Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 165 (2019) 511–516

Procedia Computer Science

www.elsevier.com/locate/procedia

INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019, ICRTAC 2019

Embedded Bi-directional GRU and LSTMLearning Models to Predict Disasterson Twitter Data

A.Bhuvaneswari[a], J.Timothy Jones Thomas[b], P.Kesavan[c]

[a]Vellore Institute of Technology, Chennai, Tamil Nadu, India
[b]Uber Technologies Inc, Hyderabad, Telangana, India
[c]Sahaj Software Solutions Ltd, Chennai, Tamil Nadu, India

Abstract

The deep learning techniques namely Long Short Term Memory (LSTM) network and Bi-directional Gated Recurrent Unit (BGRU) network turn to be de facto to build an optimal assembly line for neural network models. The prevailing state-of-the-art approaches require a substantial amount of labeled data detailed to an unambiguous event in the training phase. In this paper, embedded bi-directional GRU and LSTM learning models is applied for disaster event prediction that uses deep learning techniques to categorize the tweets. The performance of the proposed neural network model is evaluated on CrisisLexT26 benchmarking dataset. The resulting validation accuracy is estimated by comparing LSTM and bi-directional GRU with and without word embeddings. The experiments demonstrate the model selector choose the deep learning techniques to predict the disaster event with reasonably high accuracy.

*Keywords:*Event Prediction, Online Social Networks, Twitter, Deep Learning; Classification; Long Short-Term Memory, Gated Recurrent Unit

1. Introduction

Twitter is a popular microblogging site which enables online users to post the day-today updates in the form of tweets and keep informed of their followers [1]. The evolution of Web 2.0 has turn out to be the dynamic source of information to examine online activities of web users owing to its openness, online and easiness of accessibility. The growth of Internet is characterized to dynamic user-generated content due to the evolution of Online Social Networks (OSN) namely Twitter, Facebook, Tumblr, MySpace and Sina Weibo [2]. Events-of-interest discovery is accomplished using Twitter data source by examining the intensities of information diffusion after the mass emergency events. Usually, event trends on Twitter are instigated by an external happening such as a natural disaster and in some cases they are particular to a tweet triggered by a renownedcelebrity. Moreover, the OSN data is composed of tiny, noisy, semi-structured and unstructured content,challengingconsiderablydissimilar techniques to resolve machine learning techniques involved in information retrieval. The various real-time events can be detected in OSN using discerning text and multimedia based content posted by users [3].In Twitter, users post referred as 'tweets'

Corresponding Author Tel: +91-9894007258
Email: bhuvana.cse14@gmail.com

1877-0509 © 2019 The Authors. Published by Elsevier B.V.

10.1016/j.procs.2020.01.020

which is considered to be a vital feature shared by the followers. The enhanced length feature endowed more towards value-added area of research for Twitter stream analysis. The retrospective way of detecting trending event prediction and classifying the tweets in Twitter is deliberated to be animmense challenge. The aim is to provide a story news boards which resembles a quick snapshot of a disaster using twitter feeds. The research challenge is addressed by building a deep learning enabled disaster event prediction model using Twitter dataset that classify events using deep learning models namely LSTM and bi-directional GRU. In this paper, CrisisLex dataset [4] is used that contains a collection of tweets, which are specifically related to disasters. The dataset is used to identify tweets that contain valuable information regarding a disaster. During the experimentation, text generation and classification is applied on the CrisisLexT26 dataset, implementing various deep learning models and its performance is compared.

2. Related Work

Event detection involves several clustering and classification algorithms consisting of modules namely event-prediction, event-related query identification, event assignment, and event archive. An event occurs in online social streams which cause a large number of activities especially during disasters [5] use machine learning algorithms to classify data. It is based on the reflection of real-world happening at particulartemporal and spatial frame. A framework named Convolutional Bi-Directional LSTM used Bi-directional LSTMs and CNN for filtering unfitting conversations on real-world conversations and search queries [6].The machine learning techniques lead to an incremental and hierarchical modelling for classifying and building event theme structures at several granularities. The adaptation of Convolutional Neural Network (CNN), for training data Learning without Forgetting method [7] to train the network. Obviously, it preserves the original proficienciesto accomplish similarly to multitask learning that routines original task data.

The research work determined the disaster topics using which uses Shannon entropy during disasters in Cyber-Social Networks [8]. Another work, in which the Twitter user tweets as sensory information used Gradient Descent Boost, SVM and kNN to detect the specific event types namely earthquakes, tornadoes, and traffic jams [9]. On the other hand, a regression machine learning model[10], is addressed using Gradient Boosted Decision Trees and SVM for event detection. A news handling systemcalled TwitterStand[11], predominantlydecides to monitor Twitter microblog service to unavoidablyacquire breaking news from the tweets posted by the verified Twitter users. The unsupervised term selection was based on different ranking model with the cut-off being adaptively calculated. A resultant controllable topic graph is produced which associate the mined evolving terms with concurrent terms to attain a set of emergent topics.

Location based Twitter Tweets November 2013 [12], used SVM classification for event detection. The data was extracted using random filters from Twitter data extraction rule for substantial information mining. The classifier uses a supervised term selection method which was purely based on a user-centric threshold parameter. In a successive work [13], a feature constrained multi-task learning models is assembledusing CNN for spatio-temporal event forecasting. The long short-term memory (LSTM) deep learning model [14],decreases the service latency where mobile user demands the election information over wireless networks. Based on the literature survey, a novel framework is proposed to resolve using renowned deep learning algorithms.

3. Learning Approaches and Models

3.1. Long Short-Term Memory

Long Short-Term Memory (LSTM) units are units of anartificial recurrent neural network (RNN) composed of LSTM units are often called an LSTM network. LSTM is used in the experiments to generate text and classify the tweets. In common, the LSTM methodpicks the most likely word each time, be dependent on which class the tweet is to be classified. The LSTM is applied on the dataset beside with count vectorization and Word embedding using WordNet in which words are mapped to vectors. It implicates a mathematical embedding from a space with one dimension per word to anuninterrupted vector space with a much lower dimension. As a result, the word embedding generates a mapping which is later passed in to neural network for explicit representation of the context. The LSTM unit is a collection of cell, input gate, output gate and forget gate [18]as shown in Fig.1. The three gates regulate the flow of information into and out of the cell and the LSTM unit cell remember the values above arbitrary time intervals.

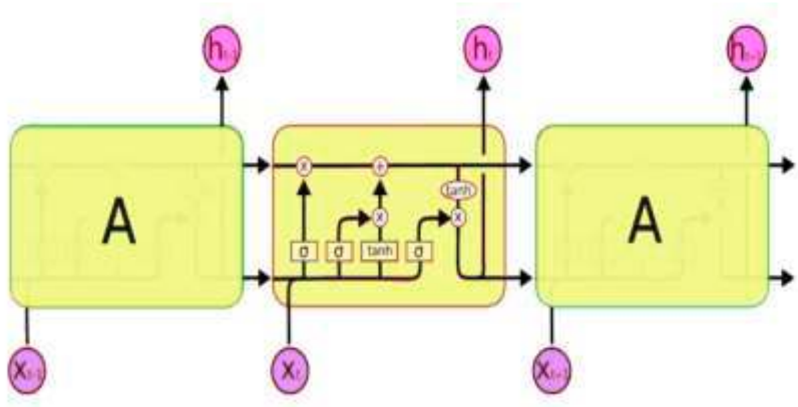

Fig.1. Long Short-Term Memory (LSTM) Model

The following equations govern our training.

$$Z_t = \sigma(W_Z[h_{t-1}, x_t]) \tag{1}$$

$$r_t = \sigma(W_r[h_{t-1}, x_t]) \tag{2}$$

$$\tilde{h}_t = \tanh(W[r_t * h_{t-1}, x_t]) \tag{3}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{4}$$

where x_{t-1}, r_{t-1}, z_{t-1}, $\tilde{h}_{t-1}$represents the input, reset gate, update gate, reset memory respectively. The following is a real tweet generated by our model and its process is given below.

$$< begin > \rightarrow h_1 = (k_1, k_2, \dots, k_{|V|}) \rightarrow Disaster \rightarrow h_2 = (k_1, k_2, \dots, k_{|V|}) \rightarrow flood \rightarrow h_3 = (k_1, k_2, \dots, k_{|V|}) \rightarrow collapsed \rightarrow h_4 = (k_1, k_2, \dots, k_{|V|}) \rightarrow < end >$$

3.2. Bi-directional Gated Recurrent Unit

Gated recurrent units (GRUs) are considered to be a gating mechanism in artificial recurrent neural networks, which was found to be similar to that of LSTM [18] as shown in Fig.2. However, GRUs has been shown to demonstrate better performance on smaller to medium quantity datasets.

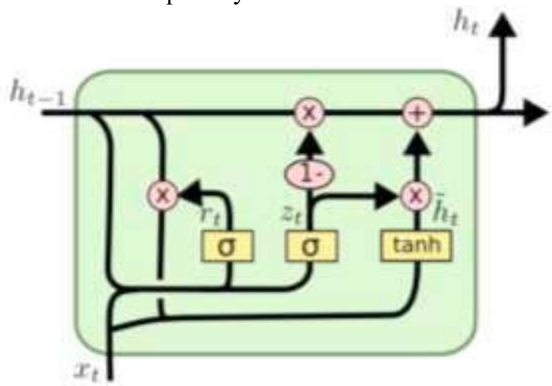

Fig.2. Gated Recurrent Unit (GRU) Model.

The word embedding is calculated using WordNet where bi-directional GRU is applied to concatenate the words of a tweet into single word vector. The placeholders are added essential to guarantee that all tweets have a conjoint length. The overall network (Right direction and Left direction) and its equations for GRU network is as follows.

Right direction: $\overrightarrow{h_t}^{(i)}$

$$\overrightarrow{z_t}^{(i)} = \partial(\overrightarrow{W}_{(i)}^{(z)} x_t^i + \vec{U}_{(i)}^{(r)} h_{t-1}^{(i)}) \tag{5}$$

$$\overrightarrow{r_t}^{(i)} = \partial(\overrightarrow{W}_{(i)}^{(r)} x_t^i + \vec{U}_{(i)}^{(r)} h_{t-1}^{(i)}) \tag{6}$$

$$\overrightarrow{\widetilde{h_t}}^{(i)} = \tanh(\overrightarrow{W}_{(i)} x_t \ + r_t \circ \vec{U}_{(i)} h_{t-1}) \tag{7}$$

$$\overrightarrow{h_t}^{(i)} = z_t{}^{(i)} \circ h_{t-1}^{(i)} + (1 - z_t{}^{(i)}) \circ \widetilde{h_t}^{(i)} \tag{8}$$

Left direction: $\overleftarrow{h_t}^{(i)}$

$$\overleftarrow{z_t}^{(i)} = \partial(\overleftarrow{W}_{(i)}^{(z)} x_t^i + \overleftarrow{U}_{(i)}^{(r)} h_{t-1}^{(i)}) \tag{9}$$

$$\overleftarrow{r_t}^{(i)} = \partial(\overleftarrow{W}_{(i)}^{(r)} x_t^i + \overleftarrow{U}_{(i)}^{(r)} h_{t-1}^{(i)}) \tag{10}$$

$$\overleftarrow{\widetilde{h_t}}^{(i)} = \tanh(\overleftarrow{W}_{(i)} x_t \ + r_t \circ \overleftarrow{U}_{(i)} h_{t-1}) \tag{11}$$

$$\overleftarrow{h_t}^{(i)} = z_t{}^{(i)} \circ h_{t-1}^{(i)} + (1 - z_t{}^{(i)}) \circ \widetilde{h_t}^{(i)} \tag{12}$$

Output:

$$y_t = softmax(U \left[\overrightarrow{h_t}^{(top)}, \overleftarrow{h_t}^{(top)} \right] + a) \tag{13}$$

wherez_t representupdate gate,r_t is reset gate,$\widetilde{h_t}$ is the reset memory,and h_t isnew memory. During the experiments using LSTM and Bi-GRU, the batch of tweets (Batch Size= 1000 numbers) in each x_t which influence each other's classification. It ispresumed that each tweet is independent of each other as counterintuitive.

4. Experiments and Results

CrisisLexT26 dataset [17] is used for the experiments and considered as a benchmarking dataset, which, as its name suggests, has 26 disasters. The proposed optimal prediction neural network model selector is initially trained on the CrisisLexT26 Dataset, with the tweet text to which disaster the tweet belongs as the attributes. The dataset containing 250K tweetsposted during 26 crisis events in 2012 and 2013, with most events having 2K-4K tweets. It has more features like information source, informativeness and information type with 250K tweets. For classification, the tweet text and the in formativeness features are used to classify if a tweet is talking about a disaster or not. The system is trained using 80% of training tweets data and tested with 20% of testing tweets data. For Topic Modelling, only the tweet text feature is enough to identify the topics.

Table 1.CrisisLexT26Dataset – Training data and Testing Data

CrisisLexT26 Dataset	**Number of Tweets**
Total Tweets	250K tweets
Valid Sampling Tweets	100 k tweets
Train Data	80 k tweets
Test Data	20 k tweets

Tensorflow[15] is used for the implementation with WordNet package using Python. As part of the pre-processing step, the dataset is also converted into a suitable form to be given to the deep learning models. The data preparation includes Stop Words removal, Punctuation removal, Stemming, Lemmatization, and Bag of words construction, count vectorization and TF-IDF Vectorization. Count Vectorization and TF-IDF vectorization are applied to the dataset after it is split into bigrams, i.e., n-grams with n=2. It yields better accuracy when considering unigrams or n-grams with n greater than 2.The unbalanced raw dataset is converted into a balanced dataset by adding a small amount of Gaussian noise to each disaster event sample. The dataset sample is exclusively pre-trained with word embedding like WordNet [16].In Recurrent Neural Networks, all the sentences in the dataset are split into words and converted using the embedding layer into word embedding. Next, the word embeddingis applied along with neural network layers - Long Short-Term Memory, Bi – directional Gated Recurrent Unit with 3-layers which isfully connected. These layers finally connect to the output layer. The accuracy of the neural networks that are trained using those layers is shown in the Fig. 3, for 5 iterations. Specifically, the GRU and LSTM made the decision globally, where the distribution of its prediction of testing set persistently counterparts the training set.Furthermore, the input dataset is randomized in such a way that tweets are batched into the neural network. It represent different disaster events and are directed towards 26 Disaster events namely 2012 Colorado wildfires, 2012 Costa Rica earthquake and so on. The randomized tweets lead to use a network with LSTM and bi-directional path GRU. The neural networks applied on the tweets give almost the same accuracy after 5 iterations, although bi-directional GRU with 3-layers produces higher accuracy when compared to LSTM.

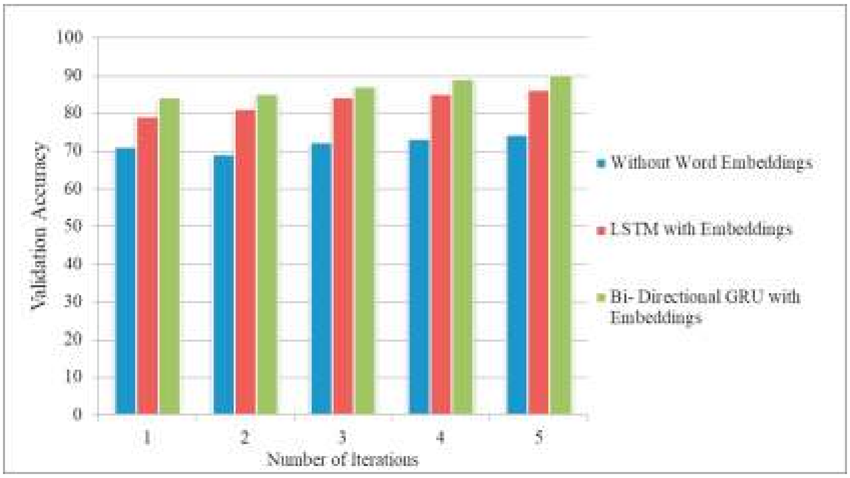

Fig.3.Validation Accuracy of disasterlearningmodels.

The experimentsare targetedtodetermine in empirical Bayes method, where global information is supportive to classify locally, even though each tweet sample is presumed to be independent of one another. It is observed from the experimental results, that there is an acute increase in validation accuracy by increasing the number of iterations for LSTM and Bi-directional GRU networks, using Word Embedding, classifier ensembles and lexicons. The prior information is deliberated, so that majority of tweet dataset belongs to any one the disaster events and routines the applied knowledge to increase its predication towards a certain disaster event class. The performance evaluation metrics are shown in Table 2for all the applied deep learning models.The results in Table 2 show that the LSTM and GRU models give high accuracy with word embedding comparatively without word embedding.

Table 2.Accuracy, Recall, F1-Score for Classifiers

Model	Overall Accuracy	Recall	F1
Without Word Embedding	0.72	0.66	0.69
LSTM + With Embedding	0.85	0.81	0.79
Bi-directional GRU + With Embedding	0.89	0.86	0.82

5. Conclusion

The deep learning models namely bi-directional GRU, LSTM and single layer network are applied on the dataset in which GRU gives the highest accuracy for classification of the CrisisLexT26 to identify disaster events. The proposed optimal prediction model selector opted LSTM for lengthy twitter feeds and GRU for short tweets respectively. Once the training set is balanced, the results are much better and more sophisticated models lead to better results despite violating a key assumption in statistics of training and testing on the same distribution. However, a better accuracy can probably be obtained by trying out different neural network configurations, with pre-trained word by tuning the hyper parameters in the other classifiers.The event classification will provide story board visualization which resembles a quick snapshot of a disaster using streaming twitter feeds. The deep learning approaches are applied for datasets collected from Twitter. As a part of future work, the system can be trained and tested for dynamic datasets which can be handled by Spark Streaming ML frameworks.

References

[1] McFedries, Paul.(2010) “Twitter tips, tricks, and tweets.” *John Wiley & Sons.*

[2] Jang, Beakcheol, and Jungwon Yoon. (2018) "Characteristics analysis of data from news and social network services." *IEEE Access***6 :** 18061-18073.

[3] Aiello, Luca Maria, Georgios Petkos, Carlos Martin, David Corney, Symeon Papadopoulos, Ryan Skraba, AyseGöker, IoannisKompatsiaris, and Alejandro Jaimes. (2013) "Sensing trending topics in Twitter." *IEEE Transactions on Multimedia***15(6):** 1268-1282.

[4] Olteanu, Alexandra, Sarah Vieweg, and Carlos Castillo. (2015) "What to expect when the unexpected happens: Social media communications across crises." *In Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, 994-1009. ACM. Vancouver, Canada.

[5] Anbalagan, Bhuvaneswari, and ChinnaiyahValliyammai. (2016) "# ChennaiFloods: Leveraging Human and Machine Learning for Crisis Mapping during Disasters Using Social Media." *In IEEE 23rd International Conference on High Performance Computing Workshops (HiPCW),* 50-59. IEEE.

[6] Tsai, Kuo Chun, Li Liliane Wang, and Zhu Han. (2018) "Caching for mobile social networks with deep learning: Twitter analysis for 2016 us election." *IEEE Transactions on Network Science and Engineering.* **20(7):**1-1.

[7] Li, Zhizhong, and Derek Hoiem. (2017) "Learning without forgetting." *IEEE transactions on pattern analysis and machine intelligence* **40(12):** 2935-2947.

[8] Anbalagan, Bhuvaneswari, and ChinnaiyahValliyammai. (2019) “Information entropy based disaster event detection framework in Online Social Networks.” *Journal of Intelligent Fuzzy System IOS Press***36:** 3981-3992.

[9] Sakaki, Takeshi, Makoto Okazaki, and Yutaka Matsuo. (2010) "Earthquake shakes Twitter users: real-time event detection by social sensors." *In Proceedings of the 19th international conference on World Wide Web*, 851-860. Raleigh, USA. ACM.

[10] Popescu, Ana-Maria, and Marco Pennacchiotti. (2010) "Detecting controversial events from twitter." *In Proceedings of the 19th ACM international conference on Information and knowledge management*, 1873-1876. ACM,

[11] Sankaranarayanan, Jagan, HananSamet, Benjamin E. Teitler, Michael D. Lieberman, and Jon Sperling. (2009) "Twitterstand: news in tweets." *In Proceedings of the 17th ACM Sigspatial International Conference On Advances In Geographic Information Systems*, 42-51. ACM.

[12] Yamada, Wataru, Daisuke Torii, Haruka Kikuchi, Hiroshi Inamura, Keiichi Ochiai, and Ken Ohta. (2015) "Extracting local event information from micro-blogs for trip planning." *In 2015 Eighth International Conference on Mobile Computing and Ubiquitous* Networking (ICMU): 7-12. IEEE.

[13] Zhao, Liang, Qian Sun, Jieping Ye, Feng Chen, Chang-Tien Lu, and Naren Ramakrishnan. (2017) "Feature constrained multi-task learning models for spatiotemporal event forecasting." *IEEE Transactions on Knowledge and Data Engineering***29(5):** 1059-1072.

[14] Lee, Wen-Yu, Winston H. Hsu, and Shin’ichi Satoh. (2018) "Learning from cross-domain media streams for event-of-interest discovery." *IEEE Transactions on Multimedia***20(1):** 142-154.

[15] Shukla, Nishant.(2018)“Machine learning with TensorFlow.” *Manning Publication.*

[16] Smedt, Tom De, and Walter Daelemans.(2012) "Pattern for python." *Journal of Machine Learning Research***13:** 2063-2067.

[17] CrisisLexT26 Dataset : Online - https://crisislex.org/data-collections.html#CrisisLexT26

[18] Brownlee, Jason. (2017)“Long Short-term Memory Networks with Python: Develop Sequence Prediction Models with Deep Learning.” Machine Learning Mastery, 2017.