Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 167 (2020) 962–968

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2019)

Improved Classification of Crisis-Related Data on Twitter using Contextual Representations

Sreenivasulu Madichetty, Sridevi M

Department of Computer Science and Engineering

National Institute of Technology,Tiruchirappalli, Tamilnadu-620015,India

Abstract

In the past few years, neural network based representations (word embeddings) have been dramatically used for Natural Language Processing (NLP) tasks. Word embeddings play a significant role in deep learning NLP tasks. Therefore, finding the best word embeddings to a specific task is an important task. A dense classifier with contextual representations (Embeddings from language models) is proposed to use for classifying crisis related data on social networks during a disaster. Three real-time Twitter datasets such as Nepal Earthquake, California Earthquake, and Typhoon hagupit are used and the performance is analyzed with parameters such as precision, recall, f1-score and accuracy. The proposed dense classifier with ELMo embeddings model gives better accuracy than the traditional classifier (Support Vector Machine) and deep learning classifiers (Convolutional Neural Network and MultiLayer Perceptron-Convolutional Neural Network with Crisis word embedding).

Keywords: ELMo Embeddings; Crisis; Dense classifier;Twitter

1. Introduction

Twitter is used as a communication medium in many social events [23] such as traffic accidents [25], recommendation system, disaster [9], etc. Classifying crisis related tweets is an important NLP task [7, 10, 15, 20, 22, 24] because different types of tweets such as rumor and spam tweets [2, 19, 4], emotion information [1, 3, 26], etc., are posted during a disaster. In traditional methods, different features are handcrafted for classifying the specific category of tweets during a crisis. Recently, deep learning models with word embeddings are used in most of the crisis-related classification tasks [15]. Word embedding plays a major role in deep learning models for classification of tweets. Therefore, the authors in [15] used CNN for classifying crisis-related data. This paper consider the problem as a multi-classification problem and uses different classes mentioned below:

1. Affected Individual (Aff_Ind) :- describes the deaths, injuries, found people, etc.

10.1016/j.procs.2020.03.395

2. Donations and Volunteering (Don_and_Vol):- describes the tweets containing information related donations and volunteering services.

3. Infrastructure and utilities (Infra_and_Uti):- denotes the infrastructure damage and utilities.

4. Sympathy and Support (Symp_and_Supp):- contains sympathy for emotional support.

5. Other Useful Information (Other_Use_Infor):- contains other useful information related to the crisis.

6. Not related or irrelevant:- denotes tweets not related to the crisis.

Table 1 shows the example tweets for all classes considered in this paper. With the advancement of Contextualized representations such as ELMo embedding gives further improvement than the Google embedding and crisis embedding. Therefore, proposed to use a dense classifier with ELMo embeddings for classifying crisis data.

Table 1: Example tweets of different classes

Classes	Example Tweets
Aff_Ind	RT @josephjett: UPDATE 10-California wine country shaken by 6.0 quake, dozens hurt via @josephjett #bonds [URL]
Don_and_Vol	RT @WayneRooney: Please support @Unicef_UK helping children in danger from #earthquake in #Nepal - [URL] PLS RT
Infra_and_Uti	RT @BBCWorld: Our reports show #NepalEarthquake destruction, including epicentre [URL] [URL]
Symp_and_Supp	Dua's for all those affected by the earthquakes in India,Nepal & amp Bhutan. Stay safe & amp help others in any form. #Equake [URL]
Other_Use_Infor	Live: Nepal cabinet meets to seek foreign help, 114 feared dead after massive quake via @firstpostin [URL]
Not_rel	RT @Smallotsbigwine: The Napa Earthquake Wine Bucket Challenge. [URL] [URL]

The contributions are summarized as follows:

1. ELMo embeddings with dense classifier are used for classifying the crisis related tasks during a disaster.
2. Extensive experiments are performed on different datasets such as California and Nepal Earthquake, Typhoon hagupit and evaluated using different parameters.

This paper is organized as follows: Section 2 explains literature survey of the crisis related tasks. Section 3 explains the ELMo embeddings with dense classifier. The fourth section describes the experiments and performance analysis of the proposed model. Finally, Section 5 concludes the paper.

2. Related Work

Traditional methods of NLP classification tasks[10, 21, 22, 24] used handcrafted features in . In [24], the authors used bags-of-words models for classifying the situational information during a disaster. The features such as uni-grams, bi-grams, Parts-Of-Speech tags (POS), etc., are dependent on the vocabulary of the tweets. Similarly, the authors in [10] developed a system which is named Artificial Intelligence Disaster Response (AIDR), and it uses uni-gram and bi-gram features for classifying crisis related data during a disaster. In [22], the authors mined the informative words from the tweets as features for detecting the resource tweets during a disaster. The authors [21, 20] developed a low-level lexical and syntactic features which are independent of vocabulary words for classifying the situational information during a disaster. The low-level lexical features has presence of wh-words, presence of numerals, count of non-situational words, etc., and syntactic features has count of personal pronouns, count of intensifiers, etc. SVM classifier is used for classifying the situational information and summarizes the tweets using the content words. Later, it is extended to Non-English tweets. In [5], the authors developed a classification-ranking approach for detecting the fact-checkable microblogs during disaster. In their method, hand-crafted features and SVM classifier with linear kernel is used for classification. However, these methods focus only on specific class tweets and need feature engineering for classification.

Word embeddings received much attention due to the automatic extraction of features from the tweets. The authors in [7, 16] used the CNN with word embeddings and showed outstanding performance than hand-crafted features in crisis-related tasks during a disaster. The BOW hand-crafted features, and the classifiers such as SVM and ANN are used. The work [15] uses CNN and MLP-CNN with crisis embedding for classifying the crisis related data. Crisis embedding used a skip-gram model of word2vec tool [14] from an enormous corpus crisis related tweets nearly 57,908 tweets. In[13], the authors developed a deep learning model based on the word embeddings for detecting the informative tweets during disaster. They used the crisis specific related word embeddings for detecting the informative tweets during disaster. In [6], the authors developed a information retrieval methodologies based on the combination of word embeddings and character embeddings for extracting the resource need and availability tweets during disaster. They showed that their method performs well than word2vec. The main drawback of this crisis word embeddings is that it gives the same word vector for the same word in a different context. So, it doesn't disambiguate the words in a different context of a tweet. A new ELMo embedding overcomes this problem with a simple dense classifier for classifying crisis-related data during a disaster. The existing methods for classifying crisis related tweets are summarized and tabulated in Table 2.

3. Proposed Method

This section describes a dense classifier with ELMo embeddings [18] for classifying crisis related data. The block diagram of the proposed method is shown in Fig. 1. The steps involved in the proposed method are given below:

1. Tweets are given as an input to the Embedding layer.
2. Tweets are splits into two parts. (i). 80% of tweets are used for training and (ii). 20% of tweets are used for testing.
3. Training tweets are sent to the embedding layer.
4. Embedding layer gives the word vectors from the tweets.
5. Word vectors are sent to the denser classifier for training the model.
6. Trained model is used for predicting the label of the testing tweets.
7. The output of the dense classifier can be predicted by the class label (either crisis related tweet or not).

Embedding layer and dense classifier are explained in the sections 3.1 and 3.2 respectively.

3.1. Embedding Layer

Embedding layer divides the tweets into words and generates the vector for each word in a tweet by using pre-trained ELMo model. It combines all the word vectors in a tweets by averaging all the word vectors into fixed size (1024) representation of a tweet. Pre-trained ELMo Embeddings are learned using Bi-directional Language Models (BiLM) and trained on a large text corpus. Unlike other word embeddings [11], it learns word vector based on the context of a words in a tweet. Therefore, ELMo embeddings are known as context-dependent representations or contextualized representations. The lower and higher layers of BiLM captures the syntax and semantics of a words in a tweet. It uses the character convolutions of a word in a tweets. Therefore, it don't have out-of-vocabulary problem and gives the different word vector for the same word in a different contexts unlike traditional word embeddings. Consider an example word 'bank', this word gives different meaning in a different contexts, in some contexts it is a financial institution and sometimes river bank. The output of embedding layer is sent to the dense layer for further processing.

Table 2: Summary of classification of crisis related tweets

Author & Year	Features	Problem statement	Limitations
Verma et al. [24] & 2011	Bag-of-words model	Classifying situational and non-situational information	Features are vocabulary dependent.
Imran et al. [10] & 2014	n-gram features	User defined categories.	It doesn't focus on specific class tweets or multi-classes.
Rudra et al. [21] & 2015	Low-level lexical and syntactic features.	Extracting situational information.	It doesn't work for new vocabulary (which are not present in lexicon).
Sreenivasulu M et al. [22] & 2017	Informative features	Identifying the resources.	Features don't work properly for multi-classification tasks because these are problem specific.
Rudra et al. [20] & 2018	Low-level lexical and syntactic features	Extracting situational information for hindi tweets	These features don't work for multi-classes.
Barnawal et al. [5] & 2019	Bag-of-words model	Identification of Fact-checkable microblogs.	It doesn't focus on different classes of tweets and only focused on Fact-checkable micro-blogs.
Caragea et al. [7] & 2016	high dimensional one hot vectors	Identifying informative messages	It doesn't semantic relationship between the words.
Nguyen et al. [15] & 2017	Crisis word embeddings	Classification of crisis related tweets into multiple categories.	It doesn't handle out-of-vocabulary words.
Sreenivasulu M et al. [13] & 2019	Detection of informative tweets.	Crisis word embeddings	It doesn't used multi-classification approach and doesn't handle out-of-vocabulary words.
Basu et al. [6] & 2019	Word embeddings and character embeddings	Extracting resource needs and availabilities of tweets	It gives same word vector for the same word in different contexts and doesn't differentiate based on the contexts.

3.2. Dense Classifier

A dense classifier is used for ELMo embeddings similar to [15] where CNN is used as a classifier for crisis embeddings for training and testing the data. The data is taken from the Embedding layer. It contains two dense layers namely dense layer with Rectifier Linear Unit (ReLU) activation function and another one is dense layer with softmax function. The first dense layer contains n_1 number of neurons and it produces n_1 dimensional vector output. The generated n_1 dimension is fed to the second dense layer. Second dense layer contains n_2 number of neurons which is equal to the number of classes. It generates the output of n_2 dimensional vector for each tweet. The n_2 dimensional vector predicts the label based on the maximum value. The values of n_1 and n_2 are set to be 256 and 6.

4. Experiment results and analysis

This section describes the datasets used for experiments and performance analysis of a dense classifier with ELMo embeddings. Experiments are performed using the Keras [8] library and tensor-flow hub of python language. Early stopping criteria is applied by using validation accuracy, the number of epochs is set to 25, the batch size of 128 is considered and Adam optimizer is used for optimization.

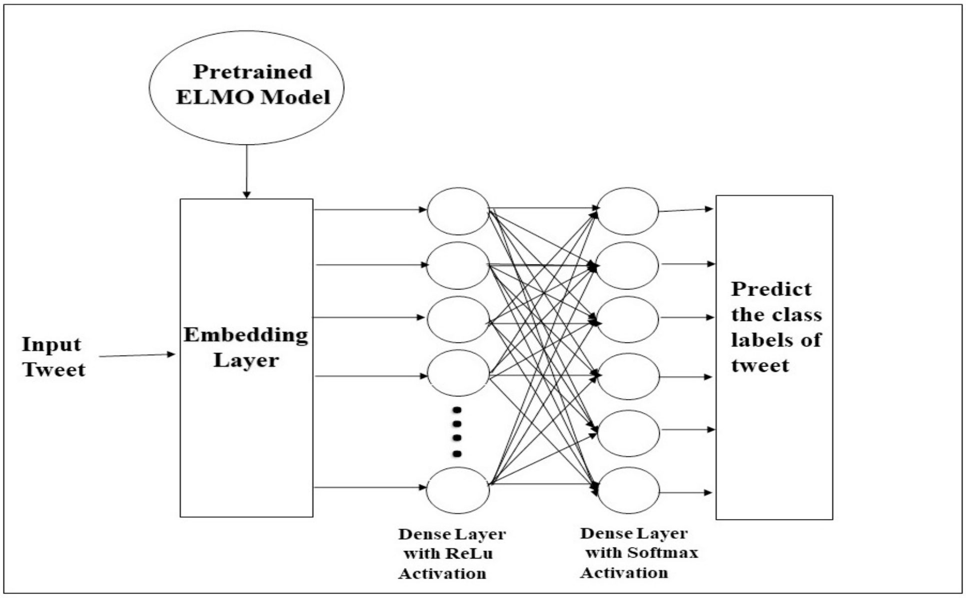

Fig. 1: Block diagram of Dense Classifier using ELMo Embeddings for classifying crisis related tweets

4.1. Datasets

The datasets such as Nepal earthquake, California Earthquake and Typhoon Hagupit are used and they were collected from the multiple sources such as CrisisNLP [12], CrisisLex [17] and AIDR [10]. For CrisisNLP and CrisisLex paid workers are used for labeling the tweets in different classes similar to [15]. AIDR volunteers labeled the tweets for different classes in different crisis. However, in this work, three datasets such as Nepal Earthquake, California Earthquake and Typhoon Hagupit are used for experiments. The details of the disaster are shown in Table 3. Out of the data, 70% of tweets are used for training, 10% of tweets are used for validation and 20% of tweets are used for testing.

Table 3: Details of disaster dataset

Events	Total Number of Labeled Tweets
Nepal Earthquake	12140
California Earthquake	2012
Typhoon Hagupit	11487

4.2. Performance Analysis

The accuracy, precision, recall and f1-score parameters are used to evaluate the performance of the dense classifier with ELMo embeddings for classifying the crisis related data. The proposed model is compared with the existing approach [15] such as SVM, CNN and MLP-CNN with crisis embeddings. Table 4 shows the precision, recall and f1-score values for all classes on Nepal Earthquake, California earthquake and Typhoon Hagupit dataset. And also reported results micro-average, macro-average and weighted average of all the parameters. It is observed that macro-average precision of Nepal Earthquake have highest value compared to the other datasets.

And also observed that weighted average value of precision, recall and f1-score got similar range of values for Nepal and california earthquake but different range for typhoon hagupit. Table 5 shows the accuracies of the dense

Table 4: Dense Classifier with ELMo Embeddings using different parameters on Nepal Earthquake, California Earthquake and Typhoon Hagupit

	Nepal Earthquake			California Earthquake			Typhoon Hagupit		
Classes	**Precision**	**Recall**	**f1-score**	**Precision**	**Recall**	**f1-score**	**Precision**	**Recall**	**f1-score**
Aff_Ind	75	35	48	90	81	85	67	4	8
Don_and_Vol	66	54	59	73	73	73	0	0	0
Infra_and_Uti	68	18	28	67	71	69	46	8	14
Symp_and_Supp	72	25	38	71	83	77	83	26	40
Other_Use_Infor	79	96	86	83	88	85	91	99	95
Not_rel	50	2	4	11	2	3	40	6	11
Micro-Average	78	78	78	79	79	79	91	91	91
Macro-Average	69	38	44	66	66	65	55	24	28
Weighted-Average	76	78	73	76	79	77	88	91	88

classifier with ELMo embeddings on three different datasets. It is observed that from Table 5 typhoon hagupit got highest values compared to earthquake datasets. It might be the reason of typhoon hagupit tweets have the words that are present in a different contexts.

Table 5: Accuracy of different events using dense classifier with ELMo embeddings

Events	**Accuracy**
Nepal Earthquake	77.57
California Earthquake	78.79
Typhoon Hagupit	90.90

The existing methodologies such as SVM classifier with Bag-of-words (BoW) model features are used as baseline-1, CNN with crisis word embeddings [15] are used as baseline-2 and MLP-CNN [15] is used as another baseline (baseline-3). MLP-CNN outperforms the SVM with BoW and CNN on three datasets for classifying crisis related tweets into multiple categories. ELMo embeddings with dense classifier give the best accuracy for classifying crisis-related data compared to existing three baselines and it is reported in Table 6.

Table 6: Comparison between the existing methods and proposed method on three disaster dataset using Accuracy parameter

Methods	**Nepal Earthquake**	**California Earthquake**	**Typhoon Hagupit**
SVM	70.45	75.66	75.45
CNN [15]	72.98	77.80	81.82
MLP-CNN [15]	73.19	76.85	82.12
Proposed Method	**77.57**	**78.79**	**90.90**

5. Conclusion and Future work

In this paper, a simple classifier model with ELMo embeddings have been proposed for classifying crisis-related data during a disaster. They are categorized into Affected Individual, Donations and Volunteering, Infrastructure and utilities, Sympathy and Support, Other Useful Information and irrelevant to disaster. The performance of proposed method are compared with traditional word representations on different disaster text corpora for multi-classification tasks. The current study is limited to in-domain (training and testing with same dataset) only. In the future, it can be extended for cross-domain (training and testing with past and future data) also for different disaster datasets. Further, there is a scope to improve the performance of the model by using advanced deep learning architectures for classifying crisis-related data during a disaster.

References

[1] Torkildson, M.K., Starbird, K., Aragon, C., (2014). Analysis and visualization of sentiment and emotion on crisis tweets, in: International Conference on Cooperative Design, Visualization and Engineering, Springer. pp. 64–67.

[2] Mondal, T., Pramanik, P., Bhattacharya, I., Boral, N., Ghosh, S., (2018). Analysis and early detection of rumors in a post disaster scenario. Information Systems Frontiers 20, 961–979.

[3] Madisetty, S., Desarkar, M.S., (2017)a. An ensemble based method for predicting emotion intensity of tweets, in: International Conference on Mining Intelligence and Knowledge Exploration, Springer. pp. 359–370.

[4] Gupta, H., Jamal, M.S., Madisetty, S., Desarkar, M.S., (2018). A framework for real-time spam detection in twitter, in: 2018 10th International Conference on Communication Systems & Networks (COMSNETS), IEEE. pp. 380–383.

[5] Barnwal, D., Ghelani, S., Krishna, R., Basu, M., Ghosh, S., (2019). Identifying fact-checkable microblogs during disasters: a classification-ranking approach, in: Proceedings of the 20th International Conference on Distributed Computing and Networking, ACM. pp. 389–392.

[6] Basu, M., Shandilya, A., Khosla, P., Ghosh, K., Ghosh, S., (2019). Extracting resource needs and availabilities from microblogs for aiding post-disaster relief operations. IEEE Transactions on Computational Social Systems 6, 604–618.

[7] Caragea, C., Silvescu, A., 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.

[8] Chollet, F., et al., (2015). Keras. `https://github.com/fchollet/keras`.

[9] Imran, M., Castillo, C., Diaz, F., Vieweg, S., (2015). Processing social media messages in mass emergency: A survey. ACM Computing Surveys (CSUR) 47, 67.

[10] Imran, M., Castillo, C., Lucas, J., Meier, P., Vieweg, S., (2014). Aidr: Artificial intelligence for disaster response, in: Proceedings of the 23rd International Conference on World Wide Web, ACM. pp. 159–162.

[11] Imran, M., Mitra, P., Castillo, C., (2016)a. 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), European Language Resources Association (ELRA), Paris, France.

[12] Imran, M., Mitra, P., Castillo, C., (2016)b. Twitter as a lifeline: Human-annotated twitter corpora for nlp of crisis-related messages. arXiv preprint arXiv:1605.05894 , 1638–1643.

[13] Madichetty, S., Sridevi, M., (2019). Detecting informative tweets during disaster using deep neural networks, in: 2019 11th International Conference on Communication Systems & Networks (COMSNETS), IEEE. pp. 709–713.

[14] Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J., (2013). Distributed representations of words and phrases and their compositionality, in: Advances in neural information processing systems, pp. 3111–3119.

[15] Nguyen, D.T., Al Mannai, K.A., Joty, S., Sajjad, H., Imran, M., Mitra, P., (2017). Robust classification of crisis-related data on social networks using convolutional neural networks, in: Eleventh International AAAI Conference on Web and Social Media.

[16] Nguyen, D.T., Mannai, K.A.A., Joty, S., Sajjad, H., Imran, M., Mitra, P., (2016). Rapid classification of crisis-related data on social networks using convolutional neural networks. arXiv preprint arXiv:1608.03902 .

[17] Olteanu, A., Castillo, C., Diaz, F., Vieweg, S., (2014). Crisislex: A lexicon for collecting and filtering microblogged communications in crises, in: Eighth International AAAI Conference on Weblogs and Social Media.

[18] Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L., (2018). Deep contextualized word representations, in: Proc. of NAACL.

[19] Madisetty, S., Desarkar, M.S., (2018). A neural network-based ensemble approach for spam detection in twitter. IEEE Transactions on Computational Social Systems 5, 973–984.

[20] Rudra, K., Ganguly, N., Goyal, P., Ghosh, S., (2018). Extracting and summarizing situational information from the twitter social media during disasters. ACM Transactions on the Web (TWEB) 12, 17.

[21] Rudra, K., Ghosh, S., Ganguly, N., Goyal, P., 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.

[22] Sreenivasulu, M., Sridevi, M., (2017). Mining informative words from the tweets for detecting the resources during disaster, in: International Conference on Mining Intelligence and Knowledge Exploration, Springer. pp. 348–358.

[23] Sreenivasulu, M., Sridevi, M., (2018). A survey on event detection methods on various social media, in: Recent Findings in Intelligent Computing Techniques. Springer, pp. 87–93.

[24] Verma, S., Vieweg, S., Corvey, W.J., Palen, L., Martin, J.H., Palmer, M., Schram, A., Anderson, K.M., (2011). Natural language processing to the rescue? extracting" situational awareness" tweets during mass emergency., Citeseer. pp. 385–392.

[25] Zhang, Z., He, Q., Gao, J., Ni, M., (2018). A deep learning approach for detecting traffic accidents from social media data. Transportation research part C: emerging technologies 86, 580–596.

[26] Madisetty, S., Desarkar, M.S., (2017)b. Nsemo at emoint-2017: an ensemble to predict emotion intensity in tweets, in: Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 219–224.