D-Sieve: A Novel Data Processing Engine for Efficient Handling of Crises-Related Social Messages

Soudip Roy Chowdhury
University of Paris-Sud & Inria
Paris, France
soudeep@gmail.com

Hemant Purohit
Kno.e.sis Center, Wright State
University
Dayton, OH, USA
hemant@knoesis.org

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

ABSTRACT

Existing literature demonstrates the usefulness of system-mediated algorithms, such as supervised machine learning for detecting classes of messages in the social-data stream (e.g., topically relevant vs. irrelevant). The classification accuracies of these algorithms largely depend upon the size of labeled samples that are provided during the learning phase. Other factors such as class distribution, term distribution among the training set also play an important role on classifier's accuracy. However, due to several reasons (money / time constraints, limited number of skilled labelers etc.), a large sample of labeled messages is often not available immediately for learning an efficient classification model. Consequently, classifier trained on a poor model often *misclassifies* data and hence, the applicability of such learning techniques (especially for the online setting) during ongoing crisis response remains limited. In this paper, we propose a *post-classification* processing step leveraging upon *two* additional content features- *stable hashtag association* and *stable named entity association*, to improve the classification accuracy for a classifier in realistic settings. We have tested our algorithms on two crisis datasets from Twitter (Hurricane Sandy 2012 and Queensland Floods 2013), and compared our results against the results produced by a "best-in-class" baseline online classifier. By showing the consistent better quality results than the baseline algorithm i.e., by correctly classifying the misclassified data points from the prior step (false negative and false positive to true positive and true negative classes, respectively), we demonstrate the applicability of our approach in practice.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

Keywords

post processing, classification, hashtag stability, named entity recognition, content analysis, social data analytics

1. INTRODUCTION

During crises, a variety of valuable information is generated on social media platforms such as Twitter by the affected people; by people who may need help (e.g., food, shelter, medical assistance, etc.) or by people who donate or offer services[2, 11, 4]. Moreover, formal disaster response organizations such as government agencies, public health care NGOs, and military play a significant role in disaster response. During such disasters, situation-sensitive requirements arise and formal disaster response agencies look for actionable and tactical information in real-time to effectively estimate early damage assessment, and to launch relief efforts accordingly.

In order to facilitate rapid information access, real-time analysis of social data is essential. Due to high volume and velocity of social media data streams, manual analysis of data is impossible and often cause filter-failure problem [8]. To overcome this problem, hybrid approaches are developed in which humans and machines work together to process high-speed streams (e.g. [5]). One such approach is the supervised classification of social media messages in which machines are trained using a handful of human-provided training samples.

Despite its success as compared to unsupervised approaches, and limitations such as scarce training data and non-uniform class distribution, the classification accuracy of a supervised classification approach can still be improved via exploring content of the messages in-depth. For instance, with the help of named-entity recognition (NER)[13] technique, one can extract embedded entities such as the names of persons, organizations, and locations from a message. Moreover, the presence of hashtags (user-defined topics) and their occurrence within a specific time-frame can be further leveraged to add important meta-information to social media messages. This can ultimately be used to filter important messages. However, techniques such as NER, when applied to high-speed social media data streams (16k tweets/minute recorded during the hurricane Sandy 2012), can suffer from scalability issues and hence, not suitable as a first component in a processing workflow.

In this paper, we introduce a novel yet scalable social media data processing engine called *D-Sieve* . The *D-Sieve* engine leverages content-based features of crisis-related messages, specifically hashtags, and named entities, to improve the classification results of a supervised approach. Concretely, we use the results of a supervised classification approach as an input to our *D-Sieve* engine. The engine builds a stable hashtag set as well as a stable named entity set by

using the true positive and true negative instances in the classified data. Further, the *D-Sieve* engine relies on these stable sets to improve the overall classification accuracy by examining and correctly classifying each instance of the test dataset. We experiment on two crises datasets from Hurricane Sandy in 2012, and Queensland Floods in 2013 to demonstrate the applicability of our approach. Specifically, following are our key contributions:

- We introduce a novel data processing method *D-Sieve* , which can be used as a post-classification component in a real-time supervised classification setting.

- We leverage the use of hashtags in the crisis-related messages and develop a stable hashtag association algorithm.

- We introduce a stable NER association measure that extracts entities from the short text messages, and mentioned URLs and computes similarity with important entities specific to true positive or negative classified behavior.

- With extensive experiments with different settings, we demonstrate the superiority of our system against the baseline classification algorithm.

Rest of the paper is organized as follows. We discuss related work to our approach and social media during emergencies in section 2, preliminaries for the *D-Sieve* engine in section 3 and its functional architecture in section 4, followed by experiments in section 5, and results in section 6 with conclusion in section 7.

2. RELATED WORK

Information collected from social media during a crisis situation can be effective to make appropriate decisions for crisis response, if processed in a timely and correct manner [16]. For this reason, social media information processing has emerged as a hot topic during the last few years and many approaches have been developed, ranging from crisis event detection leveraging burst detection algorithms to classification of messages using supervised and unsupervised techniques (for a detailed survey refer to [4]).

Among other techniques, processing of social media messages using supervised classification techniques has received wide adoption, leveraging various learning algorithms, such as support vector machine (SVM), naive bayesian, and random forest [16, 11][2]. Various factors can effect the classification accuracy of a supervised classification algorithm. For instance, size of the training set, class distribution inside the training set, term distribution over time [9], among others. Moreover, in [14], authors have reported different performances of classifications during various phases of a crisis event (i.e. pre, during, post). In this paper, authors have argued that the nature of the message content changes over the lifetime of an event and there exists an association between the content of a message and the phase of a crisis. Thus, by analyzing the content of a message (n-gram features), one can classify/associate a message to its corresponding crisis phase.

Existing work, as stated above, relies only on the content (n-grams features) of messages that are communicated during a crisis for classifying those messages to an appropriate class. The classification accuracy for most of these existing techniques largely depends upon the *richness* or the quality of the content. In case of tweets, due to the restriction imposed by the platform on the message length, often the message content lacks clarity and richness and hence that affects the classification accuracy of the existing content-based classification algorithm. Moreover, in situations where less numbers of labeled training data available for a classifier or in an online classification scenario, existing content based approach again fail to produce highly accurate classification results. To address these issues, in our approach, we aim to produce better classification results by leveraging features related to the hashtags and the webpage content associated with the URL mentioned in a message. Previous work [3] has used hashtags in a message in order to understand the sentiment of a message, while in [1] authors have further explored the usage of hashtags during different crisis situations and have further claimed that by looking at the hashtag usages in a message, one can reason about the crisis instance. Likewise, classification technique for messages based on the named entity in the message content is also explored in the existing literature. One of such recent work on NER techniques that focuses on classifying social messages by utilizing entity knowledge bases is described in [13]. Existing knowledge-bases for NER are also available as a service, such as OpenCalais[12]. However, the challenges with only NER-based classification approach arise while handling the usage of informal English language in the social media messages. In our approach we leverage both the hashtag and NER features for designing an efficient post-classification processing engine, *D-Sieve* .

Next, we explain the detailed algorithms and design of *D-Sieve* .

3. PRELIMINARIES

We represent social media messages as a tuple $\langle H, P, U, T \rangle$, consisting of hashtags H, text content P, URL U and the time of post T for messages. Apart from T, other features are optional for a message i.e., they may or may not occur all together in a message. In our approach, we have hypothesized that there exists an implicit association among messages for a crisis type/subtype and its corresponding features. Based on this hypothesis, we have designed our system, as a *post-classification* processing step. This system leverages two novel features, *stable hash-tag association* and *stable named entity association* to improve the accuracy of a baseline supervised learning classification algorithm. We further argue that the algorithms in the *D-Sieve* are generic enough to be applicable in different settings (both for the *online* and *offline* classification scenario) of a supervised classification technique. Before we explain about the algorithms behind the *D-Sieve* , let us first explain the preliminaries behind the two feature metrics that we exploit in *D-Sieve* .

3.1 Stable Hashtag Association

Similar to [15], in *D-Sieve stable hashtags set* represents the set of most frequent occurring hashtags (tags) among all messages for a crisis. The relative frequency for the occurrence of any item in a stable hashtag set does not vary over time. Suppose k messages $o_1, o_2, .., o_k$ are posted in time sequence and H_i represents a set of hashtags per ith message. Based on the relative frequency of occurrences of hashtags among k messages, the *relative hashtag frequency distribu-*

tion or *rfd* is determined. Based on the moving average score of *rfd*s $F(H_i)$ for all tags H in the k message sequence, we define the *stable rfd*, which measures how stable is the tag assignments for a crisis over its k messages.

DEFINITION 1. *Given a parameter ω, such that $2 \leq \omega \leq (k-1)$. ω represents the moving average score of the hashtags frequencies for a crisis instance, after it has received k messages, denoted by*

$$m(k,\omega) = \frac{1}{\omega} \sum_{j=k-\omega+2}^{k} s(F(j-1), F(j))$$

where s is the cosine similarity *between two consecutive* rfd*s.*

A set ϕ contains all hashtags among the k messages for which the value of $m(k,\omega)$ is greater than some predetermined threshold. We call ϕ, as the stable hashtag set for the k messages. Given these priors, we define the hashtag-based stability metrics for an input message as follows:

DEFINITION 2. *If ϕ_{TP} and ϕ_{TN} denote the sets of stable hashtags with respect to the true positive and true negative classes respectively for messages related to a crisis, then the* stable hashtag association metric *for an input message $\langle H_i, P_i, U_i, T_i \rangle$ is denoted by V_H. Such that*

$$V_H = \frac{|(\phi_{TP} \backslash \phi_{TN}) \cap T_i| - |(\phi_{TN} \backslash \phi_{TP}) \cap T_i|}{|T_i|}, \text{ where } -1 \leq V_H \leq 1$$

DEFINITION 3. *If the value of V_H is found to be > 0, then the current input message is tagged as* true positive*, if V_H is found to be < 0 then the current message is tagged as* true negative*, and if $V_H = 0$, we don't ascertain any class label to the current message based on its stable hashtag association metrics value.*

3.2 Stable Named Entity Association

Named entities in the content of a message provides the ability to interpret contextual meaning of a message, in contrast to merely processing word n-grams from the text as feature without understanding their meaning. We identify named entities in the text content P of each message in our corpus using existing knowledge base. In the design of *D-Sieve* , we have used OpenCalais [12] as our knowledge base for identifying entities. Further, we resolve URL U mentioned in the message and crawl its webpage. We identify entities in the title and content of the webpage also.

DEFINITION 4. *Suppose, E_{TP} denotes a set of entities spotted in a message and any resolved URL content among all the messages belonging to the true positive class, and similarly, E_{TN} corresponding to the true negative class. Suppose I_E denotes the set of entities that are extracted from the current message text as well as from the title and content of any existing URL in that. The stable named entity association metrics V_E is defined as follows:*

$$V_E = \frac{|(E_{TP} \backslash E_{TN}) \cap I_E| - |(E_{TN} \backslash E_{TP}) \cap I_E|}{|I_E|}, \text{ where } -1 \leq V_E \leq 1$$

DEFINITION 5. *If the value of V_E is found to be > 0, then the current input message is tagged as* true positive*, if V_E is found to be < 0 then the message is tagged as* true negative*, and if $V_E = 0$, we don't ascertain any class label to a message based on its value of stable named entity association metrics.*
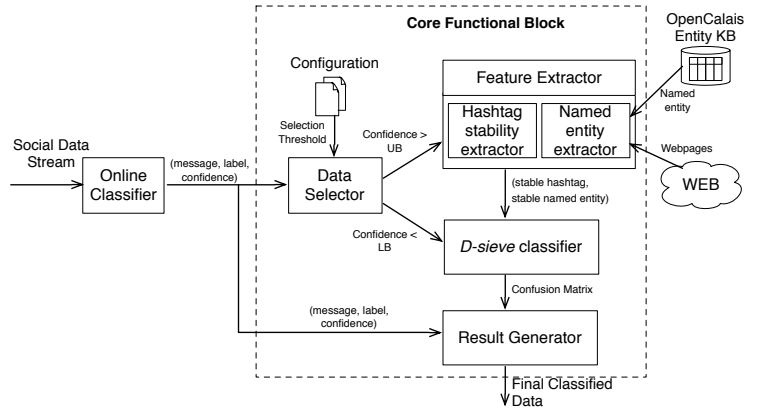

Figure 1: Functional architecture of *D-Sieve*

3.3 Weighted Aggregated Feature Score

DEFINITION 6. *Finally, we calculate the weighted average of these two feature association metrics to get the final aggregated score S_C for a message. S_C for an input message is defined as follows:*

$$S_c = W_H . V_H + W_E . V_E, \text{ where } -1 \leq S_c \leq 1$$

The features weights (W_H, W_E) are chosen to restrict the value of S_c to the range [-1, 1]. As a rule of thumb, in D-Sieve we set the values for W_H and W_E as 0.5.

DEFINITION 7. *If the value of S_c is found to be > 0, then the current input message is tagged as* true positive*, if the S_c is found to be < 0 then the message is tagged as* true negative*, and if $S_c = 0$, we don't ascertain any class label to the message.*

4. FUNCTIONAL ARCHITECTURE

D-Sieve uses the above two feature metrics for classifying messages, which otherwise would have misclassified by a machine-learning classifier. In other words, *D-Sieve* extracts features as explained in Section 3 from the true positive (TP) and true negative (TN) results of the classifier, and uses these features to correctly classify false positive (FP) and false negative (FN) results into true positive or true negative classes. Figure 1 depicts the core functional blocks for the *D-Sieve* .

The *Online classifier* classifies messages collected from social media data streams (e.g. Twitter streaming API) into one or more pre-defined classes with the help of a learned model. The output of the online classification step consists of the original message, classifier's predicted label/class for the message and the classifier's confidence. The confidence value conveys the confidence level of the classifier in predicting a label for the given input message.

Next, the *data selector* takes as input the result produced by the classifier and a configuration file that consists of *selection threshold* parameters. Selection threshold parameters contain both the upper bound (UB) and the lower bound (LB) for the classification confidence values, which are consulted by the data selector module to decide whether classifier results should be used for the feature extraction or for further label determination by *D-Sieve* algorithms. Data points (both TP and TN results) for which the classification
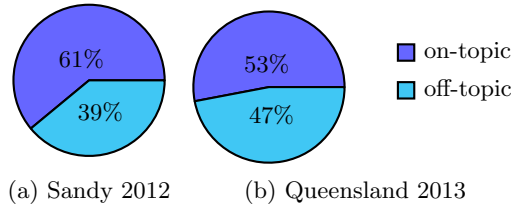
(a) Sandy 2012   (b) Queensland 2013

Figure 2: Distributions of labels among 10k sample tweets

confidence values are greater than the value of UB selection threshold, are considered as an input for the *feature extractor* module. *D-Sieve* considers data points (FP, FN, TP and TN with classification confidence < LB) having confidence value less than the LB selection threshold, as the candidate set for improving classification accuracy.

The *feature extractor* determines the stable hashtag set and the stable named entity set (using the external knowledge base OpenCalais) from the TP and TN data points. *D-sieve classifier* takes both the stable hashtag set, and the stable named entity set, using which it processes the data points in the candidate set for improvement and calculates the corresponding feature metrics as shown in the Section 3. Based on the weighted aggregated scores, a suitable classification label is assigned to an input message of the candidate set. Finally, *result generator* aggregates the new labels produced by the *D-Sieve* classifier with the old one and produces the final classification result.

5. EXPERIMENTAL SETUP

Although, algorithms in the *D-Sieve* are agnostic to underlying social media platform, due to unavailability of labeled data set from other social media platforms, in this paper we have experimented with datasets from Twitter.

We use CrisisLexT6 [10] dataset collected during two different crises. The first dataset corresponds to the tweets collected using keywords (hurricane, hurricane sandy, frankenstorm, #sandy) during the Hurricane Sandy in 2012. The second dataset corresponds to the tweets collected using keywords (#qldflood, #bigwet, queensland flood, australia flood) during the QueensLand Floods in Australia in 2013.

Each dataset contains 10k labeled tweets, which were selected uniformly at random from the raw tweets collections. Labeling is performed using the Crowdflower crowdsoucing platform, where crowd workers were asked to examine each tweet message and to determine whether a tweet is *on-topic* or *off-topic*. Figure 2 shows the distribution of labels in both datasets.

5.1 Supervised classification

Although the algorithms in the *D-Sieve* are well applicable to both online and offline classification settings, for the purpose of our experiments, we chose online supervised classification setting.

In order to mimic an online setting, we divided a labeled crisis dataset into two sets having 5k tweet messages in each. We trained a classifier using 80% as training and 20% as test set from the first set of 5k labeled messages. We use InfoGain with Ranker algorithm to select *top 800 content based features* (uni-gram and bi-gram), and used the Random Forest [7] as our classification algorithm to learn the

best-fit models. The AUC score for Sandy is 78% and 83% for Queensland. The choice of our classification algorithm is influenced by its best performance in various classification task in crisis domain in the past [11, 6].

Further, the trained model is used to classify the second set of tweet messages (i.e. the remaining 5k tweets) for each crisis dataset. Like real-time online streaming setting, each tweet from the second set is processed one-by-one through the corresponding classifier (trained on the first set) to obtain a label (on-topic or off-topic) with a confidence score.

5.2 Configuration settings for the D-Sieve

In our experimentation setup, the following configuration parameter settings have produced the optimal results when tested against both the dataset. We describe them below:

Upper and lower bound for classifier's confidence are set to select the best representative subset of the dataset that are used for the feature extraction and for the re-labeling purposes via the proposed approach. Based on our observation on the online classifier's output for both the crises datsets, we have fixed the upper bound for the confidence interval as 80% and the lower bound as 70%. Messages (true positives as well as true negatives) for which the classifier confidence is more than 80% are considered for extracting the stable hashtag and named entity association features, while the other subset of the messages (true positives or true negatives with confidence less than 70% and false negatives or false positives) are subjected to the re-labeling process by the *D-Sieve* classification algorithm.

Cosine similarity threshold for hashtag stability is set to select the most stable and representative subset of hashtags among the true positive and true negative dataset used by the *D-Sieve* during feature extraction. In our settings, we have found that the optimal sets of stable hashtags (containing 24 (0) for Sandy, and 21 (7) for Queensland in the true positive (negative) set respectively) are retrieved from the input messages, when the correlation co-efficient are set to 0.997 and 0.752 for true positive and true negative sets respectively, for both the crisis datasets.

Threshold for feature scores; as mentioned in the Section 3, the threshold for the feature scores (V_H and V_E respectively) are set to 0. That is, if they are found to be > 0, then the current input message is tagged as *true positive*, if the S_c is found to be < 0 then the message is tagged as *true negative*, and if it is 0, then the *D-Sieve* does not ascertain any class label to the message.

6. RESULTS

Among the selected subset of 5k tweet messages for each of the crisis dataset, the part of data that satisfies the upper and lower bound of classification constraints is found to be only 68% and 70% of the 5k tweets for Sandy and Queensland datasets, respectively. From the selected subset of the data, we have chosen the TP and TN class samples having the classification confidence value > 80% to extract the stable hashtag and stable named entity sets. Based on the extracted features we process each of the message in the candidate set for improvement i.e., TP and TN classes having classification confidence < 70% and also the FP and FN classes.

Figure 3

Predicted Label

(a) Results of online classification on Sandy dataset

Actual Label	on-topic	off-topic
on-topic	2682	612
off-topic	109	0

(b) Results of online classification on Queensland dataset

	on-topic	off-topic
on-topic	1294	2227
off-topic	6	11

(c) Results of online classifier's + stable hashtag association feature on Sandy

	on-topic	off-topic
on-topic	2784	510
off-topic	109	0

(d) Results of online classification + stable hashtag association on Queensland

	on-topic	off-topic
on-topic	2059	1462
off-topic	4	11

Predicted Label

(e) Results of online classification + stable named entity association feature on Sandy

Actual Label	on-topic	off-topic
on-topic	2851	443
off-topic	109	0

(f) Results of online classification + stable named entity association feature on Queensland

	on-topic	off-topic
on-topic	2426	1095
off-topic	6	11

(g) Results of online class-action + stable hashtag association+ stable named entity association features on Sandy

	on-topic	off-topic
on-topic	2884	410
off-topic	109	0

(h) Results of online classification + stable hashtag association+ stable named entity association features on Queensland

	on-topic	off-topic
on-topic	2614	907
off-topic	5	12

Figure 3: Confusion matrices with different experiment settings
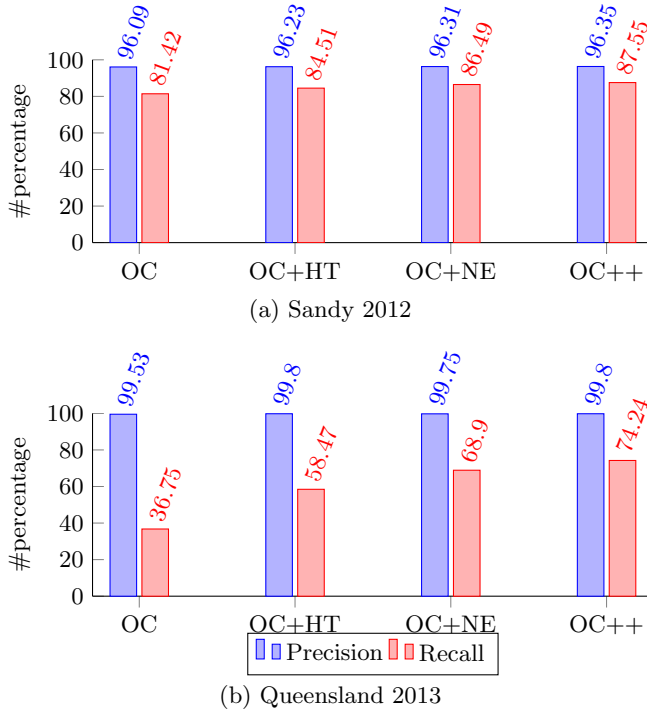

(a) Sandy 2012

(b) Queensland 2013

Figure 4: Classification results with and without D-Sieve algorithms

Figure 3 shows the confusion matrices derived from the results of the classifications without and with the *D-Sieve* algorithm. It is quite evident from the figure that both the features that we use in the *D-Sieve* are proven to be quite useful in finding out the TP class samples from the FN classified messages. However, it is to be noted that due to the sparsity of hashtags and URLs among the data along with lack of TN classes in our candidate improvement set, we could not establish the value of our algorithms in determining the TN class from the FP classified samples (especially for the Sandy dataset). The reason behind the exceptionally good results of our algorithm for the Queensland dataset are the uniform distribution of hashtags and URL among tweets, which helped *D-Sieve* algorithm to determine the true classes for the candidate test set based on their features. It is also to be noted that the reason behind the consistent superior quality results of classification with stable named entity association feature is due to the fact that with enhanced content (tweet message content + webpage content corresponding to the URL in the message), the classifiers can derive a profound semantically relevant entities for associations present in tweet messages, and thus can classify content with higher precision.

Figure 4 shows the accuracy results of classification algorithms with and without the *D-Sieve* . In the Figure we use the abbreviations OC, OC+HT, OC+NE, OC++ to indicate online classification, online classification with stable hashtag association, online classification with stable named entity association and online classification with weighted aggregate of both stable hashtag and named entity feature association, respectively. It is to be noted that in our experiments we have used the "best-in-class" classifier (cf. Section 5.1), which has produced classification result with over 96% and 99% precisions for the subset of our dataset from the Sandy and Queensland datasets respectively, the accuracy improvements for precisions are not evident from the result.

However, since our algorithms in the *D-Sieve* is able to correctly classify FN classes to TP, we could improve drastically the recall values for both the datasets (over 6% for the Sandy and over 37% for the Queensland dataset). It is to be underlined that in this paper, the efficiency of algorithm is reflecting the performance of our system with a sparse dataset (with non-uniform distributions of hashtags and URLs) in an online classification setting. In an offline setting with more uniform distribution of hashtags and URL among the messages, this efficiency metrics might go even higher.

7. CONCLUSION

In this paper, we have described the *D-Sieve* to enhance the supervised classification techniques for social media messages, especially during crises, by demonstrating the added values in improving results for the two crisis datasets, Hurricane Sandy and Queensland Floods. Our algorithms involve exploitation of content characteristics in the social media messages to derive two new features, *s*table hashtag and *s*table named entity association. The stable hashtag association captures the stable average sets of hashtags corresponding to true positive and negative classes, and then find their similarity in a candidate message that is likely misclassified (false positive or false negative). Similarly, the stable named entity association feature computes the similarity of entities spotted in the candidate misclassified message with the entities in the true positive class samples and true negative class samples. We use external knowledge base for entity spotting that helps in associating meaning to the special tokens in the message, beyond considering them at the syntactic level as word n-grams. Our results showed a consistent trend of improvements (over 6 and 37% higher recall values for Sandy and Queensland datasets) over the accuracies produced by a "best-in-class" online classifier.

Future work. In future, we plan to add deeper semantic similarity of entities between candidate misclassified messages and true positive/negative sets. We also plan to extend *D-Sieve* experiments for the online/offline settings for multiclass classifications as well as temporal implications that lead to different data size to train initial/base classifier. We will investigate performance improvement when the dataset size is low, and also latency issue for feature extraction process (specifically URL based) by using parallelization. Further, we also wish to experiment our algorithms more on other social media datasets such as Facebook and G+.

8. REFERENCES

[1] A. Bruns and Y. E. Liang. Tools and methods for capturing twitter data during natural disasters. *First Monday*, 17(4), 2012.

[2] M. A. Cameron, R. Power, B. Robinson, and J. Yin. Emergency situation awareness from twitter for crisis management. In *Proceedings of the 21st international conference companion on World Wide Web*, pages 695–698. ACM, 2012.

[3] D. Davidov, O. Tsur, and A. Rappoport. Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 241–249. Association for Computational Linguistics, 2010.

[4] M. Imran, C. Castillo, F. Diaz, and S. Vieweg. Processing social media messages in mass emergency: A survey. *arXiv preprint arXiv:1407.7071*, 2014.

[5] 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.

[6] 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.

[7] A. Liaw and M. Wiener. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[8] N. Morrow, N. Mock, A. Papendieck, and N. Kocmich. Independent evaluation of the ushahidi haiti project. *Development Information Systems International*, 8:2011, 2011.

[9] F. Mourão, L. Rocha, R. Araújo, T. Couto, M. Gonçalves, and W. Meira Jr. Understanding temporal aspects in document classification. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pages 159–170. ACM, 2008.

[10] A. Olteanu, C. Castillo, F. Diaz, and S. Vieweg. Crisislex: A lexicon for collecting and filtering microblogged communications in crises. *Proceedings of ICWSM*, 2014.

[11] H. Purohit, C. Castillo, F. Diaz, A. Sheth, and P. Meier. Emergency-relief coordination on social media: Automatically matching resource requests and offers. *First Monday*, 19(1), 2013.

[12] T. Reuters. Opencalais web service, 2009.

[13] A. Ritter, S. Clark, Mausam, and O. Etzioni. Named entity recognition in tweets: An experimental study. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 1524–1534, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

[14] 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.

[15] X. S. Yang, D. W. Cheung, L. Mo, R. Cheng, and B. Kao. On incentive-based tagging. In *Proceedings of the 2013 IEEE International Conference on Data Engineering (ICDE 2013)*, ICDE '13, pages 685–696, Washington, DC, USA, 2013. IEEE Computer Society.

[16] J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power. Using social media to enhance emergency situation awareness. *IEEE Intelligent Systems*, 27(6):52–59, 2012.