describes the details of the crisis while the second one just talks irrelevant things). By analyzing the content carefully, we find that the language styles are different between the two tweets. The emotional index is higher in the first one while the subjectivity is stronger in the second one. Therefore, we distinguish between the situation-aware and the other tweets according to their language attributes. For this reason, our approach involves two main steps: feature extraction and model learning (the overall structure of our methodology is shown on Fig. 1.

2.1 Feature Extraction

In this section, we extract several types of content-based features from each tweet to detect whether it is situation-aware or not.

Linguistic Features. According to the research in the sociology and psychology, linguistic features can reflect the mental activities and behavioral intention of posters. We extract four kinds of linguistic attributes: subjectivity, part-of-speech, tenses and lexical density. Specifically, we use Textblob[1] to calculate the subjectivity scores of tweets. Each subjectivity score ranges within $(0, 1)$ and higher score indicates stronger subjectivity of tweet. We analyze the tense of tweets by measuring the different verb tenses: ‘past tense’, ‘present participle’, ‘base form’ and ‘past participle’ using the Stanford’s Part-of-speech (POS) tool.[2] The values of verb tenses are calculated by counting the occurrences of corresponding words in the content. Lexical density is measured by keywords such as ‘verb’, ‘adverb’, ‘symbol’ and ‘number’ (also done by POS). Finally, the linguistic features add up to a 43-dimensional vector.

[1] https://textblob.readthedocs.io/en/dev/.
[2] https://nlp.stanford.edu/software/tagger.shtml.

Table 1. Mn (Mean), Std (Standard deviation) of the linguistic representative attributes across the Situation-Awareness (SA) and Non-Situational Awareness (Non-SA) tweets. PRP = personal pronoun, SYM = symbol, VBD = verb past tense, RBS = superlative adverb, SUB = subjectivity; JJ = adjective; NN = nourns; CD = cardinal number.

Feature types		Linguistic features							
Subtypes		PRP	SYM	VBD	RBS	SUB	JJ	NN	CD
Situational awareness	Mn	0.45	0.01	0.86	0.46	0.01	1.05	3.69	0.35
	Std	0.78	0.12	0.99	0.79	0.15	1.03	2.47	0.69
Non-situational awareness	Mn	1.01	0.005	1.01	0.72	0.03	0.94	1.90	0.20
	Std	1.12	0.01	1.17	0.68	0.09	1.03	1.72	0.57

Emotional Features. In order to estimate the emotional features in tweets, we analyze three aspects: emotional states, emotional intensity and content polarity. The emotional states are measured by 10 emotional terms in tweets such as ‘joy’, ‘sadness’, ‘anger’, ‘nervousness’, ‘fear’, ‘disgust’, etc. The score of each term is estimated using the Empath API [5].[3] We analyze the emotional intensity of each tweet by averaging the arousal scores of its words given in ANEW dictionary (Affective Norms for English Words).[4] Polarity is measured in a score between $(0, 1)$ by analyzing the whole tweet using the Textblob API. Finally, we get a 17-dimensional vector for this part.

Table 2. Mn (Mean), Std (Standard deviation) of the emotional representative attributes across the Situation-Awareness (SA) and Non-Situational Awareness (Non-SA) tweets.

Feature types		Emotional features						
Subtypes		Arousal	Polarity	Sad	Nervous	Angry	Joy	Disgust
Situational awareness	Mn	5.26	0.03	0.03	0.15	0.20	0.003	0.009
	Std	0.85	0.08	0.08	0.26	0.31	0.07	0.097
Non-situational awareness	Mn	4.40	0.07	0.01	0.06	0.04	0.005	0.002
	Std	1.26	0.04	0.07	0.13	0.12	0.06	0.049

Entity-Based Features. Besides the former two types of features, we also take the entity-based features into consideration. Typically, some entities are associated closely with the relevant tweets during mass emergencies, e.g., ‘government’, ‘journalism’, and ‘Medical Emergency’. We choose up to 17 entity-topics (each topic contains several entity words which are related to that topic) and measure the corresponding scores in each tweet using the Empath API.

[3] http://empath.stanford.edu/.
[4] https://tomlee.wtf/2010/06/16/anew/.

Table 3. Mn (Mean), Std (Standard deviation) of the entity-based representative attributes across the Situation-Awareness (SA) and Non-Situational Awareness (Non-SA) tweets.

Feature types		Entity features					
Subtypes		Journalism	Government	Medical	Injury	Sympathy	Weak
Situational awareness	Mn	0.015	0.02	0.015	0.03	0.023	0.07
	Std	0.09	0.13	0.128	0.18	0.15	0.28
Non-situational awareness	Mn	0.008	0.004	0.008	0.01	0.015	0.03
	Std	0.015	0.07	0.015	0.12	0.13	0.19

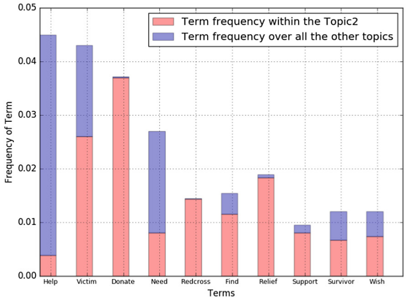
Topical Features. Although the specific locations and events can be different in various crises, the core content of tweets posted fall into several common topics such as relief, blessing, seeking for help, news report, etc. Previous research [22–24] has already shown topical features can be a strong complement for the handcrafted attributes due to its ability of capturing the latent semantic features in social media. Then we apply latent Dirichlet allocation (LDA) model to extract the topics in tweets. With the empirical evaluations, the optimal number of topics is set as 7 in our work. We show the several top frequent words for each clustered topic in Table 4. From the table, we can observe that the clustered words in each topic are significantly different. Topic 2 is closely related to relief (e.g. support, redcross, donate, etc.) while the topic 3 contains many news report related words such as: http, bbc, hit, news, etc. To demonstrate the effectiveness of word clustering from LDA, we choose topic 2 as example (as it is related to relief) and show the top-10 most relevant term frequencies within it compared to the overall frequencies in Fig. 3(a). From the figure, we can observe that topic 2 accounts for the overwhelming proportion of the overall ten term frequencies meaning that extracted topics are separated well.

In order to give an intuitive insight, we select some representative terms from the former three kinds of features by calculating the means and standard deviations of them in 'SA' and 'Non-SA' Tweets in Tables 1, 2 and 3. We can clearly observe that there exists distinguishable feature distribution patterns between 'SA' and 'Non-SA' tweets. Figure 3(b) shows the significant difference of topic distributions between the 'SA' and 'Non-SA' tweets. We further characterize the distribution differences in tweets based on the categories as following:

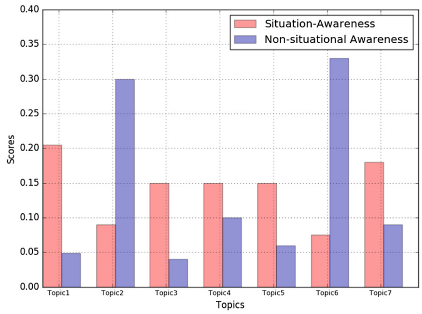
Feature Analysis. *Linguistic Feature*: We can observe that 'SA' tweets show significantly higher mean value than 'Non-SA' ones in nouns(+94%) and cardinal numbers(+75%) while the latter shows 200% and 124% higher values in subjectivity and personal pronoun respectively. Presumably, the different distributions of linguistic features conform to the common sense as situation-aware tweets always provide more objective content with nouns and cardinal numbers during crisis.

Table 4. LDA vocabulary distribution

LDA topic	Vocabulary
Topic1	*need,volunteer, share, tweet, http, evacuation, safe, affect...*
Topic2	*help, victim, donate, need, redcross, find, relief, support...*
Topic3	*http, video, photo, facebook, bbc, hit, news, youtube, ...*
Topic4	*day, trend, hours, weather, tonight, place, night, update...*
Topic5	*suspect, caught, police, attack, arrest, rip, tragedy, fire, kill...*
Topic6	*thank, love, good, better, happy, well, final, tomorrow, hope...*
Topic7	*lol, joke, want, sigh, really even, guy, more, extremely...*



(a)



(b)

Fig. 3. (a) Top-10 most relevant term frequencies within topic 2 compared to the term frequencies over all the other topics; (b) Topic distributions between 'Situation-awareness' and 'Non-Situational Awareness' tweets;

Emotional Feature: From the chosen emotional features, it is notable that 'SA' tweets show higher mean value in all of the negative emotions (e.g. sad(+200%), angry(+400%), nervous(+150%).etc) than 'Non-SA' ones while the values of joy and polarity are higher in the latter. We can infer that 'SA' tweets tend to associate with strong negative emotions due to distress caused by crisis while 'Non-SA' tweets generally talk about irrelevant topics.

Entity Feature: As for entity features, the mean values of 'SA' tweets are consistently higher than 'Non-SA' tweets in all the chose entity topics, which confirms assumption in the previous feature extraction part. It is reasonable that situation-aware contains more crisis-related entities.

Topical Feature: By analyzing the topic distributions among two categories, we can find that the mean value are much higher in 'Non-SA' tweets in topic 2 and topic 6 while 'SA' tweets get higher values in other topics. The words appearing in the topic 2 and 6 can be regarded as 'support and blessing' related terms while the other topics contain many detail-related terms which can transmit situation-aware information.

2.2 Situation-Awareness Identification

To identify the situation-aware tweets in each crisis. we propose a 1D-CNN model in an empirical data-driven manner. Deep learning has been proven to achieve promising performance in various domains such as image processing [10] and speech recognition [11], as a result of its ability to model high-level representations and capture complex relationships of data via a stacking multiple layered architecture. Recently, it is also adapted to the field of social media content analysis [12, 13]. We show the overall structure of our model in Fig. 4.

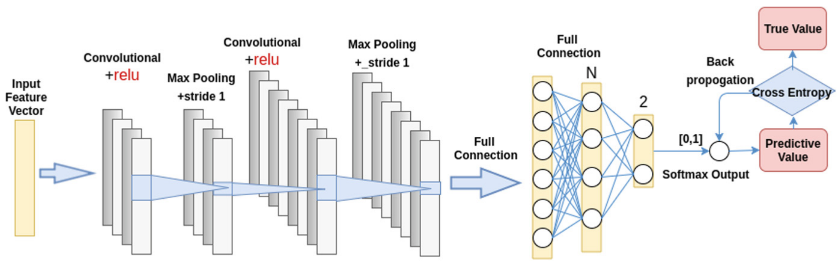


Fig. 4. 1DCNN architecture

We will explain the details of our model in the following part. The list of quality features above are used in a form of vectors as input of our model. The specific structure of our model is as follows: The first and third layers are convolutional layers, the second and forth are max-pooling layers, and the fifth and sixth are two fully-connected layers. The first and third convolutional layers in our model take a set of $F1$ and $F2$ independent filters respectively and slide them over the whole feature vector with stride size Fs and filter length Fl. Along the way, dot product is taken between the filters and chunks of the input feature. Filters are used to generate the feature vectors in each filter length. In this way, the original feature vector is projected into a stack of feature maps (vector maps in our work). Additionally, we apply the rectified linear unit (ReLU) activation function on the dot product from filters. We take the first convolutional layer as example to show the details of convolutional operation and RELU function in Eqs. 1 and 2: $Conv_i(j)$ denotes the j_{th} convolutional output from the i_{th} filter and input vector I (84 dimensional vector in our work); F_i denotes the i_{th} filter vector; $Convrelu_i(j)$ denotes the relu function result of the $Conv_i(j)$;

$$Conv_{ij} = I[j * Fs, (j * Fs + Fl)] * F_i \tag{1}$$

$$Convrelu_{ij)} = Max(0, Conv_{ij}) \tag{2}$$

After the convolutional layers, the hidden relationships between features can be combined by the dot product from each filter. The captured features can play an important role in the final performance in our model.

The multiple filters in convolutional layers can be updated automatically via the evolution of network. Followed by each convolutional layer, we add one max-pooling layer with max-filter length Pl and stride size Ps. Similar to previous part, we show the details of first max-pooling layer in Eq. 3: $Max(i, j)$ denotes the j_{th} max-pooling output from the i_{th} convolutional layer output with Relu $Convrelu_i$;

$$Maxoutput(i, j) = Max(Convrelu_i[j * Ps, (j * Ps + Pl)]) \tag{3}$$

The max-pooling layer can down-sample the feature representation, reducing its dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned. After the max-pooling layer, the important feature can be selected. The convolved and max-pooled feature vectors will be fed into two fully-connected layers (the neurons in each layer are N and 2) with dropout probability (Kp) applied for the high-level reasoning. Neurons in each fully-connected layer have full connections to all activations in the previous layer, as seen in regular Neural Networks. Their activations can therefore be computed using matrix multiplication followed by a bias offset. The process of first fully connected layer is shown in Eq. 4: L denotes the unfolded vector from the previous layer; W denotes the weight matrix with shape $(N \times L)$; B denotes the bias vector with N length; $Output(N)$ denotes the output vector of this layer;

$$Output(N) = W \times L + B \tag{4}$$

Finally, we use the softmax as the output layer of the last fully-connected layer. Each node in the softmax layer produces the probabilities of two classes including 'situation-aware' and 'non-situational aware'. The class with higher probability will be output as the prediction.

3 Experiments

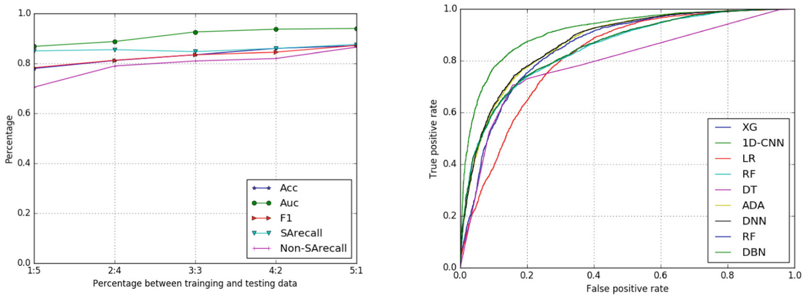
3.1 Settings

Dataset. We use the tweets dataset provided by [4], which contains six crises that occurred in English-speaking countries between October 2012 and July 2013. This dataset was collected using Twitter's API by event-related keywords and geolocations. All the tweets in this dataset have been manually classified into 'Situation-Aware' and 'Non-situational Aware'. Table 5 presents the details of our dataset, where we observe that the numbers of 'SA' and 'Non-SA' Tweets are approximately equal.

Baseline Methods. We choose the following baseline methods (the parameters of all method are tuned for optimal performance), including *Random Forrest* (the number estimators as 300, max depth as 5) [6], *Xgboosting* (learning rate as 0.05, max depth as 4 and minimum descending value of cost function as 0.1) [14], *Logistic Regression*, *Decision Tree* (max depth as 3), *Support Vector Machine* (kernel as 'poly' and C parameter as 1), *AdaBoosting* (number of estimators as

Table 5. Statistics of dataset

Crisis	# of SA Tweets	# of Non-SA Tweets
Sandy hurricane	6138	3870
Alberta flood	5189	4842
Boston bombing	5638	4364
Oklahoma tornado	4827	5165
Queensland floods	5414	4619
West texas explosion	5246	4760



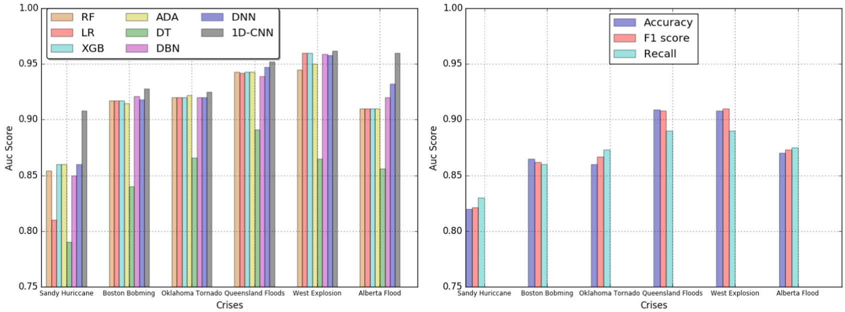
(a) (b)

Fig. 5. (a) Cross-crisis performance comparison with varied training/testing ratio. The X-axis indicates the number of crisis topic used for training, and the rest of crisis topics used for testing; (b) ROC performance over crisis topic *'Sandy Hurricane'*;

200 and learning rate as 0.3) [6]. Besides the above methods, we also choose two deep learning based methods as our baseline: *Deep Belief Network* (four hidden layers (with 110,200,400,300 units in each layer), learning rate as 0.001,activation function as relu, number of epochs as 200) and *Deep Neural Network* (four hidden layers (with 500, 600, 400, 200 in each layer), learning rate as 0.0001, training epochs as 1000).

3.2 Results

Parameter Tuning. We conducted extensive experiments to determine the optimal configuration of parameters for the 1D-CNN. There are two types of parameters in our model: the first type includes weights and biases in the model layers, which can be initiated randomly and learned afterwards from each iteration; the second type includes the parameters that should be configured manually. In particular, we select nine most common hyper-parameters for our 1D-CNN model, namely the learning rate lr, the dropout probability Kp (used to prevent overfitting), convolutional filters length Fl, number of filters in the first convolutional layer $F1$, number of filters in the third convolutional layer $F2$, stride size



(a) (b)

Fig. 6. (a) AUC performance of all methods across 6 crises. (b) Performance (accuracy, recall and F1 score) of our model over six crises

Table 6. Comparison results

	Accuracy	F1 score	AUC	Recall
Random forest [6]	0.849	0.851	0.914	0.852
Logistic regression	0.808	0.81	0.91	0.825
Xgboosting [14]	0.845	0.843	0.918	0.855
Adaboosting [6]	0.836	0.837	0.916	0.839
Decision tree	0.79	0.80	0.851	0.797
Support vector machine	0.816	0.814	0.91	0.823
Deep belief network	0.822	0.825	0.919	0.83
Deep neural network	0.847	0.848	0.92	0.846
Ours	**0.871**	**0.872**	**0.94**	**0.87**

in convolutional layers Fs, max-filter length Pl, stride size Ps in max-pooling layers and number of neurons in first connection layer N. Since the weight W and bias b in each neural layer can be learned automatically via the evolution of model network, we focus on tuning the hyper-parameters lr, Kp, Fl, $F1$, $F2$, Fs, Pl, Ps and N. In particular, we fix the number of iterations as 1000 for each experiment and try different combinations of parameters by changing the value of one parameter and keep the other eight parameters. Finally, we set $lr = 0.0001$, $Kp = 0.05$, $Fl = 3$, $F1 = 2$, $F2 = 2$, $Fs = 1$, $Pl = 2$, $Ps = 1$, $N = 1200$ as default setting of our method.

Performance Study. The aim of our work is to effectively and automatically discriminate the 'SA' tweets for a totally new crisis, hence we design our experiments by choosing five events as training data and leave the sixth event as testing data iteratively (similar to six-fold validation where validation dataset comes from a new crisis). We evaluate the performance of each classifier by four metrics: Accuracy, Recall, AUC, and F1 score. Table 6 shows the overall

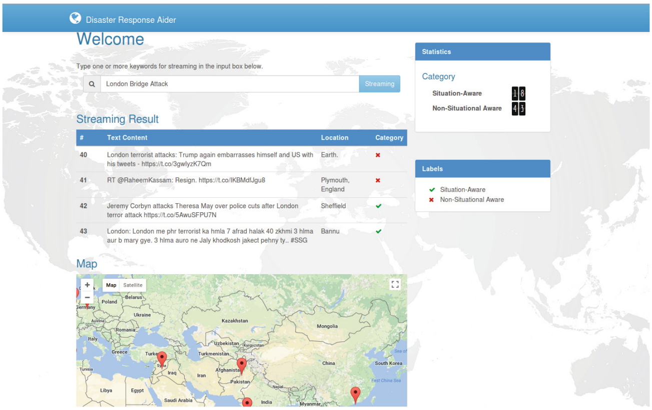


Fig. 7. Website interface

comparison results. From the table, we can observe that our method outperforms the baseline methods consistently in all the four metrics. It is notable that all of the baseline models (except decision tree) achieve more than 80% in accuracy, f1 score, recall and 90% in auc score, which confirms the high effectiveness of our feature engineering. Random forest provides best performance among the non-deep learning methods while deep neural network presents a similar performance. As for another deep learning model, deep belief network does not show better performance than other baseline methods. In particular, our model achieves improvement in recall, F1 score and accuracy by 2%, and 2.6% in AUC compared to the high baseline, which confirms convolutional layer can indeed capture some hidden but important relationships among features. In addition, we also evaluate the cross-crisis performance comparison with varied training/testing ratio, ROC performance over crisis topic 'Sandy Hurricane', AUC performance of all methods across 6 crises and details of 1D-CNN performance over six crises in Figs. 5 and 6. In the end, we choose 'Sandy Hurricane' and 'Queensland Floods' as examples and show their confusion matrix in Fig. 8.

Additionally, we build a website for interaction with our system and show web-interface in Fig. 7. We provide an input box for user to enter a crisis-related keyword or geo-location and then return the classified tweets with their location in the rolling box. We also display the geo-location of each tweet in the global map to give an insight for user. And we also provide a demo[5] based on the real crisis 'London Bridge Attack' to show our system.

4 Related Work

There are many previous work on managing social media information during disaster. Olteanu et al. [4] proposed a crisis lexicon to efficiently query twitter

[5] https://youtu.be/J1GbLppA50c.

N=10012	SA(T)	Non-SA(T)
SA(P)	3662	702
Non-SA(P)	900	4748

(a)

N=10033	SA(T)	Non-SA(T)
SA(P)	4118	501
Non-SA(P)	577	4837

(b)

Fig. 8. Confusion matrix for 'Sandy Hurricane' (a) and 'Queensland Floods' (b). The model is trained on the other 5 crises and tested on the targeted crisis. 'SA'='Situation-aware'; 'Non-SA' = 'Non-situational aware'; T = True; P = Predicted; N = total number of tweets;

to extract crisis-related messages during emergency events, which outperforms using only a set of key words chosen manually by experts. Imran et al. [7] utilized naive bayes classifier (NB) to classify the tweets into four categories and [8] implemented the conditional random field (CRF) to extract valuable information from the classified tweets. Imran et al. [15] utilized an unsupervised algorithm to capture the dynamic crisis-related topics from social media. Stowe et al. [16] presented a system for classifying disaster-related tweets with support vector machine classifier. In 2014, Imran et al. [1] proposed an artificial intelligence disaster response platform for detecting the situation-aware tweets during disaster by combining the human intelligence and machine intelligence, and they [17] study two challenges in their design: identifying which elements should be labeled and determining when to ask for annotations to be done. Similarly, Popoola [18] developed an online platform that involves citizens participation for timely information verification during natural disasters. Verma et al. [9] utilized two classifiers(naive bayes and max entropy) to detect whether a tweet is situation-aware or not. Ashktorab et al. [19] proposed a method to detect whether the tweet is damage or causality related. MacEachren [20] proposed a web-enabled geo-visual analytics approach to leveraging Twitter in support of crisis management and displaying the tweets content based on place, time and concept characteristics. Morstatter [25] developed a system for analysts with little information about an disaster to gain knowledge through the use of effective visualization techniques. This system connects human intelligence with rich data so that human clues can inform search and guide a users query to form better understanding about the disaster. Horita [21] proposed a framework which utilizes an extended model $oDMN^+$ for improving the understanding of how to leverage big data in the organizations decision-making.

Among the previous works, [1,7,9,16–19] are most related to our work. However, some of them [1,19] focus only on the application area instead of the model and feature extraction. Among the other works, [18] requires a lot of human participation to verify a disaster information and cannot automatically distinguish informative messages during disaster. [7,9,16] train and validate their models for each individual event. Those models do not show capability of handling cross-event task. [7,17] try to train cross-event model from dataset of previous events and validate it for a new event which are partly annotated by humans. In other words, the methods proposed in the previous work require data annotated by humans to train the model before it can start working. Annotation can be highly time-consuming whereas time is precious in crises. Compared to previous work, our model can automatically distinguish the situation-aware tweets during a new crisis without human participation and experimental results over real dataset show that our model can excellently handle the inter-event identification task.

5 Conclusions

In this paper, we present a convolutional neural network based model for automatic extraction of situation-aware tweets posted during crises. Our model extracts a set of text-based features of tweets: emotional, linguistic, topical and entity-based attributes, and utilizes an one-dimensional CNN to capture the hidden relationships between those features. Experimental results on a real dataset demonstrate that our model can excellently handle the cross-crisis classification task (AUC 94%, recall 87%, accuracy 87.1% and F1 score 87.2%) and perform better than a series of baseline methods. In future work, we will investigate multi-modal features of tweets and embed our current model into a multi-view learning framework. We also plan to further fuse the model into a dynamic crisis tracking system through social media streams.

References

1. Muhammad, I., et al.: AIDR: Artificial intelligence for disaster response. In: Proceedings of the 23rd International Conference on World Wide Web. ACM (2014)
2. Takeshi, S., Okazaki, M., Matsuo, Y.: Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proceedings of the 19th International Conference on World wide web. ACM (2010)
3. Jie, Y., et al.: Using social media to enhance emergency situation awareness. IEEE Intell. Syst. **27**(6), 52–59 (2012). APA
4. Alexandra, O., et al.: CrisisLex: a lexicon for collecting and filtering microblogged communications in crises. In: ICWSM (2014)
5. Ethan, F., Chen, B., Bernstein, M.S.: Empath: understanding topic signals in large-scale text. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM (2016)
6. Niharika, S., Kumaraguru, P.: Call for service: character- izing and modeling police response to serviceable requests on Facebook. In: CSCW (2017)
7. Muhammad, I., et al.: Extracting information nuggets from disaster-related messages in social media. In: ISCRAM (2013)
8. Muhammad, I., et al.: Practical extraction of disaster-relevant information from social media. In: Proceedings of the 22nd International Conference on World Wide Web. ACM (2013)
9. Sudha, V., et al.: Natural language processing to the rescue? extracting “situational awareness” tweets during mass emergency. In: ICWSM (2011)
10. Alex, K., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems (2012)

11. Souvik, K., et al.: Joint acoustic factor learning for robust deep neural network based automatic speech recognition. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE (2016)
12. Tingmin, W., et al.: Twitter spam detection based on deep learning. In: Proceedings of the Australasian Computer Science Week Multiconference. ACM (2017)
13. Duyu, T., et al.: Coooolll: a deep learning system for twitter sentiment classification. In: SemEval@ COLING (2014)
14. Tianqi, C., Guestrin, C.: Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM sigkdd International Conference on Knowledge Discovery and Data Mining. ACM (2016)
15. Muhammad, I., Castillo, C.: Towards a data-driven approach to identify crisis-related topics in social media streams. In: Proceedings of the 24th International Conference on World Wide Web. ACM (2015)
16. Kevin, S., et al.: Identifying and categorizing disaster-related tweets. In: Conference on Empirical Methods in Natural Language Processing (2016)
17. Muhammad, I., et al.: Coordinating human and machine intelligence to classify microblog communications in crises. In: ISCRAM (2014)
18. Abdulfatai, P., et al.: Information verification during natural disasters. In: Proceedings of the 22nd International Conference on World Wide Web. ACM (2013)
19. Zahra, A., et al.: Tweedr: mining twitter to inform disaster response. In: ISCRAM (2014)
20. Alan M, M., et al.: Geo-twitter analytics: Applications in crisis management. In; 25th International Cartographic Conference (2011)
21. Flvio EA, H., et al.: Bridging the gap between decision-making and emerging big data sources: an application of a model-based framework to disaster management in Brazil. Decis. Support Syst. **97**, 12–22 (2017)
22. Matthew S, G.: Predicting crime using Twitter and kernel density estimation. Decis. Support Syst. **61**, 115–125 (2014)
23. Thapen, N., Simmie, D., Hankin, C.: The early bird catches the term: combining twitter and news data for event detection and situational awareness. J. Biomed. Semant. **7**(1), 61 (2016)
24. Kar Wai, L., Buntine, W.: Twitter opinion topic model: Extracting product opinions from tweets by leveraging hashtags and sentiment lexicon. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. ACM (2014)
25. Fred, M., et al.: Understanding twitter data with tweetxplorer. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM (2013)