International Journal of Information Systems for Crisis Response and Management
Volume 11 • Issue 2 • July-December 2019

A Hybrid Domain Adaptation Approach for Identifying Crisis-Relevant Tweets

Reza Mazloom, Kansas State University, Manhattan, USA
Hongmin Li, Kansas State University, Manhattan, USA
Doina Caragea, Kansas State University, Manhattan, USA
Cornelia Caragea, University of Illinois at Chicago, Chicago, USA
Muhammad Imran, Qatar Computing Research Institute, Ar-Rayyan, Qatar

ABSTRACT

Huge amounts of data generated on social media during emergency situations is regarded as a trove of critical information. The use of supervised machine learning techniques in the early stages of a crisis is challenged by the lack of labeled data for that event. Furthermore, supervised models trained on labeled data from a prior crisis may not produce accurate results, due to inherent crisis variations. To address these challenges, the authors propose a hybrid feature-instance-parameter adaptation approach based on matrix factorization, k-nearest neighbors, and self-training. The proposed feature-instance adaptation selects a subset of the source crisis data that is representative for the target crisis data. The selected labeled source data, together with unlabeled target data, are used to learn self-training domain adaptation classifiers for the target crisis. Experimental results have shown that overall the hybrid domain adaptation classifiers perform better than the supervised classifiers learned from the original source data.

KEYWORDS

Crisis Response, Domain Adaptation, K-Nearest Neighbors, Matrix Factorization, Self-Training, Tweet Classification

INTRODUCTION

Social media is becoming a more prevalent part of our everyday life, due to the advancements in technology and virtualization. The availability of the Internet, cameras and real-time message boards at our fingertips has brought about live and parallel reporting, and witness testimonies during many events. These reports can be useful to responders and can help create awareness among the populace, especially in emergency situations (Meier, 2015; Watson, Finn, and Wadhwa, 2017). Despite the potential benefits, major response groups and organizations under-utilize these sources of information, as therein lie many administrative and technical challenges (Meier, 2013). Among the challenges, there are reliability issues associated with public and unstructured data, as well as

DOI: 10.4018/IJISCRAM.2019070101

information overload issues, as millions of messages are posted during a crisis situation (Bullock, Haddow, and Coppola, 2012).

There are many recent studies that propose the use of machine learning techniques to provide automated methods for analyzing social media data to reduce the information overload (Imran et al., 2015; Beigi et al., 2016). Machine learning techniques can help transform raw data into usable information by labeling, prioritizing and structuring data, and making them beneficial to responders and to the populace in times of need (Qadir et al., 2016). However, supervised learning algorithms rely on labeled training data to build predictive models. Accurate labeling of data for an emerging crisis is both time consuming and expensive, and, hence, it is not appropriate to assume that labeled data for a current crisis will be promptly available to be used for analysis. The lack of labeled data for emerging crisis events prohibits the use of supervised learning techniques.

To address this challenge, several works proposed to use labeled data from prior "source" crises to learn supervised classifiers for a "target" crisis (Verma et al., 2011; Imran et al., 2013; Imran, Mitra, and Srivastava, 2016). However, due to the divergence of each crisis in terms of location, nature, season, etc. (Palen and Anderson 2016), the source crisis might not accurately represent the characteristics of the target crisis (Qadir et al., 2016; Imran et al., 2015). Domain adaptation techniques (Pan and Yang, 2010; Jiang, 2008) are designed to circumvent the lack of labeled target data by making use of unlabeled target data as guideposts for the readily available labeled source data. Studies in the emergency space have shown that using domain adaptation techniques, which use target unlabeled data and source labeled data together, significantly improve classification results as compared to supervised techniques that solely use labeled source data (Li et al., 2015, 2017). Unlabeled data from the target crisis becomes more abundant as the event unfolds, and it can enable the use of domain adaptation techniques during emerging or occurring crisis events.

There are several ways in which the unlabeled target data can be used with domain adaptation techniques, including parameter-based adaptation, instance-based adaptation and feature-based adaptation (Pan and Yang, 2010). In the parameter-based adaptation, the labeled source data is used together with the unlabeled target data to identify shared parameters that result in good predictions for the target data. In the instance-based adaptation, the unlabeled target data is used to identify and/or reweigh the most relevant source labeled instances with respect to the target classification task, while in feature-based adaptation, the target unlabeled data and source labeled data are used together to find a feature representation that minimizes the difference between the two domains. Relevant prior work on crisis tweet classification using domain adaptation has relied on parameter-based adaptation. Specifically, Li et al. (2017) proposed to learn weighted source and target Naïve Bayes classifiers with the iterative methods of Expectation-Maximization (EM) (Dempster, Laird, and Rubin, 1977) and Self-Training (ST) (Yarowsky, 1995) and showed that the resulting classifiers can accurately predict the target.

Mazloom et al. (2018) proposed to use a combination of two domain adaptation approaches, specifically a hybrid between feature-based adaptation and instance-based adaptation, to reduce the variation between the two domains. First, the Alternating Nonnegative Least Squares Matrix Factorization (LSNMF) in (Lin, 2007) is used on the combined source and target data, represented using binary vectors, to create a dense and reduced conceptual representation of source and target instances. Subsequently, the k-Nearest Neighbors algorithm (kNN) is used to select a subset of the source instances which are most similar to the target instances, according to the cosine similarity calculated based on the reduced common representation. Finally, the selected subset is subsequently used to learn Naïve Bayes classifiers for the target crisis.

In this study, we propose to use a combination of all three domain adaptation approaches, specifically a hybrid between feature-based adaptation, instance-based adaptation, and parameter-based adaptation to reduce the variation between the source and target domains. As in Mazloom et al. (2018), we first use the Alternating Nonnegative Least Squares Matrix Factorization (LSNMF) Lin (2007) on the combined source and target data to create a dense and reduced conceptual representation

of source and target instances. Subsequently, the k-Nearest Neighbors algorithm (kNN) is used to select a subset of the source instances which are most similar to the target instances, according to the cosine similarity calculated based on the reduced common representation. Finally, the parameter-based adaptation approach used in Li et al. (2017) is used on the selected subset of the source labeled data combined with the available target unlabeled data. The objective is to gain an understanding of the benefits provided by the hybrid feature-instance-parameter adaptation approach, as compared to the independent feature-instance and parameter-based adaptation approaches.

As an application, we focus on the task of classifying tweets as being relevant to the event of interest or not relevant. This is one of the most basic but crucial classifications needed during a crisis event, as subsequent analysis should be done only on data relevant to the crisis in question. Furthermore, this classification is not trivial: supervised classifiers may not achieve accurate results due to domain variations.

To summarize, our main contributions are as follows:

- We extend the hybrid feature-instance adaptation approach proposed in Mazloom et al. (2018) to a hybrid feature-instance-parameter adaptation approach, where self-training is used with the feature-instance adapted source data to transfer parameters from the source crisis to the target crisis;
- As opposed to Mazloom et al. (2018), where experiments were performed only on the CrisisLexT6 (Olteanu et al., 2014a) dataset, we perform an extensive set of experiments on pairs of source-target crisis events from two datasets, CrisisLexT6 (Olteanu et al., 2014a) and 2CTweets (Schulz, Guckelsberger, and Janssen, 2017). The goal is to evaluate the feature-instance adaptation approach by comparison with approaches that make use of either feature-based adaptation or instance-based adaptation, but not both;
- We study the variation of performance with the parameters of the feature-based adaptation (specifically, the number of features, f), and instance-based adaptation (specifically, the number of neighbors, k), respectively, to identify parameters that result in good overall performance.

METHODS

The goal of this study is to design and evaluate a hybrid feature-instance-parameter adaptation approach by combining two adaptation approaches that have been successfully used in the context of classifying crisis-related tweets, specifically a feature-instance adaptation approach proposed by Mazloom et al. (2018), with the self-training adaptation approach proposed by Li et al. (2017). As a base classifier, both Mazloom et al. (2018), and Li et al. (2017) have used Naïve Bayes, a simple but powerful classifier, which does not have any tunable hyper-parameters. Furthermore, Li et al. (2017) have shown that simple Bernoulli Naïve Bayes classifiers are very appropriate for analyzing crisis-related data posted on social media, as they give results competitive with those of more sophisticated algorithms which have tunable hyper-parameters. However, other studies in the crisis domain have successfully used Random Forest (Imran, Mitra, and Srivastava 2016), an ensemble-type classifier, which is generally known to reduce the variance (Genuer 2012). Given the success of both Naïve Bayes and Random Forest classifiers, in this study we investigate both classifiers in the context of the hybrid feature-instance-parameter adaptation approach.

We will next review the feature-instance adaptation approach introduced in (Mazloom et al. 2018), and then describe the combination of this approach with a self-training approach, similar to the one proposed in Li et al. (2017).

Feature-Instance Adaptation Approach

Given a source and target pair of crises, the goal is to adapt the source data by reducing the variance with respect to the target data, and then train a Naïve Bayes or Random Forest classifier on the adapted

source data. The source adaptation is guided by the target unlabeled data. More specifically, a hybrid feature-instance adaptation approach is used to select a subset of the source instances, which are most similar to the target instances. First, the target instances are used to construct a target vocabulary *V*, which is used to represent both source and target data as bag-of-words binary vectors. As part of the feature adaptation step, the resulting data matrix *D* is decomposed using the popular Least Squares Non-Negative Matrix Factorization (LSNMF) proposed by Lin (2007). The implementation of this method is available in Python under the "nimfa" package. Intuitively, the decomposition will produce a reduced dense representation of the data, which is more suitable for identifying similar instances as compared to the sparse binary representation (Guo and Diab 2012).

As part of the instance adaptation step, the reduced representation is used to identify source instances that are most similar to the target. More precisely, for each target (unlabeled) instance, we calculate the cosine similarity to the source instances and select the *k* nearest neighbors from the source. If two different target instances have the same source instance among the *k* nearest neighbors, the selected subset of the source may contain duplicate instances. Mazloom et al. (2018) experimented with two settings, one in which duplicate neighbors were retained (i.e., seen as reweighing source instances that are close to many target instances), and another one in which duplicates are removed (so that each instance is used with the same weight). Experimental results showed that the setting where duplicates are retained gives better results overall, and thus, we only use this setting in the current study. Mazloom et al. (2018) also demonstrated that results from the joint feature-instance adaptation are superior to that of their standalone versions and provide better consistency throughout different datasets. Our experimental results regarding this matter had similar results to their work hence, only the joint adaptation results are discussed further in this study.

Finally, Naïve Bayes and Random Forest algorithms, respectively, are used to learn classifiers from the selected subset of the source. For Naïve Bayes, Mazloom et al. (2018) also experimented with two settings: one in which the Gaussian Naïve Bayes algorithm is used on the reduced representation of the selected source instances, and another one in which the Bernoulli Naïve Bayes algorithm is used on the original binary representation of the selected source instances. Experimental results showed that the Bernoulli Naïve Bayes classifiers are better than the Gaussian Naïve Bayes classifiers. In preliminary examination with Random Forest, we observed similar results. Therefore, in this study, we train Bernoulli Naïve Bayes and Random Forest classifiers on the original binary representation of the selected source instances. The resulting classifiers are evaluated on a separate target test dataset.

Self-Training Adaptation

Self-training is an iterative parameter-based approach for adapting a source classifier to a target event. Initially, a classifier is learned from the source labeled data, *SL*. At each iteration, the current classifier is used to label the available unlabeled target data *TU*, and the most confidently labeled target instances *n* in each class are used to update the classifier. This process is repeated for a fixed number of iterations. A parameter γ controls the contribution/weight of the source instances as opposed to the target instances at each iteration *i*. Intuitively, with each iteration, *i*, the weight gradually shifts from the source instances towards the target instances.

The classifier adaptation works naturally for Naïve Bayes, where the probability estimates can be easily adjusted as new training data becomes available. However, this is not the case for Random Forest classifiers, where the structure of a tree might be altered when new training data is added. Instead of changing the existing trees, or building the whole random forest classifier from scratch, it is preferable to add new trees to the existing classifier. This is the approach followed in this study. This approach is preferable to rebuilding the entire forest which significantly reduces efficiency and scalability for larger datasets.

Combining Feature-Instance Adaptation With Self-Training

We propose a hybrid feature-instance-parameter adaptation approach by combining the feature-instance adaptation with a variant of the self-training approach used by Li et al. (2017), as described above. Specifically, instead of directly using the Bernoulli Naïve Bayes and Random Forest classifiers on the selected source labeled data, the iterative process of self-training is used to adapt the parameters of the classifiers learned from the selected source, reverted to the binary representation, based on target unlabeled data. In the case of the feature-instance adaptation, the resulting classifiers are evaluated on a separate target test data.

The hybrid feature-instance-parameter approach is summarized in Algorithm 1.

Algorithm 1. Hybrid feature-instance-parameter adaptation

```
Input: Source labeled data, SL, target unlabeled data, TU, and
       target test data, TT.
1: Construct a vocabulary, V, using the target unlabeled data, TU.
2: Represent source, SL, and target, TU, as V -dimensional binary
   vectors, and create the combined source and target data matrix, D.
3: Feature adaptation: Perform the Least Squares Non-Negative
   Matrix Factorization (LSNMF) on the combined source and
   target data matrix, D, resulting in a reduced f -dimensional
   representation of the paired events.
4: Instance adaptation: Using the reduced representation, for each
   target instance in TU, find the k nearest source instances in
   SL, and add them (binary representation) to the adapted source,
   Adap- SL, while retaining duplicates.
5: Self-training: Use the adapted source data, Adap- SL, and
   target unlabeled data, TU (binary representation), to
   iteratively learn a target classifier.
6: Evaluate the resulting hybrid classifier on the target test
   data, TT.
```

Dataset

This study will use two datasets to evaluate the proposed hybrid adaptation approach. The first dataset, CrisisLexT6 (T6) (Olteanu et al., 2014b), is a collection of tweets collected from six disasters that occurred in United States, Canada and Australia between the period of October 2012 and July 2013. The dataset contains approximately 10,000 tweets per disaster labeled as related to a disaster ("on topic") or not ("off topic"). The second dataset, 2CTweets (2C) (Schulz, Guckelsberger, and Janssen, 2017), is a collection of tweets about incidents, such as car crash, fire or shooting, which happened in 10 different cities. Tweets were labeled as incident related (Yes) or not (No).

The datasets were originally collected though the Twitter API using keywords, geographic location and the affected areas of their respective crises. The data was pre-processed using the procedure described in Li et al. (2015), which comprises of replacing URLs, usernames, and emails with placeholders, as well as removing non-printable ASCII characters, re-tweets, and duplicate tweets. Twelve source-target pairs are used for each dataset in the experiments. For the CrisisLexT6 dataset, the events are paired based on chronological order and the pairs are selected to match the pairs in (Mazloom et al., 2018). For the 2C dataset, each event in a dataset is paired with all the other events, and a subset of pairs is selected as follows: First, pairs were ranked based on the accuracy results produced by the supervised Bernoulli Naïve Bayes classifier learned from the original source. Subsequently, a set of pairs was selected to ensure a wide range of accuracy values, while avoiding

using the same event as source or as target in too many pairs. The goal was to obtain a subset of pairs that is representative for the set of all possible pairs.

Each pair was converted to a binary representation using the target vocabulary, which consists of approximately 1000 words on the average, after selecting words that are repeated at least ten or four times for T6 and 2C, respectively, within source and target combined. These parameters were primarily chosen based on prior work (Li et al., 2015; Mazloom et al., 2018) while considering sufficient usage and feature loss minimization. This would result in pairs with about 1000 features for analysis. Details about the events in each dataset and number of instances/features in each event are shown in Table 1. The specific pairs of events used for the two datasets are shown in Table 3.

Table 1. List of the events in the CrisisLexT6 dataset (top) and the 2CTweets dataset (bottom) used in the experiments, together with information about the number of instances/features in each event.

CrisisLexT6 Crisis		Instances			Features
Abbreviation	Disaster	Related	Not Related	Total	
SAN	Sandy Hurricane	3870	6138	10008	1380
QUE	Queensland Floods	4619	5414	10033	1242
BOS	Boston Bombings	4364	5648	10012	1317
OKL	Oklahoma Tornado	5165	4827	9992	1143
WES	West Texas Explosion	4760	5246	10006	1239
ALB	Alberta Floods	4841	5186	10030	1322
2CTweets Crisis		**Instances**			**Features**
Abbreviation	**Disaster**	**Related**	**Not Related**	**Total**	
BOS	Boston Bombings	604	2216	2820	1648
BRI	Brisbane	698	1898	2587	1287
CHI	Chicago	214	1270	1484	862
DUB	Dublin	199	2616	2815	1384
LON	London	552	2444	2996	1673
MEM	Memphis	361	721	1082	771
NYC	NYC	413	1446	1859	1119
SAN	SanFrancisco	304	1176	1480	935
SEA	Seattle	800	1404	2204	1375
SYD	Sydney	852	1991	2843	1601

Table 2. List of the selected pairs used for the analysis results in the CrisisLexT6 dataset (top) and 2CTweets dataset (bottom) used in the experiments

Pairs	CrisisLexT6											
Source	BOS	BOS	BOS	OKL	QUE	QUE	QUE	SAN	SAN	SAN	SAN	SAN
Target	ALB	OKL	WES	ALB	ALB	BOS	OKL	ALB	BOS	OKL	QUE	WES
Pairs	**2CTweets**											
Source	SYD	LON	DUB	BRI	CHI	SAN	NYC	SEA	LON	BRI	LON	CHI
Target	DUB	CHI	SYD	SAN	BOS	LON	MEM	NYC	MEM	BOS	SEA	BRI

Table 3. Queensland Flood related example tweets from the CrisisLexT6 dataset where Bernoulli Naïve Bayes classification after adaptation has correctly classified the tweets contrary to the baseline (Original). f denotes instance adaptation only, k denotes feature adaptation, and f - k denotes feature-instance adaptation. The check symbol (✓) signifies correct classification while the cross symbol (×) signifies misclassification. Parameter adaptation for the examples below had similar results.

CrisisLexT6 Tweets	*Original*	*f*	*k*	*f - k*
USERNAME Dear AziaddictsAU and friends down under in Queensland. It's all over the news once again that the worst flood... URL ...	×	✓	×	×
USERNAME My thoughts go out to all of the flood and fire victims around Australia, stay strong	×	×	✓	×
DTN Japan: Gladstone flood victims returning home: Floodwaters are receding at Gladstone, in central Queensland,... URL	×	×	×	✓
#Australia | #Queensland | #Flood | #Tornado | Tape your windows in a cross-shape; windows are shattering due to pressure. Don't sit close.	×	×	✓	✓
News Flood crisis unfolding in Lockyer Valley: A serious flood crisis is emerging in Queensland's Lockyer Valle... URL	×	×	×	×

In each pair, the first event will be used as the source to adapt, and the second will be used as the target to classify.

Experimental Setup

This section states the research questions that motivated the study. It also describes baselines, the evaluation strategy, and technical details of the approaches used, including their configuration setup in the experiments conducted.

Research Questions

Our experiments are designed to answer the following research questions:

RQ1: Does the feature-instance adaptation approach show similar patterns on both datasets used in the study, specifically, T6 and 2C datasets? How do the results of the feature-instance adaptation compare with the results of supervised classifiers learned from the source data without any adaptation?

RQ2: Can the self-training parameter adaptation help improve the results of the feature-instance adaptation approach? Similarly, can the feature-instance adaptation improve the results of the self-training approach used on the original source labeled data together with the target unlabeled data?

RQ3: Between Bernoulli Naïve Bayes and Random Forest, which classifier benefits most from the hybrid feature-instance-parameter adaptation approach?

RQ4: What specific source-target pairs benefit the most from the adaptation?

RQ5: Does the proposed approach benefit the datasets similarly? What characteristics makes a dataset more suitable for analysis with the proposed feature-instance-parameter adaptation approach?

Baselines

We compare the proposed hybrid feature-instance-parameter approach against the following baselines:

- Supervised Bernoulli Naïve Bayes and Random Forest classifiers learned from the binary representation of the source and evaluated on the test target data;
- Feature-instance adaptation with Bernoulli Naïve Bayes and Random Forest classifiers, where we first use the binary representation of the source and target to find a reduced dense representation, and subsequently learn classifiers from the selected source subset with the binary representation;

- Self-training parameter adaptation with Bernoulli Naïve Bayes and Random Forest as base classifiers, on the original source-target binary data, without any feature-instance adaptation.

Evaluation Strategy

The 5-fold cross-validation strategy was used for evaluation. This is achieved by creating five random folds from the target unlabeled data (using the stratified splitting mechanism that ensures that the overall data distribution is maintained in each fold). In a cross-validation experiment, four folds are used for the adaptation, as target unlabeled data *TU* (together with the whole source labeled data *SL*), while the fifth fold is used as target test *TT*. The folds are then rotated such that each fold is used as the test fold exactly once. As a different *TU* set is used for adaptation, there will be a different adapted source at each rotation, which in turn creates a different classifier for each test fold. For a particular source-target pair, the accuracy results are averaged over five folds. Furthermore, for a global visual analysis corresponding to a dataset, the accuracy results are averaged over all the pairs in the dataset.

Classification Setup

Two base classifiers are used in the experiments, Bernoulli Naïve Bayes and Random Forests, as explained in the previous section (Methods). The classification algorithms are always used with the binary representation of the data, both when the original source data is used and also when a selected subset of the source data is used. Bernoulli Naïve Bayes does not have any hyper-parameters that need to be set. For Random Forest, the number of trees used is set to 100. The values of the other parameters are the default values.

Matrix Factorization Setup

Each source-target pair of events is represented by an instance-feature data matrix consisting of binary BOW (Bag-of-Words) vectors, based on the target vocabulary. Using LSMNF, the matrix is reduced from approximately 1,000 features into matrices of 30, 50, 100, 200, and 500 features, while retaining the same instance count.

K-Nearest Neighbors Setup

A key element in the feature-instance adaptation approach is the k -NN step. Given a similarity metric, in our case, cosine similarity, this process will select the k nearest neighbors for each instance of the target unlabeled set, *TU*, from the entire source labeled set, *SL*. The following k values were used in the experiments: 1, 3, 5, 7, 9, and 11, to understand the effects of this parameters on the results. Using this method of instance selection, especially for higher values of k, the selected source will potentially include duplicates. Given the prior work in Mazloom et al. (2018) we will keep duplicate instances in the selected source.

Self-Training Setup

The self-training classifiers have several hyper-parameters that can be tuned, including the number of iterations i, the number of target instances n to be added to the training data at each iteration i, and the parameter γ that is used to shift the weight from the source to the target at each iteration. Based on prior work (Li, Caragea, and Caragea 2017), for the CrisisLexT6 dataset, we set n to 5, and γ to $i * 0.002$, where i is the current iteration number, and we ran the self-training approach until convergence, as it was observed that the accuracy increased when more target instances were added to the training set. For the 2CTweets dataset, where the event size is significantly smaller than the event size in CrisisLexT6, we set n to 2, γ to $i * 0.002$, and did some preliminary tuning of the number of iterations, as sometimes, the performance decreased when more target instances were added to the training set. When Random Forest is used as the base classifier, two new trees are built

using the selected target unlabeled instances and added to the classifier with weight γ. Based on preliminary experimentation, it was also observed that when the number of target trees increases by *1.5* times the source trees, initially set to 100, the integrity of the classifier's training collapses, hence, the classifier is limited by the initial number of source trees.

EXPERIMENTAL RESULTS AND DISCUSSION

The average results for different adaptation approaches, over the 12 pairs in the CrisisLexT6 dataset, are visually shown in Figure 1 for Bernoulli Naïve Bayes and Random Forest as base classifiers, respectively. Similarly, the average results for different adaptation approaches, over the 12 pairs in the 2CTweets dataset, are visually shown in Figure 2 for Bernoulli Naïve Bayes and Random Forest as base classifiers, respectively. The results are discussed in what follows, by analyzing each adaptation approach with respect to its baseline.

Feature-Instance Adaptation

The results of the feature-instance adaptation are discussed in this section and used to answer our research question *RQ1*. As part of the feature-instance adaptation, the original binary feature set for

Figure 1. Average classification accuracy results for the 12 source-target pairs from the CrisisLexT6 dataset. The left side of the graph shows the results of the following approaches: supervised learning on the original source data, Original, and feature-instance adaptation (with different values for the number of features f and different values for the number of neighbors k). For example, the two sections of 30f-1k are interpreted as follows: 30f means LSNMF has reduced the features to 30, while 1k means that one source instance was selected for each target instance. Duplicates were retained in all feature-instance adaptation experiments, and the selected source instances were remapped back to their original representation after the adaptation. The right side of the graph shows combinations of the previous approaches with ST. Results for Bernoulli Naive Bayes are shown in the top graph, while results for Random Forest are shown in the bottom graph.

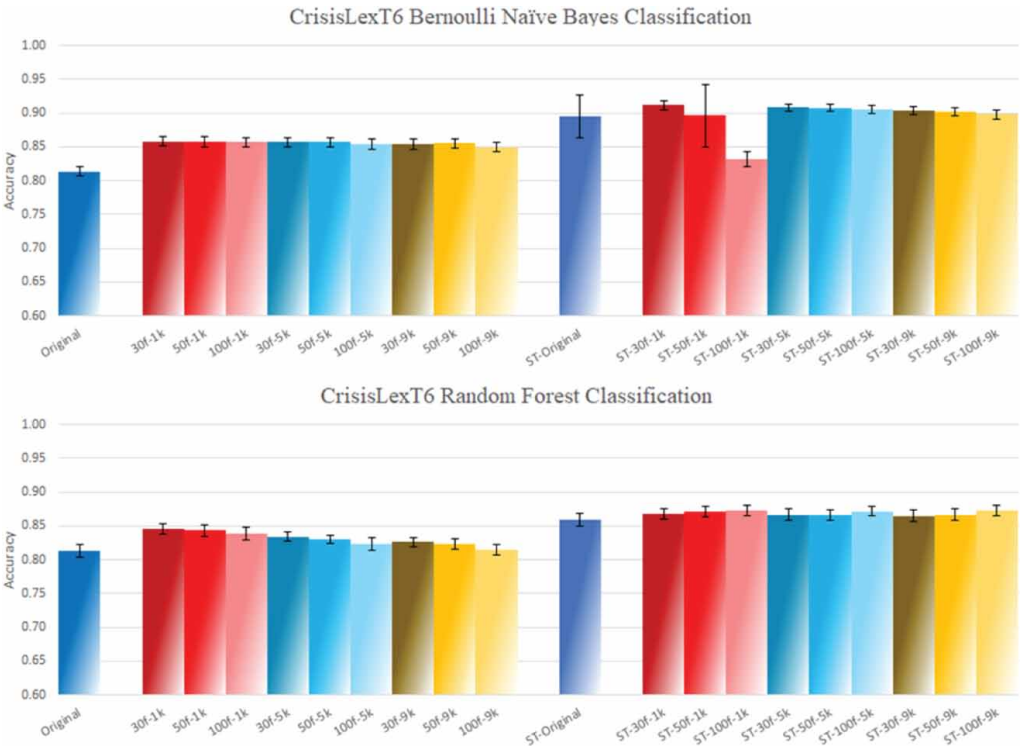
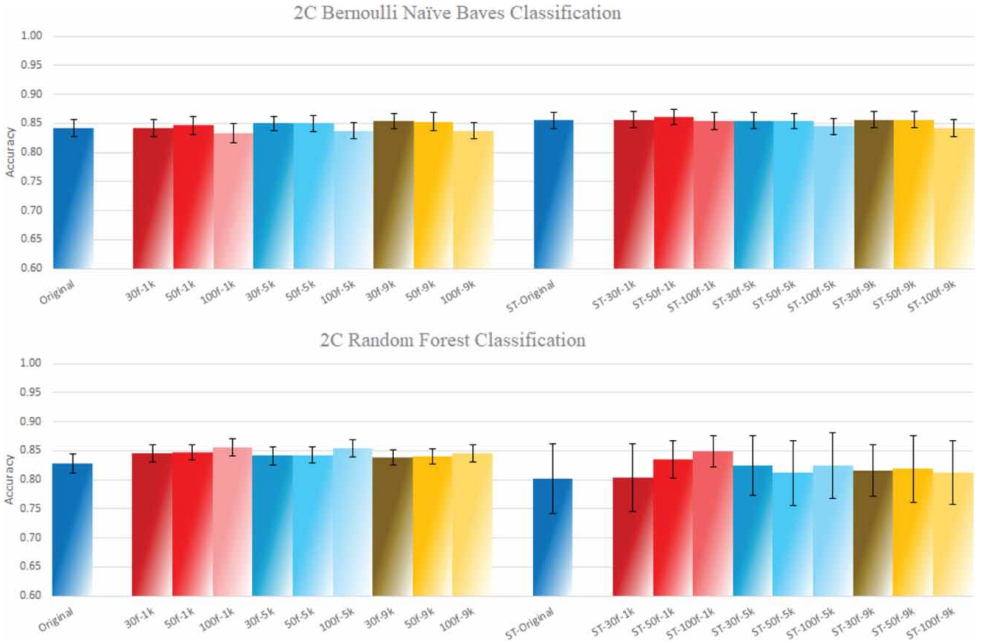


Figure 2. Average classification accuracy results for the 12 source-target pairs from the 2CTweets dataset. The left side of the graph shows the results of the following approaches: supervised learning on the original source data, Original, and feature-instance adaptation (with different values for the number of features f and different values for the number of neighbors k). For example, the two sections of 30f-1k are interpreted as follows: 30f means LSNMF has reduced the features to 30, while 1k means that one source instance was selected for each target instance. Duplicates were retained in all feature-instance adaptation experiments, and the selected source instances were remapped back to their original representation after the adaptation. The right side of the graph shows combinations of the previous approaches with ST. Results for Bernoulli Naive Bayes are shown in the top graph, while results for Random Forest are shown in the bottom graph.

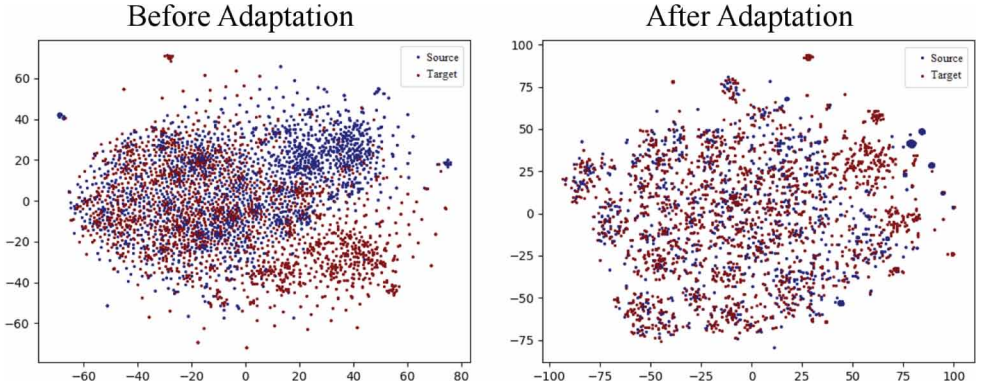


each dataset is first reduced to a denser representation using the feature adaptation technique (i.e., LSNMF). Afterwards, source instances in the reduced representation are adapted to the target using the instance adaptation technique (i.e., k-NN). As the last step, the source instances are remapped back to their original binary representation. The goal of the feature-instance adaptation is to increase the similarity of the source and target distributions.

We used the TSNE technique (Van Der Maaten 2014), before feature-instance adaptation and after feature-instance adaptation, to reduce the dimensionality of the source and target data to 2. As a result, each instance from the source and target data is represented by a two-dimensional vector. We plot the source and target instances together to visually analyze their distributions before and after adaptation. The results of the visualization are shown in Figure 3 for a pair of events in the CrisisLexT6 dataset (specifically, Boston Bombings versus the West Texas Explosion), before adaptation (left) and after adaptation (right). As can be seen, the two events have distinct distributions before adaptation, but their distributions are more similar after adaptation.

Figure 1 shows results for combinations of $f = 1,5,9$ neighbors and $k = 30,50,100$ features, respectively, as these values gave better overall results. Figure 1, in the case of CrisisLexT6, the hybrid feature-instance adaptation approach outperforms the supervised baseline by a significant margin, *Original*, for all *k, f* combinations considered, and for both Bernoulli Naïve Bayes and Random Forest classifiers. On the average, very small differences are seen between different *k, f* combinations, when Bernoulli Naïve Bayes is used, while smaller *k* and *f* values give better results when Random Forest is used.

Figure 3. Combined source (blue) and target (red) CrisisLexT6 data visualization before (left) and after feature-instance adaptation (right). It can be seen that before feature-instance adaptation (left), source (blue) instances and target (red) instances are partially separated into different clusters due to their different distributions. However, after feature-instance adaptation (right), the clusters have dispersed, and the distributions of the source and target instances have become similar. This shows that the adaptation technique has minimized variability between source and target distributions. The visualization is performed using the TSNE technique.



Somewhat similar patterns are observed for the 2CTweets dataset in Figure 2 when comparing feature-instance adaptation with the supervised baseline. However, the difference between the feature-instance adaptation and the supervised baseline is smaller than what was observed for CrisisLexT6. Furthermore, when using Bernoulli Naïve Bayes as a base classifier, a larger k values leads to better results, while in the case of Random Forest a larger f value gives better results. This may be explained by the smaller size of the 2CTweets dataset as compared to CrisisLexT6. A larger k value leads to larger subsets of selected instances, and subsequently better probability estimates for Bernoulli Naïve Bayes. In the case of Random Forest, a larger set of features leads to a more diverse set of trees in the Random Forest, and subsequently better results.

When comparing the results of the feature-instance adaptation with respect to the classifier used, overall the Bernoulli Naïve Bayes classifier gives better results than the Random Forest classifier for CrisisLexT6, while the Random Forest classifier gives slightly better results for 2CTweets. This result could also be explained by the size of the dataset used for training. When the size of the dataset is smaller, Random Forest, an ensemble classifier, can help boost the results. However, for larger datasets, the estimates obtained with Bernoulli Naïve Bayes are more accurate and lead to accurate classifiers.

Finally, it should be noted that the variance in accuracy observed across different source-target pairs is relatively small for the instance-adaptation approach, as well as for the original supervised classifiers. This observation suggests that the adaptation approach is able to consistently improve the results without causing much variation to the overall data distribution when compared to the *Original* data. Examples provided in Table 3 demonstrate that keyworks (Australia) and hash tags are best identified by the instance adaptation approach, denoted by k, whilst feature adaptation best captures news concept and negative sentiments. Consequently, instance-feature adaptation makes use of the best from both techniques.

Hybrid Feature-Instance-Parameter Adaptation

The self-training classification method was used as a parameter adaptation approach and combined with the feature-instance adaptation. As can be seen in Figure 1, for the CrisisT6 dataset, both Bernoulli Naïve Bayes and Random Forest classifiers benefit from parameter adaptation with self-training. Specifically, for Bernoulli Naïve Bayes, on the average, the results of the self-training classifier, ST-Original, are better than the results of the supervised baseline, Original, by 8%. Similarly, for Random Forest, on the average, the results of the self-training classifier, ST-Original, are better than the results

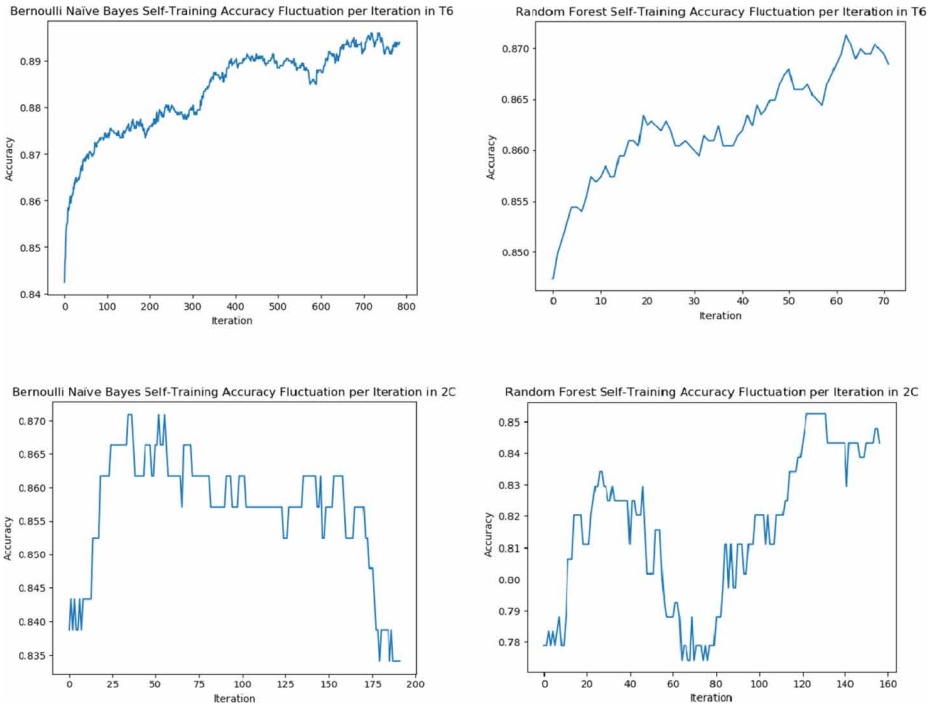
of the supervised baseline, Original, by 6%. Furthermore, the results of ST-Original are better than the results of the feature-adaptation approaches for both Bernoulli Naïve Bayes and Random Forest classifiers, showing that parameter adaptation is better than the feature-instance adaptation when each approach is used by itself. However, when combining feature-instance adaptation with self-training, accuracy is further increased. The increased accuracy of the combined feature-instance and parameter adaptation supports the idea that each adaptation approach is able to transfer complementary information from the source, and together they provide a better training base for the target when a large enough sample set is available.

As can be seen in Figure 2, the 2CTweets dataset presents a similar pattern when the Bernoulli Naïve Bayes is used, although the increase produced by self-training is smaller than the increase observed for CrisisLexT6. When the Random Forest classifier is used, the results produced by self-training, ST-Original, are worse on the average than the results produced by the supervised baseline, Original, and a high variance is observed between different source-target pairs. The results improve when self-training is used in combination with the feature-instance adaptation, but overall, they are still worse than their feature-instance adaptation counterparts. Intuitively, the poor performance of self-training, in this case, can be explained by the size of the data and the way the self-training is implemented for Random Forest. Specifically, given the small size of the data, we only add two target instances from each class to the training set (four instances all together). Those instances are used to create two new trees, that are added to the previous trees in the Random Forest. Given that the two new trees are based on four instances, they will be somewhat biased. On the other hand, we observed that adding more target instances to the training set results in incorrectly labeled instances being used, which also leads to poor performance. Different ways to implement the Random Forest classifier with self-training will be explored in future work. It should be noted that this is not an issue for the Bernoulli Naïve Bayes, where the counts are simply updated based on the new training data.

Thus, the answer to our research question RQ2 is that the feature-instance approach generally improves the results of the self-training approach (by comparing ST-Original versus feature-instance adaptation). Similarly, the feature-instance adaptation can improve the results of self-training (by comparing ST-Original with feature-instance-parameter adaptation), especially for larger datasets. However, the two adaptation approaches, self-training and feature-instance adaptation, are very competitive by themselves and give better results than the supervised baselines in most cases. Regarding research question, RQ3, as discussed above both Bernoulli Naïve Bayes and Random Forest classifiers benefit from feature-instance adaptation. The Bernoulli Naïve Bayes classifier also benefits from self-training adaptation consistently. However, with the current implementation, the Random Forest classifier benefits from self-training only in the case of larger datasets.

Finally, to answer RQ4, it should be noted that differences between CrisisLexT6 and 2CTweets datasets lead to different behavior of the self-training algorithm when used on one dataset versus the other. Specifically, as mentioned before, the events in CrisisLexT6 have significantly larger sizes when compared to the events in the 2Ctweets dataset. This ensures that accurately labeled training instances are identified at each iteration, and thus a larger number of iterations are possible. As a consequence, the performance generally increases with the number of iterations. Furthermore, the pairs of events in CrisisLexT6 have better contextual overlap, as compared to the pairs of events in 2CTweets, a fact that also contributes to accurate labels for some target instances at each iteration. As opposed to that, the events in the 2CTweets dataset are smaller in size, and the pairs of events show less overlap. Therefore, there is a higher chance that incorrectly labeled target instances are added to the training set decreasing the performance with more data. This behavior can be seen clearly in Figure 4, where the 2C pair classification shows an unusual instability in accuracy, as more target instances are introduced at each iteration. The CrisisLexT6 pair, on the other hand, has a more stable accuracy which increases with the number of iterations. Given this behavior, the number of iterations for 2CTweets was roughly tuned for a source-target pair (as opposed to tuning for each split) in the preliminary examination. More fine-tuning may lead to better results.

Figure 4. Variation of accuracy with the self-training iterations for a representative pair of events from CrisisLexT6 (Top), and a representative pair of events from 2CTweets (Bottom). Self-Training Bernoulli Naïve Bayes is used to train the classifiers in the graphs on the left, while Random Forest is used to train the classifiers in the graphs on the right. For the pair from CrisisLexT6, five target instances per class are added at each iteration, while for the 2CTweets pair, two target instances per class are added. This figure depicts the self-training instability for the 2CTweets pair as compared to the CrisisLexT6 pair.



When comparing the feature-instance adaptation method to the baselines using Bernoulli Naïve Bayes in both datasets, the adapted results tend to significantly improve the baseline results in 83% and 89% of the cases for CrisisLexT6 and 2CTweets, respectively. Similarly, when parameter adaptation is used, the results are significant in 71% and 89% of the cases, respectively, demonstrating that parameter adaptation was more effective in 2CTweets due to the dataset's smaller size. The feature-instance adaptation was also able to significantly outperform individual instance adaptation in 61% and 42% of the cases, as well as individual feature adaptation in 60% and 40% of the cases, for CrisisLexT6 and 2CTweets, respectively. However, when using Random Forest, parameter adaptation is able to increase the significant results from 43% to 60% in CrisisLexT6, when compared to no feature-instance adaptation, while in 2CTweets the significant results decrease from 45% to 36%. As discussed above, this is most likely due to the lack of instances that are used to create new trees at each iteration of the parameter adaptation. Both adaptation sets are able to outperform individual feature and instance adaptation in both datasets by about 60% with the exception of feature-instance-parameter adaptation compared to individual instance adaptation in 2CTweets.

RELATED WORK

Machine learning algorithms have been used to help responders sift through the huge amounts of crisis data and prioritize information that may be useful for response and relief. Some studies have used images to identify such data (Alam, Imran, and Ofli, 2017; Li et al., 2019) whereas others have relied on the more abundant text data (Verma et al., 2011; Caragea et al., 2011; Vieweg, 2012; Terpstra et al.,

2012; Purohit et al., 2013; Imran et al., 2013; Caragea et al., 2014; Ashktorab et al., 2014; Imran and Castillo 2015; Sen, Rudra, and Ghosh, 2015; Huang and Xiao, 2015; Imran, Chawla, Castillo, 2016; Derczynski et al., 2018; Zahera, Jalota, and Usbeck, 2018), or in some cases a combination of both image and text data (Jomaa, Rizk, and Awad, 2016; Zhang et al., 2019; Mouzannar, Rizk, and Awad 2018). For example, Imran et al. (2013) used conditional random fields and Karami et al. (2019) used sentiment analysis and topic modeling to find tweets within specific situational awareness categories. Sen, Rudra, and Ghosh (2015) used Support Vector Machine (SVM) classifiers to differentiate between situational and non-situational tweets. Huang and Xiao (2015) introduced a detailed list of situational awareness categories, divided based on three stages of a disaster (preparedness, emergency response, and recovery), and used k-Nearest Neighbors, Logistic Regression and Naïve Bayes classifiers to automatically classify tweets with respect to their defined categories. And others such as Zade et al. (2018) use different mediums such as surveys and interviews to aid in the effort.

While research on supervised machine learning in the area of emergency response has shown that it is possible to automatically classify disaster-related data, it has also emphasized one of the most important challenges that precludes the use of supervised machine learning in real time in an emerging crisis situation: the lack of labeled data to train reliable supervised models as the crisis unfolds. To address this challenge, several works proposed to use labeled data from prior "source" crises to learn supervised classifiers for a "target" crisis (Verma et al., 2011; Imran, Mitra, and Srivastava, 2016; Caragea, Silvescu, and Tapia, 2016; Nguyen et al., 2017). One drawback of this approach is that supervised classifiers learned in one crisis event, does not generalize well to other events (Qadir et al., 2016; Imran et al., 2015), as each event has unique characteristics (Palen and Anderson, 2016). These problems are widely known as domain adaptation or transfer learning (Kouw, 2018). Domain adaptation approaches (Pan and Yang 2010; Jiang 2008) that make use of unlabeled data from the target disaster in addition to label data from a source disaster are desirable. Some recent works (Li et al., 2015; Li, Caragea, and Caragea, 2017; Li et al., 2017) have shown that the use of domain adaptation approaches can significantly improve the results of the supervised classifiers learned from source only. According to Pan and Yang (2010), domain adaptation is achieved by performing parameter adaptation, feature adaptation or instance adaptation. A comprehensive description of works in each category can be found in (Pan and Yang, 2010).

In the space of disasters, the domain adaptation approaches proposed by Li et al. (2015, 2017) can be seen as parameter-based adaptation approaches. Mazloom et al. (2018) proposed a hybrid feature-instance adaptation approach and tested it on the CrisisLexT6 dataset using the Bernoulli Naïve Bayes. Mazloom et al. (2018) demonstrated the complementary strengths of feature and instance adaptation approaches, which can significantly improve classification when combined and used with supervised approaches such as Bernoulli Naïve Bayes.

To the best of our knowledge, there are no feature-instance-parameter adaptation approaches that have been used for classifying disaster related data. As a consequence, we first extend the work by Mazloom et al. (2018) by studying another dataset (2CTweets) and another base classifier (Random Forest), and subsequently we introduce a hybrid feature-instance-parameter approach and study it using both CrisisLexT6 and 2CTweets as datasets, with both Bernoulli Naïve Bayes and Random Forests as base classifiers.

CONCLUSION AND FUTURE WORK

Social media data taken from sources such as Twitter contain invaluable data which can be used in times of crisis and emergency situations to improve response and awareness. Despite many supervised learning approaches being proposed, not many agencies and groups use these approaches to identify useful information, due to lack of labeled data for training the supervised models. In this study, we proposed a simple but powerful feature-instance-parameter adaptation approach to first reduce the variation between source and target disasters, and subsequently use self-training on the modified

source together with target unlabeled data to address the scarcity of the labeled data in the target domain. Experimental results on pairs of events from two disaster datasets, CrisisLexT6 and 2CTweets, using Bernoulli Naïve Bayes and Random Forest classifiers, show that the proposed approach can help improve the results of self-training used by itself, and also the results of the feature-instance adaptation approach used by itself. However, the two adaptation approaches, self-training and feature-instance adaptation, are also powerful by themselves and give very competitive results in some cases, and better results than the supervised baselines in most cases. Between Bernoulli Naïve Bayes and Random Forest, the Bernoulli Naïve Bayes classifiers work better with the self-training approach, although both benefit from the feature-instance adaptation approach. In terms of datasets, the results show that a larger number of sources labeled and target unlabeled instances are beneficial, especially when self-training is used. Furthermore, the incremental updates that were used to implement self-training work well in the case of Bernoulli Naïve Bayes but not so well in the case of Random Forest.

In future work, more experiments can be done using different classifiers, including deep learning classifiers, on the selected source data. Furthermore, different matrix factorization and clustering approaches (potentially, with different distance metrics) can be explored. Finally, different domain adaptation classifiers, including different implementations of self-training, can be used.

ACKNOWLEDGMENT

The computing for this project was performed on the Beocat Research Cluster at Kansas State University. We thank the National Science Foundation for support [grant number CNS-1429316], which partially funded the cluster. We also thank the National Science Foundation for support [grant numbers IIS-1802284, IIS-1741345, IIS-1526542 and CMMI-1541155]. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either express or implied, of the National Science Foundation. We also wish to thank our anonymous reviewers for their constructive comments.

REFERENCES

Alam, F., Imran, M., & Ofli, F. (2017). Image4act: Online social media image processing for disaster response. *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017* (pp. 601-604). ACM. doi:10.1145/3110025.3110164

Ashktorab, Z., Brown, C., Nandi, M., & Culotta, A. (2014). Tweedr: Mining twitter to inform disaster response. *Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management ISCRAM '14*, University Park, PA. Academic Press.

Beigi, G., Hu, X., Maciejewski, R., & Liu, H. (2016). An overview of sentiment analysis in social media and its applications in disaster relief. In Sentiment Analysis and Ontology Engineering (pp. 313–340). Springer. doi:10.1007/978-3-319-30319-2_13

Bullock, J. A., Haddow, G. D., & Coppola, D. P. (2017). *Homeland security: the essentials*. Butterworth-Heinemann.

Caragea, C., McNeese, N., Jaiswal, A., Traylor, G., Kim, H.-W., Mitra, P., . . . Yen, J. (2011). Classifying Text Messages for the Haiti Earthquake. *Proceedings of the 8th International Conference on Information Systems for Crisis Response and Management ISCRAM '11*, Lisbon, Portugal. Academic Press.

Caragea, C., Silvescu, A., & Tapia, A. H. (2016). Identifying Informative Messages in Disasters using Convolutional Neural Networks. *Proceedings of the 13th Proceedings of the International Conference on Information Systems for Crisis Response and Management*, Rio de Janeiro, Brasil, May 22-25. Academic Press.

Caragea, C., Squicciarini, A. C., Stehle, S., Neppalli, K., & Tapia, A. H. (2014). Mapping moods: Geo-mapped sentiment analysis during hurricane sandy. *Proceedings of the 11th Proceedings of the International Conference on Information Systems for Crisis Response and Management,* University Park, PA, May 18-21. Academic Press.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B. Methodological*, *39*(1), 1–38. doi:10.1111/j.2517-6161.1977.tb01600.x

Derczynski, L., Meesters, K., Bontcheva, K., & Maynard, D. (2018). Helping crisis responders find the informative needle in the tweet haystack.

Genuer, R. (2012). Variance reduction in purely random forests. *Journal of Nonparametric Statistics*, *24*(3), 543–562. doi:10.1080/10485252.2012.677843

Guo, W., & Diab, M. (2012). A simple unsupervised latent semantics based approach for sentence similarity. *Proceedings of the First Joint Conference on Lexical and Computational Semantics* (pp. 586–590). Association for Computational Linguistics.

Huang, Q., & Xiao, Y. (2015). Geographic situational awareness: Mining tweets for disaster preparedness, emergency response, impact, and recovery. *ISPRS International Journal of Geo-Information*, *4*(3), 1549–1568. doi:10.3390/ijgi4031549

Imran, M., & Castillo, C. (2015). Towards a data-driven approach to identify crisis-related topics in social media streams. *Proceedings of the 24th International Conference on World Wide Web* (pp. 1205–1210). ACM. doi:10.1145/2740908.2741729

Imran, M., Castillo, C., Diaz, F., & Vieweg, S. (2015). Processing social media messages in mass emergency: A survey. *ACM Computing Surveys*, *47*(4), 67. doi:10.1145/2771588

Imran, M., Chawla, S., & Castillo, C. (2016). A Robust Framework for Classifying Evolving Document Streams in an Expert-Machine-Crowd Setting. *Proceedings of the 18th International Conference on Data Mining (ICDM)*, Barcelona, Spain. Academic Press. doi:10.1109/ICDM.2016.0120

Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., & Meier, P. (2013). Practical extraction of disaster-relevant information from social media. *Proceedings of the 22nd International World Wide Web Conference, WWW '13*, Rio de Janeiro, Brazil, May 13-17 *(pp. 1021–1024)*. Academic Press. doi:10.1145/2487788.2488109

Imran, M., Mitra, P., & Srivastava, J. (2016). Cross-Language Domain Adaptation for Classifying Crisis-Related Short Messages. *Proceedings of the 13th Proceedings of the International Conference on Information Systems for Crisis Response and Management*, Rio de Janeiro, Brasil, May 22-25. Academic Press.

Jiang, J. (2008). A literature survey on domain adaptation of statistical classifiers. Retrieved from http://sifaka.cs.uiuc.edu/jiang4/domainadaptation/survey3

Jomaa, H. S., Rizk, Y., & Awad, M. (2016). Semantic and Visual Cues for Humanitarian Computing of Natural Disaster Damage Images. *Proceedings of the 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)* (pp. 404–411). IEEE. doi:10.1109/SITIS.2016.70

Karami, A., Shah, V., Vaezi, R., & Bansal, A. (2019). Twitter speaks: A case ofnational disaster situational awareness. *Journal of Information Science*.

Kouw, W. M. (2018). An introduction to domain adaptation and transfer learning.

Li, H., Caragea, D., & Caragea, C. (2017). Towards Practical Usage of a Domain Adaptation Algorithm in the Early Hours of a Disaster. *Proceedings of the 14th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2017)*, France. Academic Press.

Li, H., Caragea, D., Caragea, C., & Herndon, N. (2018). Disaster response aided by tweet classification with a domain adaptation approach. *Journal of Contingencies and Crisis Management*, *26*(1), 16–27.

Li, H., Guevara, N., Herndon, N., Caragea, D., Neppalli, K., Caragea, C., & Tapia, A. H. et al. (2015, May). Twitter Mining for Disaster Response: A Domain Adaptation Approach. *Proceedings of the 12th International Conference on Information Systems for Crisis Response and Management*, Kristiansand, Norway. Academic Press.

Li, X., Caragea, D., Caragea, C., Imran, M., & Ofli, F. (2019). Identifying Disaster Damage Images Using a Domain Adaptation Approach. *Proceedings of the 16th International Conference on Information Systems for Crisis Response and Management* (ISCRAM), Valencia, Spain. Academic Press.

Lin, C.-J. (2007). Projected gradient methods for nonnegative matrix factorization. *Neural Computation*, *19*(10), 2756–2779. doi:10.1162/neco.2007.19.10.2756 PMID:17716011

Maaten, Van Der, L. (2014). Accelerating t-SNE using tree-based algorithms. *Journal of Machine Learning Research*, *15*(1), 3221–3245.

Mazloom, R., Li, H., Caragea, D., Imran, M., & Caragea, C. (2018). Classification of Twitter Disaster Data Using a Hybrid Feature-Instance Adaptation Approach. *Proceedings of the 15th Annual Conference for Information Systems for Crisis Response and Management*, ISCRAM Rochester, NY, May 20-24. Academic Press.

Meier, P. (2013). Crisis Maps: Harnessing the Power of Big Data to Deliver Humanitarian Assistance. Forbes Magazine. Retrieved from https://www.forbes.com/sites/skollworldforum/2013/05/02/crisis-maps-harnessing-the-power-of-big-data-to-deliver-humanitarian-assistance/#73b91c24115c

Meier, P. (2015). *Digital Humanitarians: How Big Data Is Changing the Face of Humanitarian Response*. Boca Raton, FL: CRC Press, Inc. doi:10.1201/b18023

Mouzannar, H., Rizk, Y., & Awad, M. (2018). Damage Identification in Social Media Posts using Multimodal Deep Learning. *Proceedings of the 12th International Conference on Information Systems for Crisis Response and Management*. Academic Press.

Nguyen, D. T., Al-Mannai, K., Joty, S. R., Sajjad, H., Imran, M., & Mitra, P. (2017). Robust Classification of Crisis-Related Data on Social Networks using Convolutional Neural Networks. *Proceedings of the 11th International AAAI Conference on Web and Social Media (ICWSM)*, Montreal, Canada. Academic Press.

Olteanu, A., Castillo, C., Diaz, F., & Vieweg, S. (2014a). CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises. *Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014*, Ann Arbor, MI, June 1-4, 2014. Academic Press.

Olteanu, A., Castillo, C., Diaz, F., & Vieweg, S. (2014b, May). Crisislex: A lexicon for collecting and filtering microblogged communications in crises. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*. AAAI Press.

Palen, L., & Anderson, K. M. (2016). Crisis informatics-New data for extraordinary times. *Science*, *353*(6296), 224–225. doi:10.1126/science.aag2579 PMID:27418492

Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. doi:10.1109/TKDE.2009.191

Purohit, H., Castillo, C., Diaz, F., Sheth, A., & Meier, P. (2013). Emergency-relief coordination on social media: Automatically matching resource requests and offers. *First Monday*, *19*(1). doi:10.5210/fm.v19i1.4848

Qadir, J., & Ali, A., ur Rasool, R., Zwitter, A., Sathiaseelan, A., & Crowcroft, J. (2016). Crisis analytics: Big data-driven crisis response. *Journal of International Humanitarian Action*, *1*(1), 12.

Schulz, A., Guckelsberger, C., & Janssen, F. (2017). Semantic Abstraction for generalization of tweet classification: An evaluation of incident-related tweets. *Semantic Web*, *8*(3), 353–372. doi:10.3233/SW-150188

Sen, A., Rudra, K., & Ghosh, S. (2015). Extracting situational awareness from microblogs during disaster events. *Proceedings of the 2015 7th International Conference on Communication Systems and Networks (COMSNETS)*. IEEE. doi:10.1109/COMSNETS.2015.7098720

Terpstra, T., De Vries, A., Stronkman, R., & Paradies, G. L. (2012). *Towards a realtime Twitter analysis during crises for operational crisis management* (pp. 1–9). Burnaby: Simon Fraser University.

Verma, S., Vieweg, S., Corvey, W. J., Palen, L., Martin, J. H., Palmer, M., . . . Anderson, K. M. (2011). Natural Language Processing to the Rescue? Extracting "Situational Awareness" Tweets During Mass Emergency. *Proceedings of the Fifth International Conference on Weblogs and Social Media*, Barcelona, Spain, July 17-21. Academic Press.

Vieweg, S. E. (2012). Situational awareness in mass emergency: A behavioral and linguistic analysis of microblogged communications [PhD diss.]. University of Colorado at Boulder.

Watson, H., Finn, R. L., & Wadhwa, K. (2017). Organizational and Societal Impacts of Big Data in Crisis Management. *Journal of Contingencies and Crisis Management*, *25*(1), 15–22. doi:10.1111/1468-5973.12141

Yarowsky, D. (1995). Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. *Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics*, Stroudsburg, PA (pp. 189–196). Association for Computational Linguistics. doi:10.3115/981658.981684

Zade, H., Shah, K., Rangarajan, V., Kshirsagar, P., Imran, M., & Starbird, K. (2018). From Situational Awareness to Actionability: Towards Improving the Utility of Social Media Data for Crisis Response. *Proceedings of the ACM on Human-Computer Interaction*. Academic Press. doi:10.1145/3274464

Zahera, H. M., Jalota, R., & Usbeck, R. (2018). DICE@ TREC-IS: Combining Knowledge Graphs and Deep Learning to Identify Crisis-Relevant Tweets. Academic Press.

Zhang, C., Fan, C., Yao, W., Hu, X., & Mostafavi, A. (2019). Social media for intelligent public information and warning in disasters: An interdisciplinary review. *International Journal of Information Management*, *49*, 190–207. doi:10.1016/j.ijinfomgt.2019.04.004

Reza Mazloom, M.S. is a Research Assistant at the Institute of Computation Comparative Medicine, Kansas State University with an M.S. in Computer Science. His research interests are machine learning, computational biology, big data, bioinformatics, and data mining.

Hongmin Li, Ph.D candidate, Department of Computer Science at Kansas State University. Her research interests are mainly related to machine learning more specifically domain adaptation approaches on big crisis social media data.

Doina Caragea, Ph.D., is a Professor at Kansas State University. Her research and teaching interests are in the areas of machine learning, data mining, data science, information retrieval and text mining, with applications to crisis informatics, security informatics, recommender systems, and bioinformatics. Her projects build upon close collaborations with social scientists, security experts and life scientists, and aim to provide practical computational approaches to address real-world challenges. Dr. D. Caragea received her PhD in Computer Science from Iowa State University in August 2004, and was honored with the Iowa State University Research Excellence Award for her work. She has published more than 100 refereed conference and journal articles. Her research has been supported by several NSF grants.

Cornelia Caragea is an Associate Professor of Department of Computer Science, University of Illinois at Chicago, and also an Adjunct Associate Professor of Department of Computer Science, Kansas State University. Her research interests are in artificial intelligence, machine learning, information retrieval, and natural language processing.

Muhammad Imran is a Research Scientist at the Qatar Computing Research Institute (QCRI) where he leads the Crisis Computing team under the Social Computing group from both science and engineering directions. His interdisciplinary research focuses on natural language processing, text mining, human-computer interaction, applied machine learning, and stream processing areas. Dr. Imran has published over 70 research papers in top-tier international conferences and journals including ACL, SIGIR, IJHCI, ICWSM, WWW, ASONAM, and ACM HT. Three of his papers received the "Best Paper Award" and two "Best Paper Runner-up Award." He has been serving as a Co-Chair of the Social Media Studies track of the ISCRAM international conference since 2014 and has served as Program Committee (PC) for many major conferences and workshops.