

# NLPRL at WNUT-2020 Task 2: ELMo-based System for Identification of COVID-19 Tweets

Rajesh Kumar Mundotiya, Rupjyoti Baruah, Bhavana Srivastava, Anil Kumar Singh

Department of Computer Science and Engineering

Indian Institute of Technology (BHU), India

Varanasi

{rajeshkm.rs.cse16, rupjyotibaruah.rs.cse18}@iitbhu.ac.in

{bhavanasrivastava.rs.cse17, aksingh.cse}@iitbhu.ac.in

## Abstract

The Coronavirus pandemic has been a dominating news on social media for the last many months. Efforts are being made to reduce its spread and reduce the casualties as well as new infections. For this purpose, the information about the infected people and their related symptoms, as available on social media, such as Twitter, can help in prevention and taking precautions. This is an example of using noisy text processing for disaster management. This paper discusses the NLPRL results in Shared Task-2 of WNUT-2020 workshop. We have considered this problem as a binary classification problem and have used a pre-trained ELMo embedding with GRU units. This approach helps classify the tweets with accuracy as 80.85% and 78.54% as F1-score on the provided test dataset. The experimental code is available online<sup>1</sup>.

## 1 Introduction

The Coronavirus disease (officially named COVID-19 by the World Health Organization or WHO on February 11, 2020) is still spreading worldwide, although the numbers in some countries have decreased. In mid-June 2020, there was fear and panic all around the world related to the COVID-19 outbreak. Twitter has been used for gathering information about crisis communications (Cho et al., 2013). There have been a massive number of tweets about the pandemic. One estimate puts the number at about four million COVID-19 English tweets daily (Lamsal, 2020).

For our task, the tweets can be classified in two categories: informative and uninformative. Informative tweets contain some useful information about the pandemic and the affected people and

they can help in managing the spread of the pandemic. The rest can be treated as uninformative tweets. The informative tweets provide information about recovered, suspected, confirmed, and deceased cases, and possibly also about a person's location or travel history. The majority of tweets are uninformative. This kind of classification is an example of text classification, which is a core task in many areas of Natural Language Processing (NLP).

## 2 Related work

Continuous availability of vital number of Twitter posts has been useful for developing more accurate and reliable classification methods for noisy text. Conventional NLP techniques and machine learning-based classification methods do not seem to perform well with Twitter data. Jiang et al. (2018) described a work on identifying health-related Personal Experience Tweets (PET) by combining word embedding and an LSTM neural network that demonstrated significant improvement (with  $p < 0.01$ ) in performance measures of accuracy, precision, recall, F1-score, and ROC/AUC over the conventional methods in identifying PETs.

In the past few months, many researchers have tried out several mathematical and statistical models to predict novel Coronavirus transmission (Zhao et al., 2020; Shim et al., 2020; Benvenuto et al., 2020). Since the COVID-19 dataset is of a time series nature, it is natural to use sequential networks to extract the patterns from it. Some studies have used LSTM networks to forecast the spread of infectious diseases such as the current COVID-19 epidemic (Chimmula and Zhang (2020); Bandyopadhyay and Dutta (2020); Huang et al. (2020); Tomar and Gupta (2020); Pal et al. (2020)).

Dubey (2020), who analyzed the country-wise sentiment analysis of tweets and emotions of the

<sup>1</sup>[https://github.com/Rajesh-NLP/Identification\\_of\\_informative\\_COVID-19\\_English\\_Tweets-WNUT-20](https://github.com/Rajesh-NLP/Identification_of_informative_COVID-19_English_Tweets-WNUT-20)

people from 12 countries (11th March 2020 to 31st March 2020) revealed that countries like Australia, Belgium, and India were tweeting about COVID-19 with a positive sentiment and people in China had negative sentiments about the same.

Arora et al. (2020) proposed deep learning models to predict the number of COVID-19 positive cases in 32 states of India and its Union Territories. Using RNN based LSTM cells and its variants such as deep LSTM, convolutional LSTM, and bi-directional LSTM as predictive models led to the conclusion that at present Bi-directional LSTM gives the best results, and convolutional LSTM gives the lowest results based on prediction errors.

Contextual embeddings are known to provide better results on the basic NLP problems such as Sentiment analysis (Wang et al., 2020; Müller et al., 2020; Kruspe et al., 2020; Akbik et al., 2018). In this paper, we made an effort to automatically identify whether a COVID-19 English tweet is informative or not, using a model based on a contextual embedding called ELMo (Peters et al., 2018). The dataset of 10K COVID-19 English tweets was provided by the organizers of the shared task.

### 3 Problem Statement

The goal of WNUT-2020 Task 2, named as the “Identification of informative COVID-19 English tweets”, is to classify COVID-19 tweets into informative and uninformative categories. This is a binary classification task to learn  $F : X \rightarrow Y$  where  $X = \{X_1, X_2, X_3 \dots, X_m\}$  is a tweet of length  $m$  and  $Y \in \{informative, uninformative\}$ .

### 4 System Description

Word embeddings or distributed representations of words use dense, real-valued vectors to represent vocabulary words. Word embeddings have a much smaller dimension than the size of the vocabulary and carry syntactic and semantic information about the words, unlike one-hot vectors. In this paper, we use ELMo (Embeddings from Language Models) method proposed by Peters et al. (2018) for word embeddings. The dimensionality of each word vector is 2048.

The Bi-GRU model is a variation of RNN (Hermands and Schrauwen, 2013) that simultaneously models the word representation with its preceding and following information. The output of the word embedding is fed to the GRU (Cho et al., 2014) unit. This GRU unit is then passed to a linear layer for

predicting the classes.

## 5 System Training

The provided dataset (Nguyen et al., 2020) contains 10000 tweets, out of which 4719 tweets are informative, whereas the rest are labeled as uninformative. The dataset is divided into three sub-parts, each used to train, tune and test the model, respectively. The statistics for these sub-parts of the dataset, according to the class, are mentioned in Table 1.

Dataset	Informative	Uninformative
Train	3303	3697
Development	472	528
Test	944	1056

Table 1: Number of tweets in the dataset partitions used for the experiment

We have used pre-trained ELMo embedding with the size of 2048, with 512 GRU hidden units, and 0.5 as the dropout to perform our experiments. The model was trained by using the flair library (Akbik et al., 2019) at a batch size of 32 for 30 epochs, and with a learning rate of 0.1. To prevent overfitting of the model training, we have used the early-stoppage and the patience value of 3.

## 6 Results

ELMo and contextual string embeddings (Akbik et al., 2018) (released in the Flair framework<sup>2</sup>) are prominent embedding techniques that consider contextual information while generating the word embedding. We tried both an ELMo-based model and a Flair-based model. The embedding used for the Flair-based model was the Flair embedding (not the Pooled Flair embedding). The pre-trained models of these techniques are trained on the news domain for English. In our experiments on the development data, ELMo gave better results. Hence, for the submitted system for test dataset, we have used the ELMo-based model. Using the provided dataset, we obtained the 78.54% as weighted F1-score on the test dataset, whereas 84.79% by ELMo-based model and 83.60% by Flair-based model on the development dataset. The evaluation of the test and development dataset with the usual metrics are outlined in Table 2.

<sup>2</sup>[https://github.com/flairNLP/flair/blob/master/resources/docs/embeddings/FLAIR\\_EMBEDDINGS.md](https://github.com/flairNLP/flair/blob/master/resources/docs/embeddings/FLAIR_EMBEDDINGS.md)

Dataset	P	R	F1	Accuracy
Test <sub>ELMo</sub>	83.35	74.26	78.54	80.85
Dev <sub>ELMo</sub>	84.80	84.80	84.79	84.80
Dev <sub>Flair</sub>	84.11	83.10	83.60	84.10

Table 2: Classification results obtained from various evaluation metrics (Precision, Recall, weighted F1-score and Accuracy) on the Test and Development (Dev) dataset

The *informative* class F1-score obtained on the development dataset (83.69% with ELMo-based model and 83.96% with Flair-based model) are given in Table 3.

Model	Precision	Recall	F1-score
ELMo	84.78	82.63	83.69
Flair	85.34	82.63	83.96

Table 3: Classification results on the *informative* class obtained from various evaluation metrics on the development dataset for the two methods

## 7 Conclusion

This system description paper reports a simple method that leverages the ELMo embedding features to train a COVID-19 informative tweet identification system. We obtained 78.54% F1-score on the test dataset with ELMo-based approach. For future work, we will use this model with adverse regularization, in order to make it more robust. Further, the self-learning algorithm can be applied an available large monolingual corpus gathered from Twitter.

## Acknowledgments

The support and the resources provided by Param Shivay facility under the National Supercomputing Mission, Government of India at the Indian Institute of Technology, Varanasi are gratefully acknowledged.

## References

Alan Akbik, Tanja Bergmann, Duncan Blythe, Kashif Rasul, Stefan Schweter, and Roland Vollgraf. 2019. *FLAIR: An easy-to-use framework for state-of-the-art NLP*. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pages 54–59, Minneapolis, Minnesota. Association for Computational Linguistics.

Alan Akbik, Duncan Blythe, and Roland Vollgraf. 2018. *Contextual string embeddings for sequence labeling*. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1638–1649, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Parul Arora, Himanshu Kumar, and Bijaya Ketan Panigrahi. 2020. Prediction and analysis of covid-19 positive cases using deep learning models: A descriptive case study of india. *Chaos, Solitons & Fractals*, 139:110017.

Samir Kumar Bandyopadhyay and Shawni Dutta. 2020. Machine learning approach for confirmation of covid-19 cases: Positive, negative, death and release. *medRxiv*.

Domenico Benvenuto, Marta Giovanetti, Lazzaro Vassallo, Silvia Angeletti, and Massimo Ciccozzi. 2020. Application of the arima model on the covid-2019 epidemic dataset. *Data in brief*, page 105340.

Vinay Kumar Reddy Chimmula and Lei Zhang. 2020. Time series forecasting of covid-19 transmission in canada using lstm networks. *Chaos, Solitons & Fractals*, page 109864.

Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*.

Seong Eun Cho, Kyujin Jung, and Han Woo Park. 2013. Social media use during japan’s 2011 earthquake: how twitter transforms the locus of crisis communication. *Media International Australia*, 149(1):28–40.

Akash Dutt Dubey. 2020. Twitter sentiment analysis during covid19 outbreak. *Available at SSRN 3572023*.

Michiel Hermans and Benjamin Schrauwen. 2013. *Training and analysing deep recurrent neural networks*. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 190–198. Curran Associates, Inc.

Chiou-Jye Huang, Yung-Hsiang Chen, Yuxuan Ma, and Ping-Huan Kuo. 2020. Multiple-input deep convolutional neural network model for covid-19 forecasting in china. *medRxiv*.

Keyuan Jiang, Shichao Feng, Qunhao Song, Ricardo A Calix, Matrika Gupta, and Gordon R Bernard. 2018. Identifying tweets of personal health experience through word embedding and lstm neural network. *BMC bioinformatics*, 19(8):210.

Anna Kruspe, Matthias Häberle, Iona Kuhn, and Xiao Xiang Zhu. 2020. Cross-language sentiment analysis of european twitter messages during the covid-19 pandemic. *arXiv preprint arXiv:2008.12172*.

- Rabindra Lamsal. 2020. [Coronavirus \(covid-19\) tweets dataset](#).
- Martin Müller, Marcel Salathé, and Per E Kummervold. 2020. Covid-twitter-bert: A natural language processing model to analyse covid-19 content on twitter. *arXiv preprint arXiv:2005.07503*.
- Dat Quoc Nguyen, Thanh Vu, Afshin Rahimi, Mai Hoang Dao, Linh The Nguyen, and Long Doan. 2020. WNUT-2020 Task 2: Identification of Informative COVID-19 English Tweets. In *Proceedings of the 6th Workshop on Noisy User-generated Text*.
- Ratnabali Pal, Arif Ahmed Sekh, Samarjit Kar, and Dilip K Prasad. 2020. Neural network based country wise risk prediction of covid-19. *arXiv preprint arXiv:2004.00959*.
- Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*.
- Eunha Shim, Amna Tariq, Wongyeong Choi, Yiseul Lee, and Gerardo Chowell. 2020. Transmission potential and severity of covid-19 in south korea. *International Journal of Infectious Diseases*.
- Anuradha Tomar and Neeraj Gupta. 2020. Prediction for the spread of covid-19 in india and effectiveness of preventive measures. *Science of The Total Environment*, page 138762.
- T. Wang, K. Lu, K. P. Chow, and Q. Zhu. 2020. Covid-19 sensing: Negative sentiment analysis on social media in china via bert model. *IEEE Access*, 8:138162–138169.
- Shi Zhao, Qianyin Lin, Jinjun Ran, Salihu S Musa, Guangpu Yang, Weiming Wang, Yijun Lou, Daozhou Gao, Lin Yang, Daihai He, et al. 2020. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in china, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International journal of infectious diseases*, 92:214–217.