

Comparing the effectiveness of Convolutional Neural Network and Long Short-Term Memory Network for Disaster Based Social Media Messages—Using Thunderstorm and Cyclone as Case Studies

Annie Singla
Centre of Excellence in Disaster
Mitigation & Management
Indian Institute of Technology
Roorkee, India
asingla@dm.iitr.ac.in

Rajat Agrawal
Department of Management
Studies
Indian Institute of Technology
Roorkee, India
rajat@ms.iitr.ac.in

Nguyen Thi Dieu Linh
Science and Technology
Department
Hanoi University of Industry,
Vietnam
nguyenlinh79.haui@gmail.com

Abstract—We present a framework to ameliorate the classification of disaster-related social media messages. In the present work, we have incorporated the Convolutional Neural Network, and Long Short-Term Memory Network. To demonstrate the applicability and effectiveness of the proposed approach, it is applied to the thunderstorm and cyclone Fani dataset. The results indicate that CNN is better than the LSTM model with an accuracy score of 0.9999 (99.99%) and loss score of 0.0410. The output from the research study is helpful for disaster managers to make effective decisions on time.

Index Terms—Convolutional Neural Networks, Disaster Management, Long Short-Term Memory, Social Media.

I. INTRODUCTION

Social Media (SM) is one of the prominent tools that has been ranked the fourth most popular source of accessing emergency information [1]. During disastrous events, a lot of textual content is rapidly posted on microblogs, however only a few messages are useful and informative, the rest all are either sentimental or opinions [2]. Hence, considering the colossal and varying amount of data being broadcasted and shared through SM during the disaster, an effective text mining solution needs to be developed to distinguish the messages as *disaster-relevant* and *irrelevant*.

Convolution Neural Network (CNN) and Long Short-Term Memory Network (LSTM) network are the Deep Learning (DL) models in which the level of abstraction increases gradually by non-linear transformations of input data. Therefore, we address the challenge of assessing the relevancy of disaster-related SM messages by examining the capability of DL models of Twitter data in different disastrous events. Overall, the major contributions of this research are as follows:

- 1) We propose two datasets, i.e., a cyclone, and a thunderstorm dataset with manual label of relevance.
- 2) We present a novel framework for the classification of disaster data.
- 3) We compare two different neural networks in terms of performance and tune the models.
- 4) We computationally evaluate our model on different disaster-related datasets (thunderstorm and cyclone).

II. LITERATURE REVIEW

In the past few years, researchers have advocated the use of DL neural network model architectures over sundry domains. Zhao et al. [3] has utilized CNN to automatically extract and analyze the cyber threat intelligence from SM data. Madisetty and Desarkar [4] employ CNN based architecture to detect spam on Twitter. Salminen et al. [5] experiments with neural networks to develop an online hate classifier. Andrea et al. [6] presents an intelligent system regarding vaccine-related tweets utilizing CNN and LSTM models integrated with word embeddings. Wu et al. [7] identifies emotional labels from the psychological SM texts employing CNN-LSTM model. Ombabi et al. [8] experiments with CNN and LSTM models to analyse the sentiment from Arabic tweets. Subramani et al. [9] experiments with CNN and LSTM to identify domestic violence crisis from Facebook posts. They have been used in the medical domain. Monshi et al. [10] surveyed CNN and LSTM in generating radiological reports.

In this paper, we propose a DL based framework to address the aforementioned challenge. The proposed model will use DL techniques to classify disaster-related tweets based on the relevancy of the messages.

III. METHODOLOGY

The architectural diagram of the proposed framework is shown in Fig.1.

A. Data Acquisition

We used Twitter Search Application Programming Interface¹ to gather the tweets. We used the query #thunderstorm, #CycloneFani to fetch the tweets. Only English language tweets were collected. With the coded web-crawler in Python language, 1,48,287 thunderstorm tweets fetched from May 03, 2018, to May 10, 2018. The number of gathered cyclonic tweets are 1,72,907 dated April 27, 2019, to May 06, 2019.

B. Development of Gold Standard

We engaged three human annotators for the development of Gold Standard. The annotators were proficient in English

¹ <https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets.html>

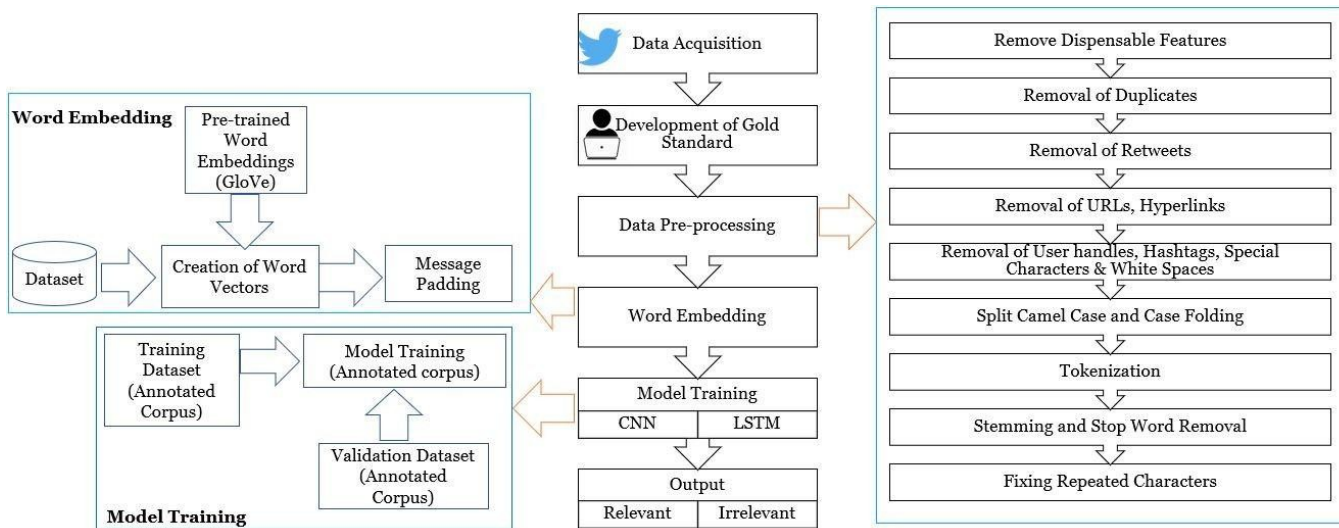


Fig. 1 Methodology of the Proposed Framework.

language and everyday users of Twitter, but none of them is an author of this paper. Given the fetched dataset to annotators, they were asked to label the tweets as relevant and irrelevant. Initially, all the three annotators divided each dataset into three equal sections and annotated the tweets individually. They annotated the tweets by inspecting carefully the terms relevant to thunderstorm and cyclone Fani. Therefore, *relevant* and *irrelevant* tweets are categorized as following:

- *Relevant tweets:* Tweets those contain information about the situation, need, non-need and availability of a vital resource like food, water, shelter, medical aid etc.
- *Irrelevant tweets:* Tweets which do not contain any information about the disaster.

C. Data Preprocessing

One of the main challenges of using SM data for Disaster Management (DM) is the unstructured and raw data. To transform the raw data into information, tweets undergo a series of pre-processing steps. The steps range from removal of irrelevant features to normalization and transforming the data into vectors. In our approach, the pandas dataframe has been used to execute the following steps of pre-processing: removal of dispensable features, removal of duplicates, removal of retweets, URLs, hyperlinks, user handles, hashtags, special characters, and white spaces, split camel case and case folding, tokenization, stemming and stop word removal, fixing repeated characters.

D. Model Details

In this study, we employ two DL models namely CNN, and LSTM. For the purpose of building the models, each dataset is partitioned into two subsets- training and validation dataset in the ratio of 80:20, respectively. During the training phase, the training dataset is fed into the model to generate output from input efficiently. The purpose of the validation dataset is to provide an unswayed evaluation of the model to fit on training data whilst tuning the hyperparameters.

IV. RESULTS

In order to experiment with different neural network models, we ran several training simulations with random combinations of hyperparameters, i.e., random grid search, to evaluate the best converging model towards the best accuracy score. Table I demonstrates the results from the validation phase. Specifically, it lists out the accuracy and loss score obtained during each simulation. With the optimized standards in mind, we chose the hyperparameters that yielded the best accuracy and lesser loss score. The simulations are run for 50 epochs. The best scores are highlighted in bold. Considering

Cyclone Fani dataset, LSTM performs better than CNN with higher accuracy score and lesser loss score. Further, considering thunderstorm dataset, CNN model outperforms LSTM model. Considering both the model's performance, the optimized CNN model yields highest accuracy and lowest loss score on thunderstorms between the two datasets (Table II, row 1).

V. CONCLUSION

Reliable and accurate framework for disaster in SM usage is of great significance but has gained scant attention as compared to other fields. However, in recent years, significant advancements have been made for developing efficient frameworks. In the paper, we described CNN and LSTM models and applied the classifier to manually annotated datasets, which are classified into relevant and irrelevant messages. We adopted well-known state-of-the-art CNN and LSTM models for the disaster domain. For both the models, we demonstrated the results with thunderstorm as well as cyclone datasets. We investigated the effect of the model architecture and hyperparameters' tuning on the model's performance to determine the best architecture. Our results demonstrate that CNN model is better than LSTM model in terms of accuracy and loss. Thus, CNN model outperforms LSTM model in assessing the relevancy of disaster-based messages with an accuracy score of 0.9999 (99.99%) and 0.0410 loss score. Hence, the research adds to the growing body of disaster research by proposing a DL based frame-

TABLE I Accuracy and Loss score on each of the datasets.

Dataset	Model	Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss	
Cyclone	CNN	5	0.9825	0.2102	0.9827	0.2010	
	CNN	10	0.9827	0.1911	0.9828	0.1909	
	CNN	25	0.9825	0.1010	0.9827	0.1008	
	CNN	50	0.9825	0.1009	0.9827	0.1008	
	LSTM	5	0.9827	1.3936	0.9828	1.3524	
	LSTM	10	0.9769	1.1610	0.9813	1.1760	
	LSTM	25	0.9790	0.3294	0.9858	0.3110	
	LSTM	50	0.9904	0.1007	0.9993	0.1006	
	Thunderstorm	CNN	5	0.9945	0.0898	0.9956	0.0691
		CNN	10	0.9955	0.0535	0.9956	0.0432
CNN		25	0.9962	0.0432	0.9978	0.0424	
CNN		50	0.9988	0.0414	0.9999	0.0410	
LSTM		5	0.8992	41.4000	0.9241	11.7506	
LSTM		10	0.9315	20.4506	0.9310	17.8893	
LSTM		25	0.9969	2.4062	0.8979	2.3794	
LSTM		50	0.8854	1.2971	0.9386	1.4972	

TABLE II Results with the optimal hyperparameter combinations for CNN and LSTM models. The bold numbers correspond to highest accuracy score and lowest loss score.

Model	Dataset	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
CNN	Thunderstorm	0.9988	0.0414	0.9999	0.0410
LSTM	Cyclone	0.9904	0.1007	0.9993	0.1006

work that efficiently assesses the disaster-based messages as *disaster-relevant* or *irrelevant*.

VI. LIMITATIONS AND FUTURE SCOPE

The results exhibit the potential of the proposed approach. There are a number of limitations and research directions to be considered for future work. First, the work is limited to static data. We envision the real-time relevancy prediction of the SM messages. Second, it is limited to natural disasters. We ideate a multi-hazard design for efficient decision-making.

REFERENCES

- [1] J. Kim and M. Hastak, "Social network analysis: Characteristics of online social networks after a disaster," *Int. J. of Info. Mgt.*, vol. 38, no. 1, pp. 86–96, 2018. [Online]. Available: <https://doi.org/10.1016/j.ijinfomgt.2017.08.003>
- [2] M. Basu, A. Shandilya, P. Khosla, K. Ghosh, and S. Ghosh, "Extracting Resource Needs and Availabilities From Microblogs for Aiding Post- Disaster Relief Operations," *IEEE Trans. on Comp. Soc. Sys.*, vol. 6, no. 3, pp. 1–15, 2019.
- [3] J. Zhao, Q. Yan, J. Li, M. Shao, Z. He, and B. Li, "TIMiner: Automatically extracting and analyzing categorized cyber threat intelligence from social data," *Comp. & Sec.*, vol. 95, 2020.
- [4] S. Madisetty and M. S. Desarkar, "A Neural Network-Based Ensemble Approach for Spam Detection in Twitter," *IEEE Trans. on Comp. Soc. Sys.*, vol. 5, no. 4, pp. 973–984, 2018.
- [5] J. Salminen, M. Hopf, S. A. Chowdhury, S. Jung, H. Almerakhi, and B. J. Jansen, "Developing an online hate classifier for multiple social media platforms," *Hum.-cent. Comp. and Info. Sci.*, vol. 6, pp. 1–34, 2020. [Online]. Available: <https://doi.org/10.1186/s13673-019-0205-6>
- [6] E. D. Andrea, P. Ducange, A. Bechini, A. Renda, and F. Marcelloni, "Monitoring the public opinion about the vaccination topic from tweets analysis," *Exp. Sys. With App.*, vol. 116, pp. 209–226, 2019. [Online]. Available: <https://doi.org/10.1016/j.eswa.2018.09.009>
- [7] Z. Wu, D. Pi, J. Chen, M. Xie, and J. Cao, "Rumor detection based on propagation graph neural network with attention mechanism," *Exp. Sys. with App.*, vol. 158, pp. 1–16, 2020.
- [8] A. H. Ombabi, W. Ouarda, and A. M. Alimi, "Deep learning CNN – LSTM framework for Arabic sentiment analysis using textual information shared in social networks," *Soc. Net. Anal. and Min.*, pp. 1–13, 2020. [Online]. Available: <https://doi.org/10.1007/s13278-020-00668-1>
- [9] S. Subramani, H. U. A. Wang, H. U. Y. Q. Vu, G. Li, and S. Member, "Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning," *IEEE Acc.*, vol. 6, pp. 54 075–54 085, 2018.
- [10] M. M. A. Monshi, J. Poon, and V. Chung, "Deep learning in generating radiology reports: A survey," *Art.Int. In Med.*, vol. 106, no. April 2019, p. 101878, 2020. [Online]. Available: <https://doi.org/10.1016/j.artmed.2020.101878>