Cross-lingual COVID-19 Fake News Detection

Jiangshu Du*, Yingtong Dou*, Congying Xia*, Limeng Cui†, Jing Ma‡ and Philip S. Yu*

*Department of Computer Science, University of Illinois at Chicago, USA
†Department of Computer Science and Engineering, Pennsylvania State University, USA
‡Department of Computer Science, Hong Kong Baptist University, HongKong
{jdu25, ydou5, cxia8, psyu}@uic.edu, lzc334@psu.edu, majing@hkbu.edu.hk

Abstract—**The COVID-19 pandemic poses a great threat to global public health. Meanwhile, there is massive misinformation associated with the pandemic which advocates unfounded or unscientific claims. Even major social media and news outlets have made an extra effort in debunking COVID-19 misinformation, most of the fact-checking information is in English, whereas some unmoderated COVID-19 misinformation is still circulating in other languages, threatening the health of less-informed people in immigrant communities and developing countries. In this paper, we make the first attempt to detect COVID-19 misinformation in a low-resource language (Chinese) only using the fact-checked news in a high-resource language (English). We start by curating a Chinese real&fake news dataset according to existing fact-checking information. Then, we propose a deep learning framework named CrossFake to jointly encode the cross-lingual news body texts and capture the news content as much as possible. Empirical results on our dataset demonstrate the effectiveness of CorssFake under the cross-lingual setting and it also outperforms several monolingual and cross-lingual fake news detectors. The dataset is available at https://github.com/YingtongDou/CrossFake.**

Index Terms—**fake news; COVID-19; cross-lingual; dataset**

I. INTRODUCTION

The booming of social media fosters various kinds of misinformation where fake news is a typical form that represents news articles that are intentionally and verifiably false [1]. Amid the unprecedented COVID-19 global pandemic, the mysterious cause of the coronavirus and its severe infectiousness have incited more fake news. For instance, many news articles falsely claim that drinking or injecting bleach could kill the COVID-19 virus, which has already made detrimental hurt to less-informed individuals[1]. The WHO has officially declared the wide-spreading COVID-19 misinformation as a "*infodemic*" and called for mitigating it [2].

Fact-checking by domain experts is a widely used approach to debunk COVID-19 misinformation. Besides professional fact-checking organizations like Politifact[2], the infodemic has prompted many social media, news outlets, and even governments to launch exclusive tools to debunk misinformation [3]. However, recent reports[3] notice that the COVID-19 misinformation has imposed threats to non-English speakers. Since those people do not consume English media, the vetted information in English is hardly accessed by them. Meanwhile, the lack of fact-checking or content moderation in some non-English media exacerbates the negative influence of misinformation. A case we have discovered is an English post falsely claiming that U.S. hospitals are preparing for 96 million coronavirus infections at the early phase of the pandemic[4], many news articles claiming the same thing are still existing on various Chinese social media platforms[5] after one year the source being debunked. Thus, it is imperative to develop an effective fake news detection model for low-resource languages.

Among previous fake news detection works [1], only a few investigated fake news under the cross-lingual or multi-lingual setting. Some papers adopted pre-trained multi-lingual encoders to encode the news in different languages [4], [5]; while a recent work utilized language-independent features to handle the multi-lingual setting [6], and another work applied transfer learning to map the monolingual word embeddings from different languages into the same space [7]. Besides the above works, several cross-lingual learning approaches have been applied to similar domains like hate speech [8] and abusive language detection [9].

Regarding the cross-lingual COVID-19 infodemic, the annotated news articles in low resource languages are scarce, and the news develops quickly across different languages with many new terms. Therefore, it is infeasible to train a monolingual model based on a low-resource language with few annotations. Moreover, the lack of news social engagement information in some low-resource languages impedes the application of social context-based fake news detectors [10].

To cope with the above challenges, we attempt to train a cross-lingual fake news detector trained solely based on a high resource language (English) COVID-19 news corpus and used to predict news credibility in a low resource language (Chinese). We start by curating a COVID-19 news dataset in Chinese based on existing fact-checking information. Then, we propose an end-to-end fake news detection framework named CrossFake based on pre-trained language models. To deal with the long news body text, it is sliced into sub-text groups before being fed into language encoders. Experimental results verify the effectiveness of proposed CrossFake comparing to monolingual and cross-lingual baselines. We also discuss the advantage and drawbacks of the proposed model using several real-world cases. We summarize our contributions as follows:

---

[1] https://nyti.ms/3j8E7OJ
[2] https://www.politifact.com/coronavirus/
[3] https://nyti.ms/3gqujOl, https://bit.ly/38BYZI7

[4] https://bit.ly/3keGK0L
[5] https://archive.vn/SVGuK, https://archive.vn/FOidI

- We collect and annotate a fine-grained cross-lingual COVID-19 fake news dataset.
- We propose an end-to-end cross-lingual fake news detector tailored to the news text properties.
- We empirically show the advantage and limitation of CrossFake comparing to mono/cross-lingual baselines.

II. DATA COLLECTION

Recent works [11]–[14] have released several annotated English COVID-19 news datasets according to the fact-checking information from recognized fact-checking institutions. In contrast, there is seldom fact-checked Chinese COVID-19 news corpus. Although [15] has curated a fact-checked Chinese COVID-19 misinformation dataset, it only contains short-text social media posts without complete news body text. We also notice some COVID-19 fact-checking websites operated by Chinese social media or government[6]. However, their fact-checking articles only analyze and verify specific claims related to COVID-19 without referring to their corresponding source news articles.

To obtain high-quality news data, we resort to annotate Chinese news manually. Given thousands of annotated English COVID-19 news in existing datasets [11], [12], we find their corresponding Chinese news as the annotated Chinese news and our data collection protocol is shown below:

Step 1. Select a piece of English news in existing datasets.
Step 2. Search the translated English title under three major Chinese news search engines[a].
Step 3. Check if there is a news title on the first page of search results that has a similar meaning as the original English news. If yes, go to Step 4; else, go to Step 1.
Step 4. Check if the content of the selected Chinese news and original English news express similar opinions towards the same event/claim. If yes, go to Step 5; else, go to Step 1.
Step 5. Collect the metadata of the selected Chinese news and add them to the dataset.

[a]https://sogou.com/, https://toutiao.com/, http://baidu.com/

Four data annotators with both English and Chinese knowledge are trained with the above protocol before collecting the data. Two annotators collect the fake news and the other two collect real news. All annotated news articles are discussed together with an independent fact-checking expert to guarantee accuracy and consistency. We have inspected more than one thousand fact-checked news spanning from January 2020 to October 2020 and have collected 86 fake and 114 real Chinese news. Note that even our dataset contains news metadata like title, author, and timestamp, we only investigate the news body text in this paper for simplicity. More dataset details are shown in Section IV-A.

III. METHODOLOGY

Problem Definition. This paper aims to predict the Chinese COVID-19 news truthfulness, while only a small number of
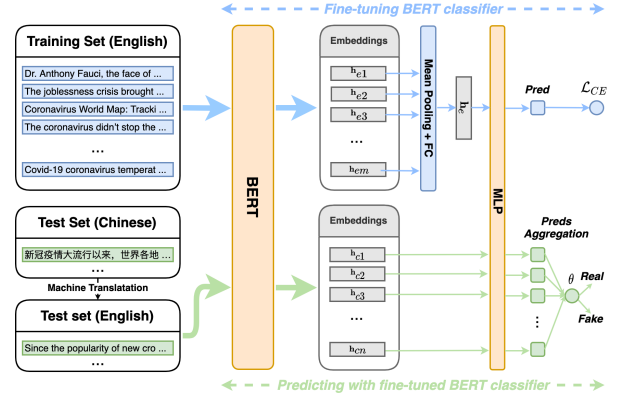
Fig. 1: The workflow of the proposed CrossFake fake news detector. We train a neural classifier based on the aggregated BERT embeddings of a fact-checked English news sub-text (i.e., token groups). To verify a Chinese news article, it is first translated into English, and the final predictions are made via aggregating all sub-text predictions.

annotated Chinese COVID-19 fake news is not enough to train a good supervised classifier. Thanks to the rich annotated English COVID-19 news corpus [11], [12], [14], we formulate our problem as a cross-lingual fake news detection task. Specifically, we define a set of English news articles as $e \in N_e$ ($c \in N_c$ for Chinese news, resp.) and their labels as $y_e \in Y_e$ ($y_c \in Y_c$ for Chinese news, resp.). The label for each news belongs to either 1 (fake news) or 0 (real news). Our objective is to train a classifier C with training data N_e and label Y_e and maximize the test accuracy of C on N_c and Y_c. The framework of the proposed CrossFake model is shown in Figure 1. Next, we elucidate how to train CrossFake and how it verifies Chinese news credibility.

Model Training. In the training phase. we encode the annotated English news using BERT [16], a language pre-training mode, as our base model. Comparing to social media posts, the news body text length is usually longer. As per our collected dataset, most of news has more than 512 tokens after tokenization, which exceeds the maximum tokens BERT can process. To tackle this problem, we decide to break the long body text into *sub-text* groups. Specifically, given the tokenized body text T_e for a piece of news e, we break T_e into a set of m token groups $TG_e = \{t_{e1}, \ldots, t_{em}\}$ sequentially where $m = \left\lceil \frac{|T_e|}{500} \right\rceil$. It is to say that one sub-text size is 500. Then, we can represent the final news embedding $\mathbf{h}_e$ as:

$$\mathbf{h}_e = \text{FC}\left(\frac{\sum_{i=1}^{m} \text{BERT}(t_{ei})}{m}\right), \tag{1}$$

where each token group $t_{ei} \in TG_e$ (sub-text) is encoded by BERT separately. A mean pooling layer and a fully-connected (FC) layer are applied over all sub-texts to yield the final news embedding $\mathbf{h}_e$. The fact-related information in news body text can be captured and retained as much as possible through the operation above.

TABLE I: Dataset statistics. Long Text% means the percentage of news articles exceeding 512 tokens after tokenization.

Dataset	Time	Lang.	Long Text%	Fake	Total
Training	Jan. - Oct. 2020	ENG	81.87%	49.23%	2840
Test		CHN	41.00%	43.00%	200

Since our data only has fake and real news, we adopt the binary cross-entropy loss function to update the classifier C:

$$\mathcal{L} = \sum_{e \in \mathcal{N}_e} - \log\left(y_e \cdot \text{ReLU}(\text{MLP}(\mathbf{h}_e))\right). \quad (2)$$

The loss function is optimized using SGD, and it is equivalent to fine-tuning the pre-trained BERT encoder.

News Verification. As we mentioned in the problem definition, we need to verify the truthfulness of Chinese news based on the classifier C trained in English. As Figure 1 illustrated, we first translate the test data (i.e., Chinese COVID-19 news) into English with Google Translator API[7] to align the input data language of C. We compared baselines that encode the Chinese directly, and their performance is worse than our approach (see more in Section IV-C). The test data tokenizer and encoder are the same as what is used on the training data. The sub-texts set for news c is $TG_c = \{t_{c1}, \ldots, t_{cn}\}$ where $n = \left\lceil \frac{|T_c|}{100} \right\rceil$. Note that we use 100 instead of 500 as the sub-text size for test data since the collected Chinese news articles are relatively shorter than English news (see Long Text% in Table I). The prediction result p_c for the test news c is obtained via the following equation:

$$p_c = \begin{cases} 1, if \ \frac{\sum_{j=1}^{n}|C(t_{cj})|}{n} \geq \theta, \\ 0, if \ \frac{\sum_{j=1}^{n}|C(t_{cj})|}{n} < \theta, \end{cases} \quad (3)$$

where we aggregate the prediction results of all sub-texts for c and θ is a classification threshold empirically set to 0.8.

IV. EXPERIMENT

A. Dataset

In this paper, only English news articles are used for the training purpose since we aim to leverage the fact-checking knowledge in a high resource language to verify news truthfulness in a low resource language (Chinese). Our training dataset consists of all English COVID-19 news from three datasets: ReCOVery [11], FakeCovid [14], and CoAID [12]. We evaluate the performance of proposed CrossFake on the Chinese COVID-19 news dataset collected in this paper as described in Section II. Table I shows the statistics of the training and test datasets.

B. Experimental Setup

Baselines We compare our proposed model together with several monolingual and cross-lingual fake news detectors:

- *Monolingual*: CSI [17], SAFE [18], and exBAKE [19] are three monolingual baselines. CSI employs an LSTM to encode the news content to detect fake news. SAFE uses

---

[7]https://pypi.org/project/google-trans-new/

TABLE II: The fake new classification performance with standard deviation over five runs. *-sub* (*-avg* resp.) represents that the final prediction is made via aggregating the prediction result of each sub-text (predicting based on the average embedding of sub-texts resp.).

Model	Accuracy	Precision	Recall	F1
CLEF [4]	$43.12_{0.41}$	$42.88_{0.43}$	$97.38_{3.89}$	$59.53_{1.15}$
EMET [5]	$45.90_{3.29}$	$42.15_{1.59}$	$70.93_{18.47}$	$51.89_{7.35}$
CSI [17]	$68.30_{1.29}$	$61.41_{1.77}$	$71.16_{5.01}$	$65.81_{2.02}$
SAFE [18]	$71.60_{2.71}$	$63.69_{3.80}$	$80.70_{3.72}$	$71.01_{1.45}$
exBAKE [19]	$64.30_{3.53}$	$55.64_{3.99}$	$92.09_{7.87}$	$68.96_{0.60}$
exBAKE-sub	$66.80_{2.91}$	$59.73_{3.05}$	$70.47_{10.34}$	$64.30_{4.88}$
CrossFake-avg	$73.60_{2.31}$	$64.84_{2.81}$	$85.35_{5.91}$	$73.51_{2.18}$
CrossFake-sub	$75.00_{3.94}$	$71.45_{5.14}$	$70.47_{7.56}$	$70.67_{5.12}$

the TextCNN [20] to encode the news textual information. exBAKE utilizes the vanilla BERT as the English text encoder. We train the models above on the English training data and evaluate them on the translated test data.

- *Cross-lingual*: CLEF [4] and EMET [5] are two cross-lingual fake news detection models. CLEF leverages Multilingual-BERT [16] to encode non-English news, and we adopt Multilingual-BERT to encode both English and Chinese data in the experiment. EMET proposes a framework to detect misleading social media posts across different languages with the multilingual transformer [21]. In our experiment, EMET only encodes news articles since our dataset does not include other data types.

Experimental Settings CSI, SAFE, exBAKE, CLEF, and CrossFake are implemented using PyTorch. For EMET, we use the author's code [5]. All baselines use the batch size of 16, and CrossFake is trained with batch size 1.

C. Result Analysis

The average fake news classification performance with standard deviations over five random runs is reported in Table II, *-sub* and (*-avg* represent two prediction approaches (predicting-aggregating and aggregating-prediction, respectively). From Table II, we have following conclusions:

1) CrossFake **has the best performance.** CrossFake-avg outperforms all baselines in accuracy and precision, CrossFake-sub obtains a further performance gain via aggregating prediction results. The good performance of CrossFake-avg benefits from the knowledge captured from long news articles, whereas exBAKE can only process the first 512 tokens of an article. By leveraging the average embedding of a long news text, the original news information can be better retained. Meanwhile, we can observe that aggregating sub-text predictions could help alleviate the bias induced by a classifier with a higher preference for fake news. Therefore, exBAKE-sub has better accuracy and precision than exBAKE.

2) CNN-based model outperforms the RNN-based model. SAFE outperforms CSI significantly in terms of accuracy (71.60% and 68.30%). It might be because sequential models like RNN and LSTM used in CSI; experience information forgetting in long sequences, which are prevalent in the news corpus. On the contrary, SAFE adopts CNN, which could extract the local critical information related to fact-checking.

3) Pre-trained multi-lingual models are ineffective. EMET and CLEF can encode Chinese news without translation. However, both of them perform poorly on the test set. CLEF's high false-positive rate (i.e., high recall, low precision) suggests it has an overfitting issue. Two reasons for the poor performance of EMET and CLEF are summarized below:

1) Multilingual-BERT and the multilingual-transformer adopted by CLEF and EMET are pre-trained on the standard corpora (Wikipedia, Reddit, etc.) and released before the COVID-19 pandemic. Therefore, they might lack the domain knowledge for the new events, and it is difficult to map the terms like "COVID-19" and their corresponding cross-lingual words to a similar space. It suggests that the pre-trained language models should be kept up-to-date to handle the emerging events.

2) Similar to exBAKE, the maximum input sequence length limits the representation capability of Multilingual-BERT and multilingual-transformer while there are many long news articles in our dataset. Besides slicing text like CrossFake, more effective approaches are demanded to encode the long news text.

V. Discussion and Limitation

Translation Quality. We find that machine translation quality is a bottleneck of cross-lingual tasks, especially in emerging events. For instance, as a novel term, the Chinese expression of "Coronavirus" is mistranslated as "new crown virus", misleading the fake news classifiers. Moreover, we have attempted to translate all English training data to Chinese and train a Chinese fake news classifier. The test performance is bad since the low translation quality harms the training data quality.

Information Location. A piece of fake news may present false information in the middle or at the end of its body text. For a fake news article discussing coconut oil can destroy Coronavirus [8], the annotated misinformation appears after a lengthy introduction, which exceeds the maximum sequence length most language models can process. Consequently, the fact-related information in this news will be discarded by those models. Therefore, a model that can capture an article's complete information is crucial for automatic fact-checking.

Dataset Size. Due to the difficulty of data collection, our test dataset size is relatively small compared to other fake news datasets. Though the results may not be generalizable, we hope our preliminary exploration and experimental results could encourage future works in this direction.

VI. Conclusion and Future Work

In this paper, we make the first attempt to detect COVID-19 fake news under a cross-lingual setting. We collect and annotate a Chinese COVID-19 news dataset and propose an end-to-end fake news detector, CrossFake, which is trained on English news and could detect most of the collected Chinese fake news after translation. Our experimental results demonstrate the advantage of encoding more news content and

the limitation of pre-trained multi-lingual encoders. We will release the annotated data with rich metadata and encourage future research on cross-lingual fake news detection, especially in designing better models to capture the knowledge across different languages. Moreover, the event-centric analysis based on our data is another research direction.

References

[1] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," *CSUR*, 2020.

[2] W. H. Organization, "Managing the covid-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation," https://bit.ly/3DjoMTM, 2020.

[3] S. Siwakoti, K. Yadav, N. Bariletto, U. Zanotti, L.and Erdogdu, and J. N. Shapiro, "How covid drove the evolution of fact-checking," *Harvard Kennedy School Misinformation Review*, 2021.

[4] D. Dementieva and A. Panchenko, "Fake news detection using multilingual evidence," in *DSAA*, 2020.

[5] S. Schwarz, A. Theóphilo, and A. Rocha, "Emet: Embeddings from multilingual-encoder transformer for fake news detection," in *ICASSP*, 2020.

[6] I. Vogel and M. Meghana, "Detecting fake news spreaders on twitter from a multilingual perspective," in *DSAA*, 2020.

[7] S. Haider, L. Luceri, A. Deb, A. Badawy, N. Peng, and E. Ferrara, "Detecting social media manipulation in low-resource languages," *arXiv preprint arXiv:2011.05367*, 2020.

[8] L. Stappen, F. Brunn, and B. Schuller, "Cross-lingual zero-and few-shot hate speech detection utilising frozen transformer language models and axel," *arXiv preprint arXiv:2004.13850*, 2020.

[9] G. Glavaš, M. Karan, and I. Vulić, "Xhate-999: Analyzing and detecting abusive language across domains and languages," in *ICCL*, 2020.

[10] V. Nguyen, K. Sugiyama, P. Nakov, and M. Kan, "Fang: Leveraging social context for fake news detection using graph representation," in *CIKM*, 2020.

[11] X. Zhou, A. Mulay, E. Ferrara, and R. Zafarani, "Recovery: A multimodal repository for covid-19 news credibility research," in *CIKM*, 2020.

[12] L. Cui and D. Lee, "Coaid: Covid-19 healthcare misinformation dataset," *arXiv preprint arXiv:2006.00885*, 2020.

[13] Y. Li, B. Jiang, K. Shu, and H. Liu, "Mm-covid: A multilingual and multidimensional data repository for combatingcovid-19 fake new," *arXiv preprint arXiv:2011.04088*, 2020.

[14] G. K. Shahi and D. Nandini, "Fakecovid–a multilingual cross-domain fact check news dataset for covid-19," *arXiv preprint arXiv:2006.11343*, 2020.

[15] C. Yang, X. Zhou, and R. Zafarani, "Checked: Chinese covid-19 fake news dataset," *arXiv preprint arXiv:2010.09029*, 2020.

[16] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *NAACL*, 2019.

[17] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection," in *CIKM*, 2017.

[18] X. Zhou, J. Wu, and R. Zafarani, "Safe: Similarity-aware multi-modal fake news detection," in *PAKDD*, 2020.

[19] H. Jwa, D. Oh, K. Park, J. M. Kang, and H. Lim, "exbake: automatic fake news detection model based on bidirectional encoder representations from transformers (bert)," *Applied Sciences*, 2019.

[20] Y. Zhang and B. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," *arXiv preprint arXiv:1510.03820*, 2015.

[21] Y. Yang, D. Cer, A. Ahmad, M. Guo, N. Law, J.and Constant, G. H. Abrego, S. Yuan, C. Tar, Y. Sung *et al.*, "Multilingual universal sentence encoder for semantic retrieval," *arXiv preprint arXiv:1907.04307*, 2019.

---

[8]https://bit.ly/3sFd1lv, https://bit.ly/2WeC4Qr