Sympathy Detection in Disaster Twitter Data

Yingjie Li
University of Illinois at Chicago
yli300@uic.edu

Seoyeon Park
University of Illinois at Chicago
spark313@uic.edu

Cornelia Caragea
University of Illinois at Chicago
cornelia@uic.edu

Doina Caragea
Kansas State University
dcaragea@ksu.edu

Andrea Tapia
Pennsylvania State University
atapia@ist.psu.edu

ABSTRACT

Nowadays, micro-blogging sites such as Twitter have become powerful tools for communicating with others in various situations. Especially in disaster events, these sites can be the best platforms for seeking or providing social support, of which informational support and emotional support are the most important types. Sympathy, a sub-type of emotional support, is an expression of one's compassion or sorrow for a difficult situation that another person is facing. Providing sympathy to people affected by a disaster can help change people's emotional states from negative to positive emotions, and hence, help them feel better. Moreover, detecting sympathy contents in Twitter can potentially be used for finding candidate donors since the emotion "sympathy" is closely related to people who may be willing to donate. Thus, in this paper, as a starting point, we focus on detecting sympathy-related tweets. We address this task using Convolutional Neural Networks (CNNs) with refined word embeddings. Specifically, we propose a refined word embedding technique in terms of various pre-trained word vector models and show great performance of CNNs that use these refined embeddings in the sympathy tweet classification task. We also report experimental results showing that the CNNs with the refined word embeddings outperform not only traditional machine learning techniques, such as Naïve Bayes, Support Vector Machines and AdaBoost with conventional feature sets as bags of words, but also Long Short-Term Memory Networks.

Keywords

Refined Word Embedding, Deep Learning, Machine Learning, Sympathy Tweets Detection

INTRODUCTION

Sympathy is an expression of one's compassion or sorrow for a difficult situation that another person is facing. Expressing sympathy for people who are facing disaster situations is one way of caring for people emotionally. S. Kim et al. (2012) discovered that the expression of certain emotions, including sympathy "influences" other people in Twitter to become more positive. Moreover, Froyum (2018) discussed that if an emotional connection between helper and helped exists, e.g., sympathy, helpers are more eager to give assistance in various kinds of formats. According to Baberini et al. (2015), there is a higher chance that people who express sympathy could potentially provide monetary helping hands since they are emotionally inclined to people who are encountering disasters status. Donations are very important to people in crisis circumstances for restoring their lives.

Since micro-blogging social media sites such as Twitter are widely used nowadays, people get real-time information by using these sites. Not only limited to personal use for acquiring knowledge, exploiting social media as information sources in groups has shed light on earning valuable resources. For example, for building "post-disaster" programs,

Table 1. Examples of Sympathy Tweets and Non Sympathy Tweets

Sympathy Tweets	Non Sympathy Tweets
RT@HotelElGanzo: We send our best wishes and thoughts out to everyone who is being affected by #HurricaneODILE - Please stay safe #Cuidense	RT @RedCross: To check on loved ones in #Baja impacted by #Odile, contact State Departments Overseas Citizen Services at 1-888-407-4747
#India deepest sympathy and respect to all those killed or injured in the flood	Floods triggered by heavy rainfall in Himalayas wreak havoc in North India, inundates 1500 villages in Uttar Pradesh http://t.co/W97qixmYQ5
Hoping for best: Powerful earthquake shakes Northern California http://t.co/gvBVrSbLgN http://t.co/aCgUeEt8ZE	#news U.S. quake: The recovery from the strongest earthquake in 25 years to strike Northern California will be... http://t.co/OxVb2dGtWj

non-profit organizations can use Twitter as information roots. More specifically, in emergency situations like earthquakes, floods and typhoons strike, Red Cross can use Twitter as a source of big-data analysis for figuring out long-term donors. According to Han et al. (2013), identifying long-term donors is crucial because they need to plan longstanding programs such as fund-raising campaigns. However, Han et al. (2013) reported the fact that when a disaster happens, one-time donations usually spike sharply. Twitter, which is a real-time posting platform, can be a good resource for extracting valuable information since we can easily notice this sharply increasing one-time donation phenomenon in real-time.

Thus, our goal in this paper is to identify Twitter postings (or tweets) that have sympathy related contents. This can be used not only for designing improved features in Twitter for providing emotional support, but can also be used for identifying users that can be candidate donors. Consequently, we learn supervised models to automatically detect sympathy in Twitter data. Our labeled Twitter data comes from a series of diverse disasters that happened during 2013 to 2015 world-wide. Examples of tweets that contain and do not contain sympathy are shown in Table 1.

Many natural language processing approaches for short text or sentence classification use Convolutional Neural Networks (CNNs) (Y. Kim 2014) and Long Short-Term Memory Neural Networks (LSTMs) to reduce the burden of feature engineering. In text processing, deep learning approaches are predominant in terms of feature detection since humans do not need to intervene in a detailed way. Deep learning approaches have achieved substantial success over traditional machine learning techniques in disaster contexts as well (V. K. Neppalli et al. 2018; Nguyen et al. 2017; C. Caragea, Silvescu, et al. 2016; Ben Lazreg et al. 2016). For example, Nguyen et al. (2017) used CNNs to classify tweets into categories related to situational awareness and showed improvements over traditional algorithms. Ben Lazreg et al. (2016) used a different deep learning approach, specifically a Long Short-Term Memory (LSTM) network, to learn a model from crisis tweets and used this model to generate snippets of information summarizing the tweets. In addition to these, V. K. Neppalli et al. (2018) reported that CNNs outperform Naïve Bayes that exploit Twitter specific features and conventional linguistic features in Twitter informative posting classification.

In this paper, we focus on sympathy detection of Twitter disaster data using Convolutional Neural Networks. Conventional word embeddings such as pre-trained word2vec, Glove, FastText, and a domain-specific word2vec trained on CrisisNLP data (Imran, Mitra, et al. 2016) are used to test the performance of our proposed model. We observe that even if two different words look similar or are semantically related, conventional word embeddings based model may not consider them as tightly correlated, and hence, return a low similarity value. For example, when we use the CrisisNLP word embeddings for measuring the similarity of words "pray" and "prayer", the similarity value that implies how much they are alike is significantly low. In order to mitigate this problem, we design a refined word embedding technique based on word stemming, where semantically similar words become closer in the vector space. Our technique is inspired from Yu et al. (2017) who used a sentiment lexicon for generating polished word embeddings. By using refined word embeddings, we achieve an F1 score of approximately 76% as compared with approximately 74% achieved by conventional word embeddings. In experiments, we also contrast CNNs with two other deep learning models, LSTM and Bi-LSTM. In summary, our contribution can be organized as follows:

- We explore CNNs with various word embeddings (pre-trained and fine-tuned on domain specific data) for sympathy detection of Twitter disaster data. Moreover, we design a word stemming based refinement model to improve word representations in vector space.

- We show that the CNNs with refined word embeddings outperform the CNNs with conventional word embeddings. In addition, we contrast the CNNs with LSTMs and Bi-LSTMs and show that the CNNs (with and without refined word embeddings) have better performance in most cases.
- Our CNN models outperform traditional approaches such as decision tree and SVM on sympathy detection task. The F1 score of proposed model is 3.63% higher than SVM, which has the best performance in traditional classifiers.

RELATED WORK

Substantial progress has been made in extracting, processing and classifying messages posted in social media during disaster events, using natural language processing and machine learning techniques (C. Caragea, Squicciarini, et al. 2014; Verma et al. 2011, Sarda and Lal Chouhan 2017, Rudra, Sharma, et al. 2018). For example, Ashktorab et al. (2014) used a combination of classification, clustering, and extraction methods to identify actionable information for disaster responders. Imran, Elbassuoni, et al. (2013) classified tweets into situational awareness categories, such as caution and advice, casualties and damage, donations, people missing, using unigrams, bigrams, POS tags, presence of URL/hashtags, tweet length, etc. Yin et al. (2012) used similar features to identify tweets related to a disaster, identify disaster type and assess the impact of a disaster. Rudra, Ghosh, et al. (2015) used lexical features to extract situational information in disasters. More recently, researchers started to explore deep learning techniques, and in particular Convolutional Neural Networks and word embeddings, for classifying disaster social media data into diverse categories such as situational awareness, informational tweets, and actionable tweets (Alam et al. 2018; Derczynski et al. 2018; V. K. Neppalli et al. 2018; Nguyen et al. 2017; C. Caragea, Silvescu, et al. 2016).

Sentiment analysis and emotion detection have also been explored in disaster contexts. For example, Nagy and Stamberger (2012) focused on sentiment detection in Twitter during the San Bruno, California gas explosion and fires from 09/2010. They used SentiWordNet together with dictionaries of emoticons and out of vocabulary words, and a sentiment-based dictionary to identify the basic sentiment of a tweet. Schulz et al. (2013) focused on the classification of human emotions into six classes: anger, disgust, fear, happiness, sadness, and surprise, from tweets related to the Hurricane Sandy. As features, they used bag of words, part of speech tags, character n-grams (for n=3,4), emoticons, and sentiment-based words compiled from the AFINN word list (Nielsen 2011) and SentiWordNet (Baccianella et al. 2010). Mandel et al. (2013) performed a demographic sentiment analysis using Twitter data during Hurricane Irene. K. Neppalli et al. (2017) explored the concept of *emotional divergence* and analyzed how likely a tweet is to be retweeted with respect to its emotional divergence value, using Twitter data from the Sandy Hurricane. *Emotional divergence*, first introduced by (Pfitzner et al. 2012), measures the diversity of the emotions expressed in a text.

Social support is one important aspect for which people join social networks, and hence, researchers investigated the types of social support (e.g., emotional support and informational support) and their benefits in these networks. For example, Bautista and Lin (2015) studied social support in tweets posted during a terrorist manhunt in Philippine, in which 44 elite policemen died, and found that informational support was more frequent compared with emotional support. Among the sub-types of informational support, the authors found that a large fraction of the posted tweets contained news-related pictures and articles. Similarly, during the Great East Japan Earthquake of March 2011, the city lost electricity and water and communication became inaccessible. Smartphone became the primary device to obtain information and communicate with each other. The author found that Twitter has the potential to disseminate important information, which is greatly beneficial during the disaster (Kaigo 2012). When the massive floods happened in Queensland, Australia, during December 2010 to January 2011, the social media platforms provided valuable information to the public. Not only for this, social media was also used as communication channels to people who were facing emergency disaster situations (Bunce et al. 2012). Emotional support in Twitter shows benefits on people's moods and behavior. Generally, when people feel emotionally supported, they tend to feel better. For example, in a study on social aspects of emotions in the general Twitter, S. Kim et al. (2012) showed that the expression of certain emotions, including *sympathy*, "influences" others to become more positive. More precisely, the authors discovered that changes in emotional states from negative to positive are influenced by topics such as greeting, sympathy and recommendation, whereas changes from positive to negative are influenced by topics such as worry, teasing, and complaint (S. Kim et al. 2012).

Against this background, we explore and contrast deep learning models, Convolutional Neural Networks (CNNs) and Long Short-Term Memory Networks (LSTMs), and refined word embeddings for *sympathy detection* in disaster-related Twitter data. Imran, Mitra, et al. (2016) presented results on a multi-class classification task for classifying disaster-related tweets into situational awareness categories of which sympathy and emotional support represents one such category. However, unlike Imran, Mitra, et al. (2016), we present a thorough analysis of sympathy detection and propose a word embedding technique that produces more suitable embeddings for our task.

METHODS

Nowadays, CNNs perform well on many text classification tasks. For example, Y. Kim (2014) showed that a CNN with a convolution and pooling layer can produce good performance on several classification tasks including sentiment analysis and question classification. Zhang and Wallace (2017) conducted extensive experiments of CNNs on sentence classification task. The authors concluded several useful suggestions that are helpful for applying CNNs to sentence classification. Aipe et al. (2018) developed a CNN model for multi-label classification of disaster dataset. Specific linguistic features are extracted from tweets to augment the original CNN model.

Based on these previous works, we use a CNN model for the classification of tweets as containing sympathy or not. Figure 1 shows our proposed architecture which is inpired from Y. Kim (2014) and Aipe et al. (2018). The CNN models generally work better when the input is given as word embeddings.

Word embeddings are trained from a very large corpus to generate effective representations for words with similar meanings. This method contributes to the progress of word representations and thus has been widely used in text classification and machine translation areas. Nowadays, it is very convenient for people to apply pre-trained word embeddings(e.g., GoogleNews and Glove) to their applications and get satisfying results. In our paper, we mainly discuss four word embeddings: CrisisNLP, GoogleNews, Glove and Fasttext. Different from the others, CrisisNLP is a word2vec embedding trained in a smaller dataset, which can be regarded as domain embedding.

A common problem existent in current word embeddings is that semantic consistency is hard to maintain in vector space, which means the most similar words of a target word in word space are irrelevant to the target word in some cases. And meanwhile, the synonyms of target word are far from the location of target word in the vector space. This problem is clearly illustrated in Figure 2a. The red numbers in this figure are cosine similarities between the target word and its synonyms and all other words are top 20 most similar words of target word 'pray'. Target word 'prays' is far from its synonyms and thus has low cosine similarities with those words. Therefore, it restricts potential improvement of some fine-tuned classifiers.

In order to address similar problems in sentiment analysis, Yu et al. (2017) proposed a refinement model to converge sentimentally similar words by using sentiment lexicon. The results show that it has good performance in both binary and fine-grained sentiment classification tasks. However, this model is only designed to improve the sentiment embedding or conventional embedding applied in sentiment analysis. Therefore, inspired by this model, a more general refinement method of conventional word embedding is proposed in this section. The procedure for refining word embeddings can be divided into three stages. First, Snowball stemmer is used to cluster words with similar meanings, which means that words with same stem are regarded as one cluster. Second, words in each cluster move close to each other in the vector space to get better representations, following the approach described below. Finally, the refined embedding is fed as the input of CNN model and yields better performance in sympathy detection task. As we can see in Figure 2b, the synonyms of the target word move much closer to the target word in the vector space with refinement model, which brings better word representations in sympathy detection.

Snowball stemmer is by far the most popular stemmer and performs better when compared with original Porter stemmer. The comparison between original Porter stemmer and Snowball stemmer is shown in Table 2. We can clearly observe that Snowball stemmer is more precise than Porter stemmer. For all selected vocabulary words in conventional word embedding, we have over 40,000 clusters in total after running Snowball stemmer and these clusters are used to improve the word representations in vector space. After we obtain the word clusters for original

Table 2. Porter stemmer vs. Snowball stemmer

Stemmer	Word cluster
Porter stemmer 'gener'	'general', 'generation', 'generator', 'generous', 'generations', 'generate', 'generally', 'generated', 'generals', 'generic', 'generously', 'generators', 'generating', 'generates', 'gener', 'generational', 'generalize', 'generalizing', 'genere', 'generalizations', 'generalization', 'generalized', 'generics', 'generative', 'generalities', 'generou', 'generality', 'generically'
Snowball stemmer 'generat'	'generation', 'generator', 'generations', 'generate', 'generated', 'generators','generating', 'generates', 'generational', 'generat', 'generative'

word embedding, we'd like words in each cluster to move closer to each other in vector space and meanwhile not too far from their original locations. An objective function is presented to minimize the total distance among each cluster, which is defined as

$$\Phi(V) = \sum_{i=1}^{n} \sum_{j=1}^{k} w_{ij} dist(v_i, v_j) \tag{1}$$

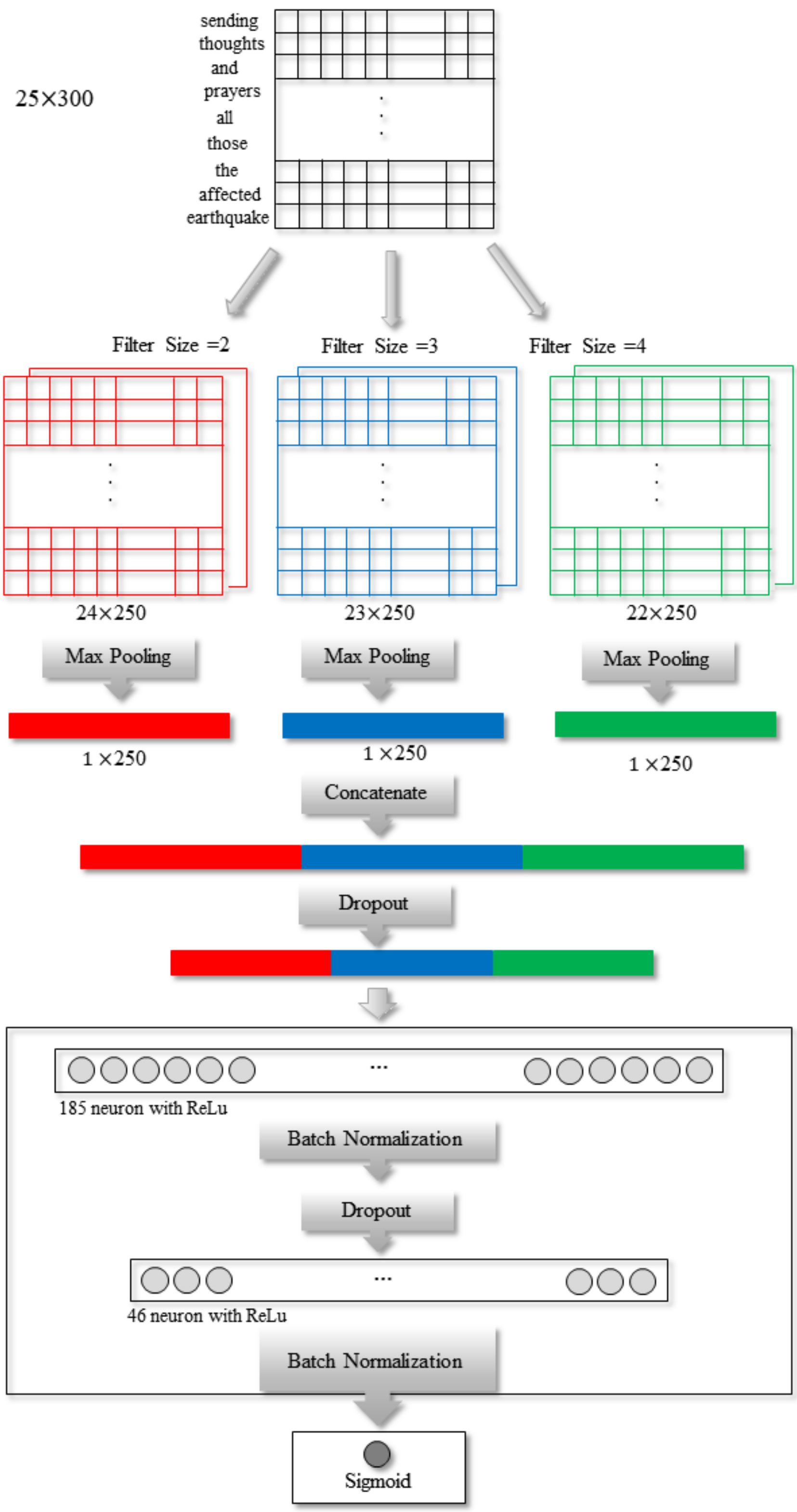
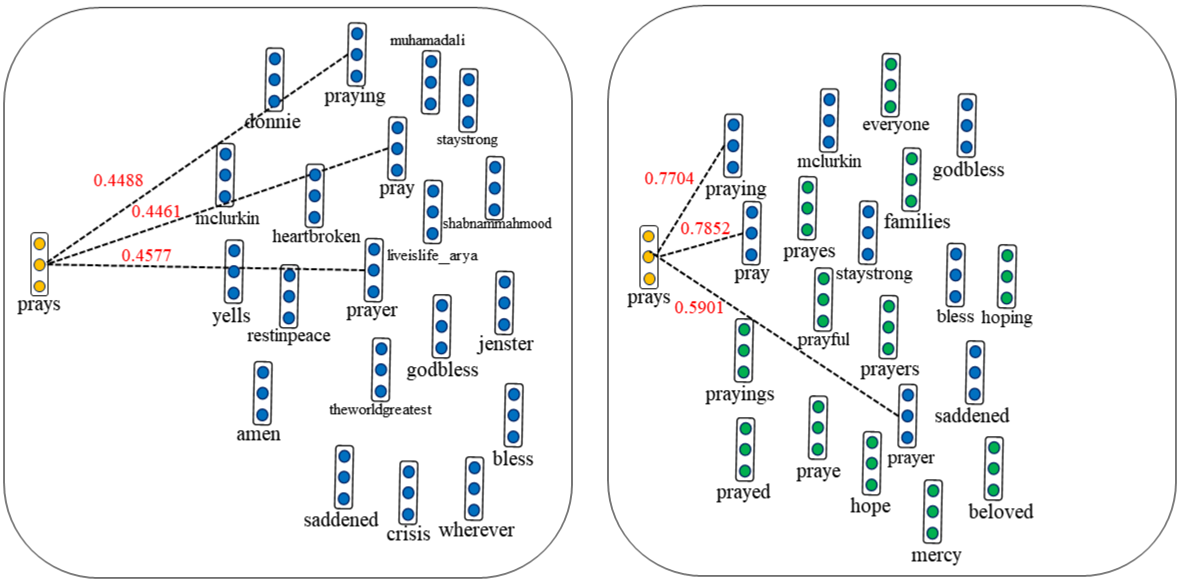


Figure 1. CNN Model



(a) Word embedding without refinement model (b) Word embedding with refinement model

Figure 2. Conceptual diagram of word space for CrisisNLP embedding

where n denotes total number of clusters and k denotes the size of each cluster, dist(v_i,v_j) denotes the distance between target word v_i and similar word v_j, w_{ij} denotes the weight of each similar word related to target word. Each target word v_i has k-1 similar words within one cluster and the number of k may vary due to different size of clusters. Square Euclidean distance is used to calculate the distance between the target word and its similar words, defined as

$$dist(v_i, v_j) = \sum_{d=1}^{D} (v_i^d - v_j^d)^2 \tag{2}$$

where D is the dimension of pre-trained word embedding and w_{ij} is defined below

$$w_{ij} = \frac{1}{order_j} \tag{3}$$

where $order_j$ denotes the order of word in each cluster. Once we finish the word stemming, we have a fixed order for words in each cluster. In order to avoid putting weights on specific word in cluster, we randomly shuffle the order of each cluster when we run refinement model. To prevent refined model from destroying good word representations, a constraint, which is defined as the first part of objective function, is added to make sure the target word can only move with small steps in vector space. The objective function is defined as

$$argmin\Phi(V) = argmin \sum_{i=1}^{n} \left[\alpha dist(v_i^{t+1}, v_i^t) + \beta \sum_{j=1}^{k} w_{ij} dist(v_i^{t+1}, v_j^t) \right] \tag{4}$$

where v_i^{t+1} denotes word vector of target word in t+1 step, parameter α and β determine how far the target word can move away from its original location in vector space. To minimize the objective function, we set $\frac{\partial \Phi(V)}{\partial v_i^{t+1}}$=0 and target word vector is updated by

$$v_i^{t+1} = \frac{\alpha v_i^t + \beta \sum_{j=1}^{k} w_{ij} v_j^t}{\alpha + \beta \sum_{j=1}^{k} w_{ij}} \tag{5}$$

We need to tune three parameters: the number of iterations, α and β. The number of iterations controls the degree to which words converge in the vector space. α and β determine how far the target word can move for each iteration. In our experiments, iteration number=1, α=10 and β=1.

Table 3. # of Tweets for each label in CrisisNLP Data

Class	Train(70%)	Val(20%)	Test(10%)	# of Tweets
Sympathy and Emotional Support	1,378	394	197	1,969
Caution and Advice	743	212	107	1,062
Displaced People and Evacuations	442	126	63	631
Donation Needs or Offeres or Volunteering Services	1,827	522	261	2,610
Infrastructure and Utilities Damage	1,296	370	185	1,851
Injured or Dead People	1,754	501	251	2,506
Missing Trapped or Found People	281	80	40	401
Not Related or Irrelevant	1,642	469	234	2,345
Other Useful Information	4,004	1,144	572	5,720
# of Tweets	13,367	3,818	1,910	19,095

DATASET

The dataset used in experiments comes from CrisisNLP (Imran, Mitra, et al. 2016). The dataset consists of tweets posted during 11 different disasters that happened during 2013 to 2015. Each tweet is annotated by paid CrowdFlower Workers using situational awareness categories. The disasters in this collection can be classified as two categories: natural disasters and medical disasters. First category has 8 labels and second category has 7 labels. In this paper, we only used natural disasters data since medical disasters data does not contain sympathy labeled tweets.

As a result, we use 9 different disasters and the total number of tweets is 19,095. The detail information of the dataset labels/categories and the number of tweets in each category are shown in Table 3. All tweets are posted only in English. Since we have a purpose of extracting sympathy tweets, we mainly focus on Twitter postings that have label of "sympathy and emotional support" and treat this as positive class. All other labels are considered as the negative class. The number of tweets containing sympathy is 1,969 and the number of non-sympathy tweets is 17,126.

EXPERIMENTS AND RESULTS

Preprocessing

This step contains removing Twitter ID, hashtags, special characters and URLs. Moreover, OOV dictionary is used to process slangs and abbreviations. Lastly, we remove words and sentences with length smaller than 3.

CNN performance

The proposed CNN model, which is shown in Figure 1, is based on Y. Kim (2014) and Aipe et al. (2018). The CNN model consists of a convolutional layer followed by a max-pooling, concatenation and two fully-connected layers with batch normalization. More specifically, with regard to convolutional layer, there are 250 feature maps for each kernel size {2,3,4} and these feature maps are then max-pooled and concatenated as the input of fully-connected layer. Two dropout layers are added to reduce overfitting of neural network and drop rate is 0.5.

In experiments, the proposed CNN model, as well as the LSTM and Bi-LSTM models, are implemented on the disaster dataset described earlier. Ten different random seeds (from 5 to 50 with interval 5) are selected to split the pre-processed dataset and the final result is the average score of the 10 experiments. The proposed refined word embedding method is implemented in four different pre-trained word embeddings and tested on all deep learning frameworks. As we can observe from Table 4, all refined models outperform original models. More specifically, for example, refined CNN improves original CNN model with word embedding CrisisNLP, word2vec (GoogleNews), Glove and Fasttext by 0.25%, 0.57%, 0.48% and 1.21% respectively. Overall, CNN is generally the best performing algorithm with or without refinement model. It produces the best performance on word2vec, Glove and FastText embedding. Even though the refined Bi-LSTM model has the best result (75.08%) in CrisisNLP embedding, however the refined CNN with FastText embedding yields a better result (75.78%).

From our experimental results, we can also conclude that there exists at most 0.95%, 1.23%, 1.44% and 1.21% increase in CrisisNLP, word2vec, Glove and FastText embedding, respectively. Table 4 also shows that FastText word embedding generally performs better than other word embeddings. Four out of six models with FastText word embedding yield better performance over the others. To conclude, the refined CNN model with FastText embedding achieves the highest F1 score, which shows that the CNN architecture has the ability to solve specific text classification tasks such as sympathy detection and sentiment analysis.

Table 4. Performance of CNN and other frameworks on sympathy detection(F1 Score)

Model	CrisisNLP	word2vec	Glove	Fasttext_Crawl
Refined CNN	74.26	**75.39**	**75.42**	**75.78**
Refined LSTM	74.64	74.07	75.15	74.38
Refined Bi-LSTM	**75.08**	75.18	74.48	75.33
CNN	74.01	74.82	74.94	74.57
LSTM	74.03	72.84	73.71	74.38
Bi-LSTM	74.13	74.03	73.44	74.44

CNN vs. Baseline Classifiers

We compare our proposed refined CNN model with baseline machine learning algorithms such as Logistic Regression, Decision Tree, Stochastic Gradient Descent(SGD), AdaBoost , Support Vector Machine(SVM) and CNN-non-static model (Kim, 2014) with word2vec embedding in Table 5. In order to guarantee the correctness of comparisons, for each baseline model, the same training and testing samples are used and the final result is the average of 10 experiments.

From Table 5, we can see that the refined CNN model performs best in recall and F1 measure, which achieve 70.79% and 75.78%, respectively. We also observe that CNN-non-static model has the best result among baseline classifiers, followed by SVM and AdaBoost. Even though Logistic Regression has the highest precision score, it performs the worst in recall due to imbalanced data distribution. In general, CNN has better precision and recall score and thus it can be applied to detect sympathy sentences in highly skewed data.

Table 5. CNN vs. Baseline Classifiers

Model	Precision	Recall	F1-Score
Refined CNN	81.83	**70.79**	**75.78**
Logistic Regression	**95.2**	43.55	59.58
Decision Tree	64.44	66.67	65.47
SGD	90.69	55.77	68.96
AdaBoost	82.9	59.31	69.09
SVM	80.17	65.66	72.15
CNN-non-static (Kim, 2014)	87.99	64.60	74.40

CONCLUSION AND FUTURE DIRECTIONS

In this paper, we explore neural networks (e.g., CNNs, LSTMs) with different word embeddings for sympathy detection from Twitter disaster data. As a starting point for finding candidate donors, we successfully achieve the task of sympathy contents tweets detection using deep learning techniques. Moreover, we present a word stemming based refinement model to improve the existing pre-trained word embeddings. Instead of training new word embeddings, we can easily integrate the proposed refinement model into conventional word embeddings and improve the F1 score of text classification task such as sympathy detection. Experiments on disaster dataset show that proposed refined CNN model produces the best performance when compared with CNN and other deep learning frameworks. The performance of our model is significantly better than traditional classifiers.

In future work, it would be interesting to test our proposed model on various classes of disaster datasets, e.g. ‘donation needs or offers’. Also, it would be very interesting to see if our refinement model can improve text classification under different backgrounds. Moreover, an exciting direction to explore is the relations between sympathy and donations. An inspection of the CrisisNLP dataset reveals that there are many tweets that contain both sympathy and donation related contents. Several tweets that show a potential relationship between sympathy and donations are provided in Table 6. Based on this, our refinement model can also be extended as adding donation encoded features in the last layer of CNN and determine whether the performance of sympathy detection changes.

Table 6. Tweets containing 'Sympathy' and mentioning 'Donations'.

RT@RealLouisAndre: Sending out all my support for all people in #Chile all the prayers are with you, also u can donate @RedCross #PrayForChile
@justinbieber thanks for the challenge but I'm not able to do it so I just donated 30$ do the #ASL hope that makes up for it
RT@DAVIDsTEA: Our hearts go out to everyone in Nepal. Until May 4th, all profits from Nepal Black tea will be donated, to help fund

ACKNOWLEDGMENTS

We thank the National Science Foundation for support from the grants 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 ocial policies, either express or implied, of the National Science Foundation. We also wish to thank our anonymous reviewers for their constructive comments.

REFERENCES

Aipe, A., Mukuntha, N., Ekbal, A., and Kurohashi, S. (2018). "Deep Learning Approach towards Multi-label Classification of Crisis Related Tweets". In: ISCRAM '18.

Alam, F., Ofli, F., Imran, M., and Aupetit, M. (2018). "A Twitter Tale of Three Hurricanes: Harvey, Irma, and Maria". In: ISCRAM '18.

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

Baberini, M., Coleman, C.-L., Slovic, P., and Västfjäll, D. (2015). "Examining the effects of photographic attributes on sympathy, emotions, and donation behavior". In: *Visual Communication Quarterly* 22.2, pp. 118–128.

Baccianella, S., Esuli, A., and Sebastiani, F. (2010). "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining". In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valletta, Malta: European Language Resources Association (ELRA).

Bautista, J. R. and Lin, T. T. (2015). "Social Support on Twitter: The Case of the Philippine's #Fallen44". In: *Cyberpsychology, Behavior, and Social Networking*.

Ben Lazreg, M., Goodwin, M., and Granmo, O.-C. (2016). "Proceedings of the Artificial Intelligence Applications and Innovations: 12th IFIP WG 12.5 International Conference and Workshops, AIAI 2016, Thessaloniki, Greece, September 16-18, 2016". In: *Information Abstraction from Crises Related Tweets Using Recurrent Neural Network*. Ed. by L. Iliadis and I. Maglogiannis. Cham: Springer International Publishing, pp. 441–452.

Bunce, S., Partridge, H., and Davis, K. (2012). "Exploring information experience using social media during the 2011 Queensland floods: a pilot study". In: *The Australian Library Journal* 61.1, pp. 34–45.

Caragea, C., Silvescu, A., and Tapia, A. H. (2016). "Identifying informative messages in disaster events using convolutional neural networks". In: *International Conference on Information Systems for Crisis Response and Management*, pp. 137–147.

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

Derczynski, L., Meesters, K., Bontcheva, K., and Maynard, D. (2018). "Helping Crisis Responders Find the Informative Needle in the Tweet Haystack". In: ISCRAM '18.

Froyum, C. (2018). ""They Are Just Like You and Me": Cultivating Volunteer Sympathy". In: *Symbolic Interaction*.

Han, B., Ryzhov, I. O., and Defourny, B. (2013). "Efficient learning of donor retention strategies for the American Red Cross". In: *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World*. IEEE Press, pp. 17–28.

Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., and Meier, P. (2013). “Practical extraction of disaster-relevant information from social media”. In: *22nd International World Wide Web Conference, WWW ’13, Rio de Janeiro, Brazil, May 13-17, 2013, Companion Volume*, pp. 1021–1024.

Imran, M., Mitra, P., and Castillo, C. (2016). “Twitter as a Lifeline: Human-annotated Twitter Corpora for NLP of Crisis-related Messages”. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Portoroz, Slovenia: European Language Resources Association (ELRA).

Kaigo, M. (2012). “Social media usage during disasters and social capital: Twitter and the Great East Japan earthquake”. In: *Keio Communication Review* 34.1, pp. 19–35.

Kim, S., Bak, J., and Oh, A. H. (2012). “Do You Feel What I Feel? Social Aspects of Emotions in Twitter Conversations”. In: *Proceedings of the Sixth International Conference on Weblogs and Social Media, Dublin, Ireland, June 4-7, 2012*.

Kim, Y. (2014). “Convolutional Neural Networks for Sentence Classification”. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1746–1751.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11, pp. 2278–2324.

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

Li, H., Herndon, N., Caragea, D., and Caragea, C. (2017). “Disaster Response Aided by Tweet Classification with a Domain Adaptation Approach”. In: *Journal of Contingencies and Crisis Management (JCCM), Special Issue on HCI in Critical Systems. In press*.

Mandel, B., Culotta, A., Boulahanis, J., Stark, D., Lewis, B., and Rodrigue, J. (2013). “A demographic analysis of online sentiment during hurricane Irene”. In: Workshop on Language in Social Media.

Morstatter, F., Lubold, N., Pon-Barry, H., Pfeffer, J., and Liu, H. (2014). “Finding Eyewitness Tweets During Crises”. In: *ACL 2014 Workshop on Language Technologies and Computational Social Science*. Baltimore, Maryland, USA.

Nagy, A. and Stamberger, J. (2012). “Crowd Sentiment Detection during Disasters and Crises”. In: Proceedings of the 9th International ISCRAM Conference.

Neppalli, K., Caragea, C., Squicciarini, A., Tapia, A., and Stehle, S. (2017). “Sentiment Analysis during Hurricane Sandy in Emergency Response.” In: *The International Journal of Disaster Risk Reduction (IJDRR 2017)*.

Neppalli, V. K., Caragea, C., and Caragea, D. (2018). “Deep Neural Networks versus Naïve Bayes Classifiers for Identifying Informative Tweets during Disasters”. In: *Proceedings of the 15th International Conference on Information Systems for Crisis Response and Management*. ISCRAM ’18. Rochester, New York, USA.

Nguyen, D. T., Al-Mannai, K., Joty, S. R., Sajjad, H., Imran, M., and Mitra, P. (2017). “Robust Classification of Crisis-Related Data on Social Networks Using Convolutional Neural Networks.” In: *ICWSM*, pp. 632–635.

Nielsen, F. Å. (2011). “A new ANEW: Evaluation of a word list for sentiment analysis in microblogs”. In: *CoRR* abs/1103.2903.

Pfitzner, R., Garas, A., and Scheweitzer, F. (2012). “Emotional Divergence Influences Information Spreading in Twitter”. In: Proceedings to the 6th International AAAI Conference on Weblogs and Social Media. AAAI Press, pp. 543–546.

Purohit, H., Castillo, C., Imran, M., and Pandev, R. (2018). “Social-eoc: Serviceability model to rank social media requests for emergency operation centers”. In: *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, pp. 119–126.

Rudra, K., Sharma, A., Ganguly, N., and Ghosh, S. (2018). “Characterizing and Countering Communal Microblogs During Disaster Events”. In: *IEEE Transactions on Computational Social Systems* 5.2, pp. 403–417.

Rudra, K., Ghosh, S., Ganguly, N., Goyal, P., and Ghosh, S. (2015). “Extracting situational information from microblogs during disaster events: a classification-summarization approach”. In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, pp. 583–592.

Sarda, P. and Lal Chouhan, R. (2017). “Extracting Non-Situational Information from Twitter During Disaster Events”. In: *Journal of Cases on Information Technology* 19, pp. 15–23.

Schulz, A., Thanh, T., Paulheim, H., and Schweizer, I. (2013). “A Fine-Grained Sentiment Analysis Approach for Detecting Crisis Related Microposts”. In: ISCRAM 2013.

Tanev, H., Zavarella, V., and Steinberger, J. (2017). “Monitoring disaster impact: detecting micro-events and eyewitness reports in mainstream and social media”. In: *Proceedings of the 14th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2017)*. France.

Verma, S., Vieweg, S., Corvey, W. J., Palen, L., Martin, J. H., Palmer, M., Schram, A., and Anderson, K. M. (2011). “Natural Language Processing to the Rescue? Extracting" Situational Awareness" Tweets During Mass Emergency.” In: *ICWSM*. Barcelona, pp. 385–392.

Yin, J., Lampert, A., Cameron, M., Robinson, B., and Power, R. (2012). “Using social media to enhance emergency situation awareness”. In: *IEEE Intelligent Systems* 27.6, pp. 52–59.

Yu, L.-C., Wang, J., Lai, K. R., and Zhang, X. (2017). “Refining word embeddings for sentiment analysis”. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 534–539.

Zahra, K., Imran, M., Ostermann, F. O., Boersma, K., and Tomaszewski, B. (2018). “Understanding eyewitness reports on Twitter during disasters”. In: *ISCRAM 2018 Conference Proceedings*, pp. 687–695.

Zhang, Y. and Wallace, B. (2017). “A Sensitivity Analysis of (and Practitioners’ Guide to) Convolutional Neural Networks for Sentence Classification”. In: *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Vol. 1, pp. 253–263.