BANANA at WNUT-2020 Task 2: Identifying COVID-19 Information on Twitter by Combining Deep Learning and Transfer Learning Models

Tin Van Huynh
University of Information Technology
VNU-HCM, Vietnam
16521827@gm.uit.edu.vn

Luan Thanh Nguyen
University of Information Technology
VNU-HCM, Vietnam
17520721@gm.uit.edu.vn

Son T. Luu
University of Information Technology
VNU-HCM, Vietnam
sonlt@uit.edu.vn

Abstract

The outbreak COVID-19 virus caused a significant impact on the health of people all over the world. Therefore, it is essential to have a piece of constant and accurate information about the disease with everyone. This paper describes our prediction system for WNUT-2020 Task 2: Identification of Informative COVID-19 English Tweets. The dataset for this task contains size 10,000 tweets in English labeled by humans. The ensemble model from our three transformer and deep learning models is used for the final prediction. The experimental result indicates that we have achieved F1 for the INFORMATIVE label on our systems at 88.81% on the test set.

1 Introduction

The rapid spread of the coronavirus (COVID-19) has caused a global health crisis. This virus is hazardous to people's health and causes a big panic all over the world. Statistics show that each day there are 4 million tweets related to COVID-19 on Twitter (Lamsal, 2020). Therefore, it is essential to keep track of the information associated with this disease. Along with the development of many social networking platforms such as Twitter and Facebook. This is the primary way that helps people capture information about COVID-19 regularly. However, there is much content appearing daily on these social media platforms. Most of them do not have information about the status of COVID-19, such as the number of suspected cases or cases near the user's area.

In this article, we present our approach at WNUT-2020 Task 2 (Nguyen et al., 2020) to identify Tweets containing information about COVID-19 on the social networking platform Twitter or not. A Tweet is believed to have information if it includes information such as recovered, suspected, confirmed, and death cases and location or travel

Tweet	Label
A New Rochelle rabbi and a White Plains doctor are among the 18 confirmed coronavirus cases in Westchester. HTTPURL	0
Day 5: On a family bike ride to pick up dinner at @USER Broadway, we encountered our pre-COVID-19 Land Park happy hour crew keeping up the tradition at an appropriate #SocialDistance.HTTPURL	1

Table 1: Several examples in the WNUT-2020 Task 2 dataset. 0 and 1 stand for INFORMATIVE and UNINFORMATIVE, respectively.

history of the patients. Specifically, we described the problem as follows.

- **Input**: Given English Tweets on the social networking platform.

- **Output**: One of two labels (INFORMATIVE and UNINFORMATIVE) predicted by our system.

Several examples are shown in Table 1

In this paper, we have two main contributions as follows.

- Firstly, we implemented four different models based on neural networks and transformers such as Bi-GRU-CNN, BERT, RoBERTa, XLNet to solve the WNUT-2020 Task 2: Identification of informative COVID-19 English Tweets.

- Secondly, we propose a simple ensemble model by combining multiple deep learning and transformer models. This model gives the highest performance compared with the single

models with F1 on the test set is 88.81% and on the development set is 90.65%.

2 Related work

During the happening of the COVID-19 pandemic, the information about the number of infected cases, the number of patients is vital for governments. Dong et al. (2020) constructed a real-time database for tracking the COVID-19 around the world. This dataset is collected by experts from the World Health Organization (WHO), US CDC, and other medical agencies worldwide and is operated by John Hopkins University. Also, there are many other COVID-19 datasets such as multilingual data collected on Twitter from January 2020 (Chen et al., 2020) or Real World Worry Dataset (RWWD) (Kleinberg et al., 2020).

Besides, on social media, the spreading of COVID-19 information is extremely fast and enormous and sometimes leads to misinformation. Shahi et al. (2020) conducted a pilot study about detecting misinformation about COVID-19 on Twitter by analyzing tweets using standard social media analytics techniques. From the researching results, the authors want to help authorities and social media users counter misinformation. Moreover, the rumors and conspiracy theories within the emergence times of COVID-19 spreading had made communities feel fearmongering and panicky, which lead to racism about COVID-19 patients and citizens from infected countries, and mass purchase of face masks as well as the shortage of necessaries, according to (Depoux et al., 2020). Thus it is necessary to identify the right information from the social media text.

3 Dataset

The dataset provided by Nguyen et al. (2020) contains 10,000 English Tweets about COVID-19, which is used to automatically identify whether a tweet contains useful information about the COVID-19 (informative) or not (uninformative). There are 4,719 INFORMATIVE tweets and 5,281 UNINFORMATIVE tweets in the dataset, and three different annotators annotate each tweet. The inter-annotator agreement calculated by Fleiss' Kappa score of the dataset is 81.80%. Also, the dataset is split into the training, development, and test sets with proportion 7-1-2. Table 2 shows the overview information about the dataset.

	INFORMATIVE	UNINFORMATIVE
Training	3,303	3,697
Development	472	528
Test	944	1,056

Table 2: Overview of the WNUT-2020 Task 2 dataset.

4 Methodologies

In this paper, we propose an ensemble method that combines the deep learning models with the transfer learning models to identify information about COVID-19 from users' tweets.

4.1 Deep neural model

We implement the Bi-GRU-CNN model, which was used for salary prediction by Wang et al. (2019) and Job prediction by Van Huynh et al. (2020), with the GloVe-300d word embedding (Pennington et al., 2014). This model consists of three main layers: the word representation layers (word embedding), the 1D Convolutional layers (CONV-1D), and the bidirectional GRU layer (Bi-GRU). This model also achieved high performances on previous study works(Wang et al., 2019; Van Huynh et al., 2019, 2020). Fig 1 illustrates the Bi-GRU-CNN model.
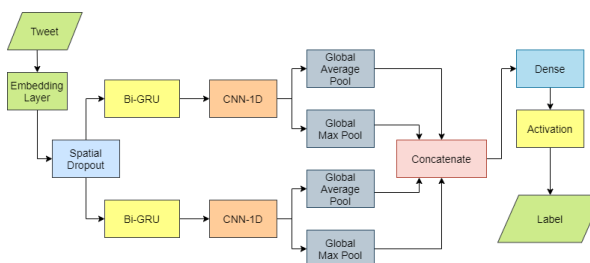

Figure 1: Overview architecture of Bi-GRU-CNN model.

4.2 Transfer learning model

Inspired by transfer learning success on many NLP tasks such as text classification (Do and Ng, 2006; Rizoiu et al., 2019) and machine reading comprehension (Devlin et al., 2019; Van Nguyen et al., 2020). In this paper, we used the SOTA transfer learning models, such as BERT (Devlin et al., 2019), RoBERTa (Liu et al., 2019), and XLNet (Yang et al., 2019) with fine-tuning techniques for the problem of identifying informative tweet about COVID-19. In our experiment, we used the pre-trained language model, as described in Table 3. All of these pre-trained models are constructed on English texts.

Transfer model	Pre-trained model
BERT	*bert_en_uncased*
RoBERTa	*roberta-base*
XLNet	*xlnet-large-cased*

Table 3: List of transformer models used in our experiment.

4.3 Ensemble method

As the success of the ensemble models of previous tasks (Van Huynh et al., 2020; Nguyen et al., 2019), we propose a simple yet effective ensemble approach with the majority voting between the outputs of four different models, including Bi-GRU-CNN, BERT, RoBERTa, and XLNet for classifying whether a tweet contains information about COVID-19 or not. Fig 2 describes our ensemble model.
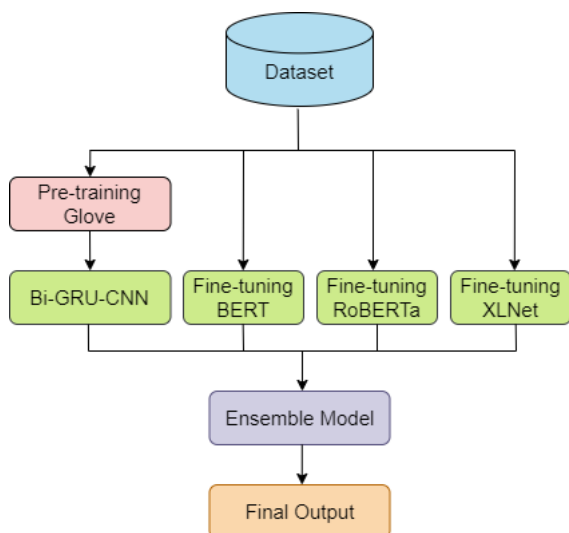

Figure 2: Overview architecture of our ensemble approach.

5 Experiment

5.1 Experimental settings

In this study, we experimented with datasets provided by WNUT-2020 Task 2. Training, development, and testing sets are divided as described in Section 3. To evaluate our models, we use four metrics include accuracy, precision, recall, and F1.

To prepare data for the model training and model evaluation phases, we perform the simple and effective pre-processing of input data as follows:

- Step 1: Converting the tweet into the lower-case strings.

- Step 2: Removing the user names in the tweets.

- Step 3: Deleting all URLs in the tweets.

- Step 4: Representing words into vectors with pre-trained word embedding sets for deep neural network models.

According to analyzing the length of the tweets in the data, we set max_length of the models to be 512 and epochs to be 15 for two models Bi-GRU-CNN and XLNet, and 3 for model BERT and RoBERTa. After searching for extensive hyper-parameter, we set learning_rate equal to 1e-3 and dropout equal to 0.2 for the Bi-GRU-CNN model and learning_rate equal to 1e-5 and dropout equal to 0.1 for three models BERT, RoBERTa, and XLNet.

5.2 Experimental results

Experimental results of the single model and the ensemble model on the development set are presented in Table 4. Specifically, in the single models, the Bi-GRU-CNN model gives the lowest performance with 85.66% by F1 and 86.10% by accuracy. The single model with the highest efficiency is XLNet, which attained 89.86% by F1 and 90.30% by accuracy. In addition, the BERT model gives the highest Precision with 89.53%, and the RoBERTa model achieved the highest Recall result with 90.74%. In particular, our recommend ensemble model gives the best performance when combining the power of single models together, which accomplished 90.65%, 91.00%, and 92.37% by F1, Accuracy, and Recall respectively, according to Table 4. Specifically, our model improved 0.79% by F1 and 0.70% by Accuracy over the most extensive single model (XLNet), and 0.63% by Recall over the RoBERTa model.

Model	Accuracy	Precision	Recall	F1
Bi-GRU-CNN	86.10	83.50	87.92	85.66
BERT	89.79	**89.53**	88.77	89.15
RoBERTa	89.90	87.47	**91.74**	89.56
XLNet	**90.30**	88.66	91.10	**89.86**
Ensemble	**91.00**	88.98	92.37	90.65

Table 4: Model performances on the development set of the COVID-19 dataset

After the system evaluation of WNUT-2020 Task 2, Table 5 displays our ensemble model results on the testing set. This result is compared with the top 5 highest teams' results and the baseline model

(BASELINE - FASTTEXT). Our model with F1 is 88.81%, 2.15% lower than the first rank team, and 13.78% higher than the baseline model. As for the results of accuracy, we get 89.40%, 2.10% lower than the first rank team, and 12.10% higher than the baseline model.

Team Name	Accuracy	Precision	Recall	F1
NutCracker	91.50	91.35	90.57	90.96
NLP_North	91.40	90.29	91.63	90.96
SupportNUTMachine	91.40	90.46	91.42	90.94
#GCDH	91.25	89.19	92.69	90.91
Loner	91.20	89.18	92.58	90.85
BASELINE - FASTTEXT	77.30	72.88	77.10	75.03
BANANA	**89.40**	**88.53**	**89.09**	**88.81**

Table 5: Comparison with our model's performance with that of other teams.

5.3 Result analysis

Fig. 3 describes the confusion matrix of the ensemble methods when predicting informative tweets about COVID-19. It can be inferred from Fig. 3 that the ability of prediction correct label on the INFORMATIVE label is higher than the UNINFORMATIVE label.
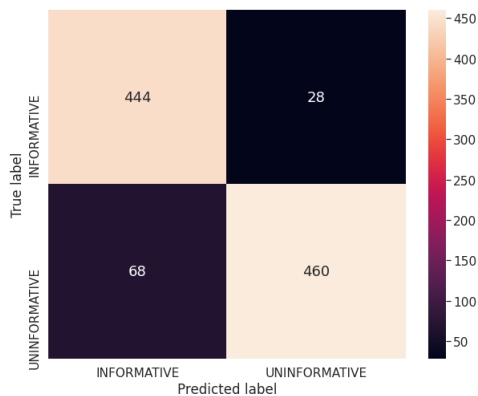

Figure 3: Confusion matrix of our ensemble model on the COVID English Tweet dataset.

In addition, Table 6 displays some error prediction examples from the dataset. Most of the wrong predictions occurred because of the appearance of special characters such as hashtag, the **HTTPURL** phrases, which stand for the URL links in the tweets. For the INFORMATION tweets, the appearance of the **HTTPURL** phrase and the hashtag **#coronavirus** make the classification model predict the wrong label. This mistake is the same for the UNINFORMATION tweets, where the appearance of **HTTPURL** phrase and the hashtags related to the Coronavirus affected the results of the prediction model.

Tweet	PL	TL
NATIONAL NEWS: Coronavirus: Linda Lusardi says COVID-19 made her want to die and turned her vomit blue HTTPURL	UN	IN
Oh, he was sick before getting on the flight back to Australia. So Vail, Denver and possibly LA (if the layover was long enough) are suspect. I bet the LA airport is where he got it :(#coronavirus	UN	IN
20/03/20 PRESS RELEASE Council Urges Consumers to be Considerate Fears about a positive coronavirus (COVID-19) case led shoppers in Fiji to begin stocking up on supplies to fill pantries. HTTPURL	UN	IN

(a) Misclassified INFORMATION examples.

Tweet	PL	TL
Vegetable market Peshawar KP.! People here are least worried about #COVID19 .! 1 infected Asymptomatic person and he would be transmitting it to complete city inside and outside ! HTTPURL	IN	UN
FYI, the state's #COVID19 stats show 395 people have been tested via state lab. But doesn't show a total for all labs. HTTPURL	IN	UN
New post (Mike Pence Celebrates Story of Great Great Grandmother Recovering from Coronavirus) has been published on Randy Salars News And Comment - HTTPURL HTTPURL	IN	UN

(b) Misclassified UNINFORMATION examples.

Table 6: Misclassified examples. PL: predicted label, TL: true label

6 Conclusion and future work

This paper has addressed our work on the WNUT-2020 Task 2: Identifying COVID-19 Information on Twitter. We proposed our ensemble model combining the deep learning models and the transfer learning models for detecting information about COVID-19 from users' tweets. Our ensemble model achieved 91.00% by accuracy and 90.65% by F1 on the development set, and achieved 89.04% by accuracy and 88.81% by F1 on the public test set, which ranked #25 in the competition.

In the future, we will improve our model's performance by exploring different features of the users' tweets and transfer learning models with fine-tuning techniques. Finally, we hope our study can be applied in practice for detecting COVID-19 from social networks to support the COVID-19 battle all over the world.

References

Emily Chen, Kristina Lerman, and Emilio Ferrara. 2020. Tracking social media discourse about the

covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health and Surveillance*, 6(2):e19273.

Anneliese Depoux, Sam Martin, Emilie Karafillakis, Raman Preet, Annelies Wilder-Smith, and Heidi Larson. 2020. The pandemic of social media panic travels faster than the COVID-19 outbreak. *Journal of Travel Medicine*, 27(3). Taaa031.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Chuong B Do and Andrew Y Ng. 2006. Transfer learning for text classification. In *Advances in Neural Information Processing Systems*, pages 299–306.

Ensheng Dong, Hongru Du, and Lauren Gardner. 2020. An interactive web-based dashboard to track covid-19 in real time. *The Lancet Infectious Diseases*, 20.

Bennett Kleinberg, Isabelle van der Vegt, and Maximilian Mozes. 2020. Measuring emotions in the covid-19 real world worry dataset. *arXiv preprint arXiv:2004.04225*.

Rabindra Lamsal. 2020. Coronavirus (covid-19) tweets dataset.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach.

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*.

Duc-Vu Nguyen, Kiet Van Nguyen, and Ngan Luu-Thuy Nguyen. 2019. Nlp@uit at vlsp 2019: A simple ensemble model for vietnamese dependency parsing. *The Sixth International Workshop on Vietnamese Language and Speech Processing VLSP 2019*.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

Marian-Andrei Rizoiu, Tianyu Wang, Gabriela Ferraro, and Hanna Suominen. 2019. Transfer learning for hate speech detection in social media. *arXiv preprint arXiv:1906.03829*.

Gautam Kishore Shahi, Anne Dirkson, and Tim A. Majchrzak. 2020. An exploratory study of covid-19 misinformation on twitter.

Tin Van Huynh, Vu Duc Nguyen, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen, and Anh Gia-Tuan Nguyen. 2019. Hate speech detection on vietnamese social media text using the bi-gru-lstm-cnn model. *arXiv preprint arXiv:1911.03644*.

Tin Van Huynh, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen, and Anh Gia-Tuan Nguyen. 2020. Job prediction: From deep neural network models to applications. In *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, pages 1–6. IEEE.

Kiet Van Nguyen, Duc-Vu Nguyen, Anh Gia-Tuan Nguyen, and Ngan Luu-Thuy Nguyen. 2020. New vietnamese corpus for machine reading comprehension of health news articles. *arXiv preprint arXiv:2006.11138*.

Zhongsheng Wang, Shinsuke Sugaya, and Dat PT Nguyen. 2019. Salary prediction using bidirectional-gru-cnn model. *Assoc. Nat. Lang. Process.*

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in Neural Information Processing Systems 32*, pages 5753–5763. Curran Associates, Inc.