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

IEEE ICDM Workshop'21

220111 Chia-Chun Ho

Outline

Introduction
COVID-19 fake news

- 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 falsely claim that drinking or injecting bleach could kill the COVID-19 virus, which has already made detrimental hurt to less-informed individuals.
- The WHO has officially declared the wide-spreading COVID-19 misinformation as a "infodemic" and called for mitigating it.

Introduction

Fake news in other languages

- Recent reports 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.

Introduction

Fake news in other languages

- A case the authors have discovered is an English post falsely claiming that US hospitals are preparing for 96 million coronavirus infections at the early phase of the pandemic.
  - Many news articles claiming the same thing are still existing on various Chinese social media platforms after one year the source being debunking.
- Thus, it's imperative to develop an effective fake news detection model for low-resource language.

Introduction

Cross-lingual fake news detection

- In previous works, 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.
- A recent work utilized language-independent features to handle the multi-lingual setting
- Another work applied transfer learning to map the monolingual word embeddings from different languages into the same space.
- Besides the above works, several cross-lingual learning approaches have been applied to similar domains like hate speech and abusive language detection.

Introduction

Cross-lingual COVID-19 fake news detection

- The annotated news articles in low resource languages are scarce, and the news develops quickly across different languages with many new terms.
  - Therefore, it's 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 language impedes the application of social context-based fake news detectors.

Introduction

CrossFake

- To cope with above challenges, 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).
- The authors curating a COVID-19 news dataset in Chinese based on existing fact-checking information.

Introduction

CrossFake

- 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's 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.

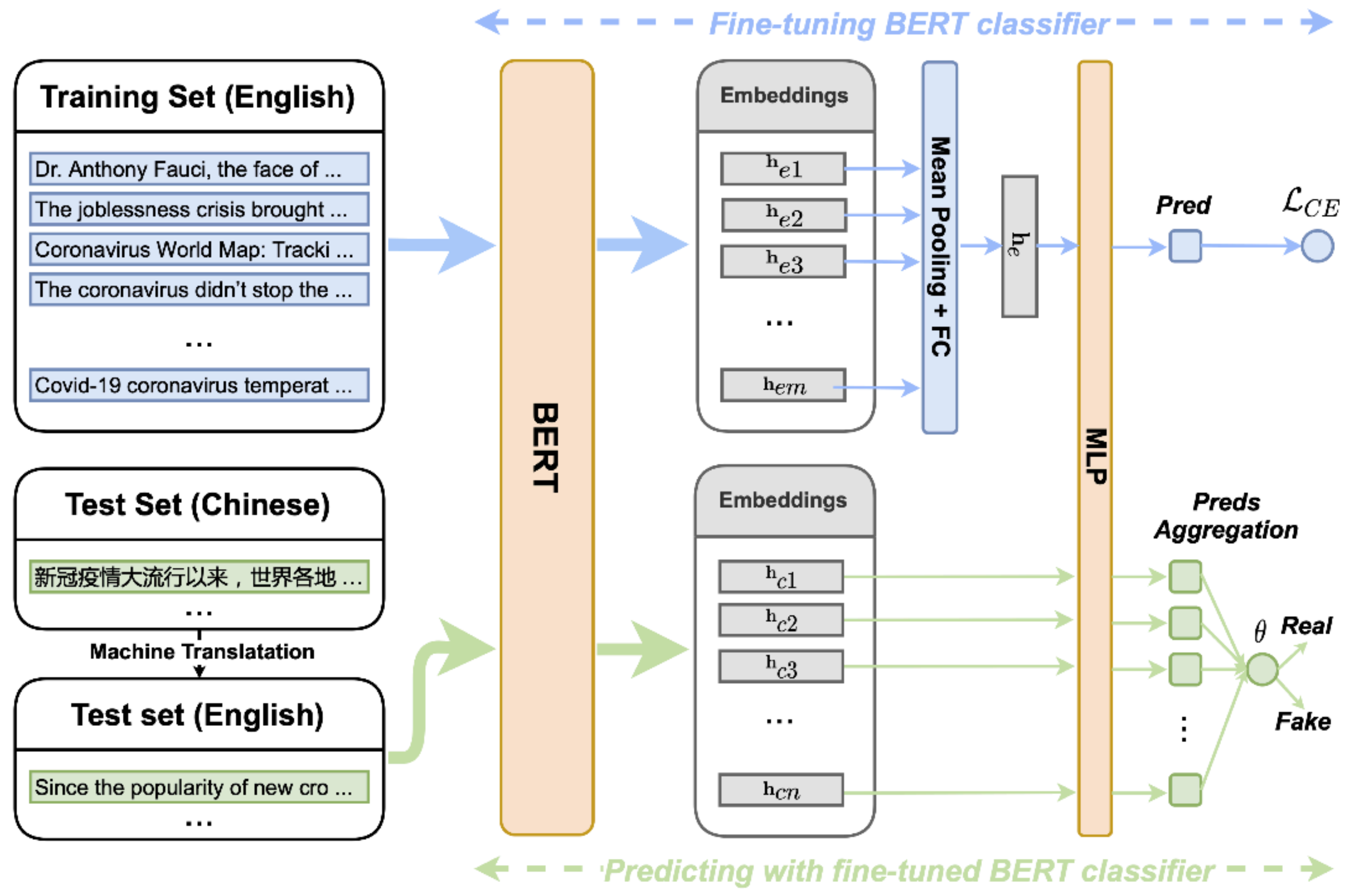


Fig. 1: The workflow of the proposed CrossFake fake news detector. We train a neural classifier based on the aggregated BERT embeddings of a fact-checked English news sub-text (i.e., token groups)). To verify a Chinese news article, it is first translated into English, and the final predictions are made via aggregating all sub-text predictions.

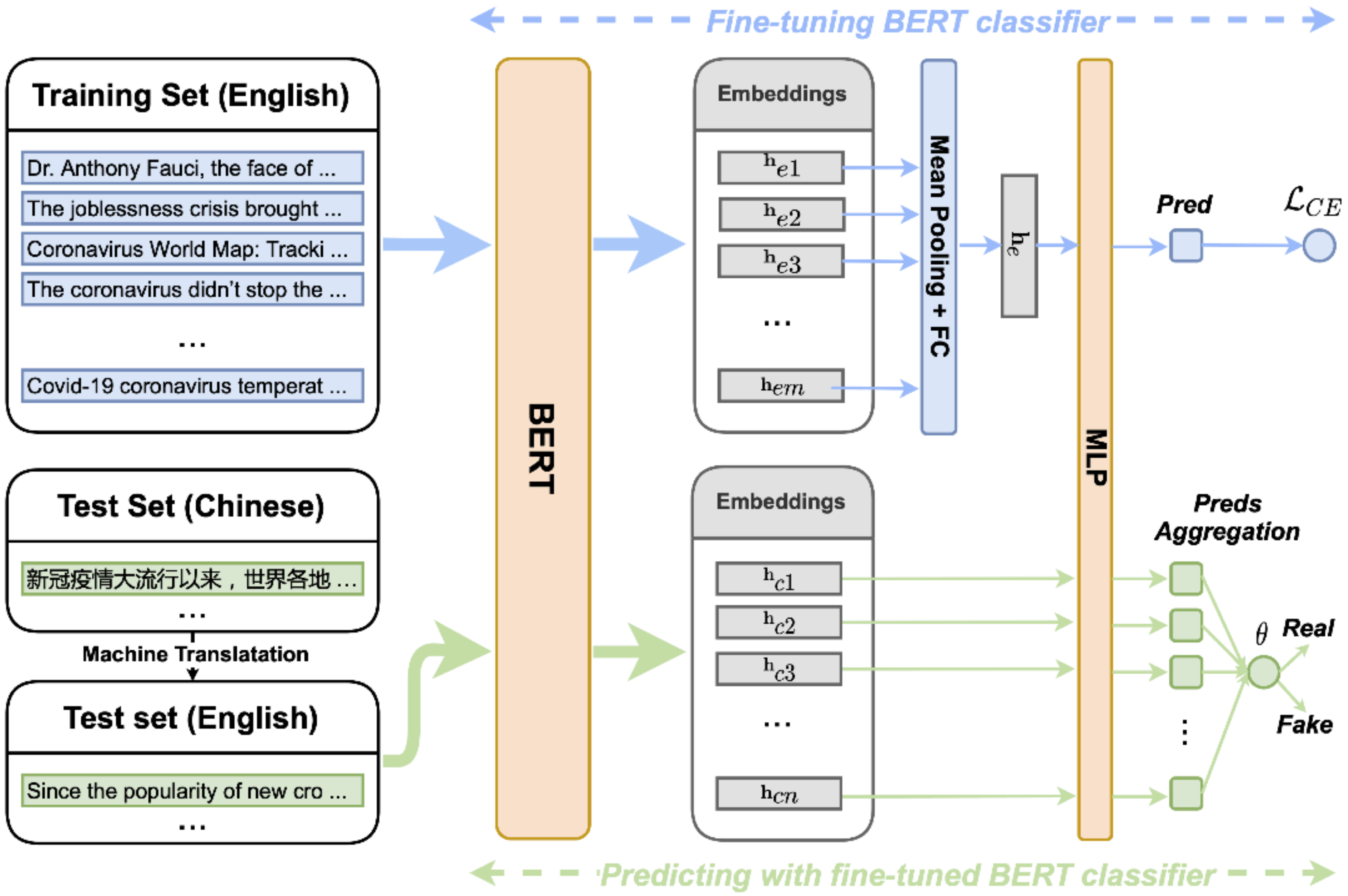
Introduction

Contribution

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

Methodology

Framework overview



Methodology

Problem Definition

- Aims to predict the Chinese COVID-19 news truthfulness, while only a small number of annotated Chinese COVID-19 fake news is not enough to train a good supervised classifier.
- Formulated problem as a cross-lingual fake news detection task.
- $e \in N_e$: English news; $c \in N_c$: Chinese news
- $y_e \in Y_e$: label of e; $y_c \in Y_c$: label of c; (1: fake news, 0: real news)
- 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.

Methodology

Model Training

- Encode the annotated English news using BERT as based model.
- Comparing to social media posts, the news body text length is usually longer.
- As per author's collected dataset, most of news has more than 512 tokens after tokenized, exceeds the max tokens BERT can process.

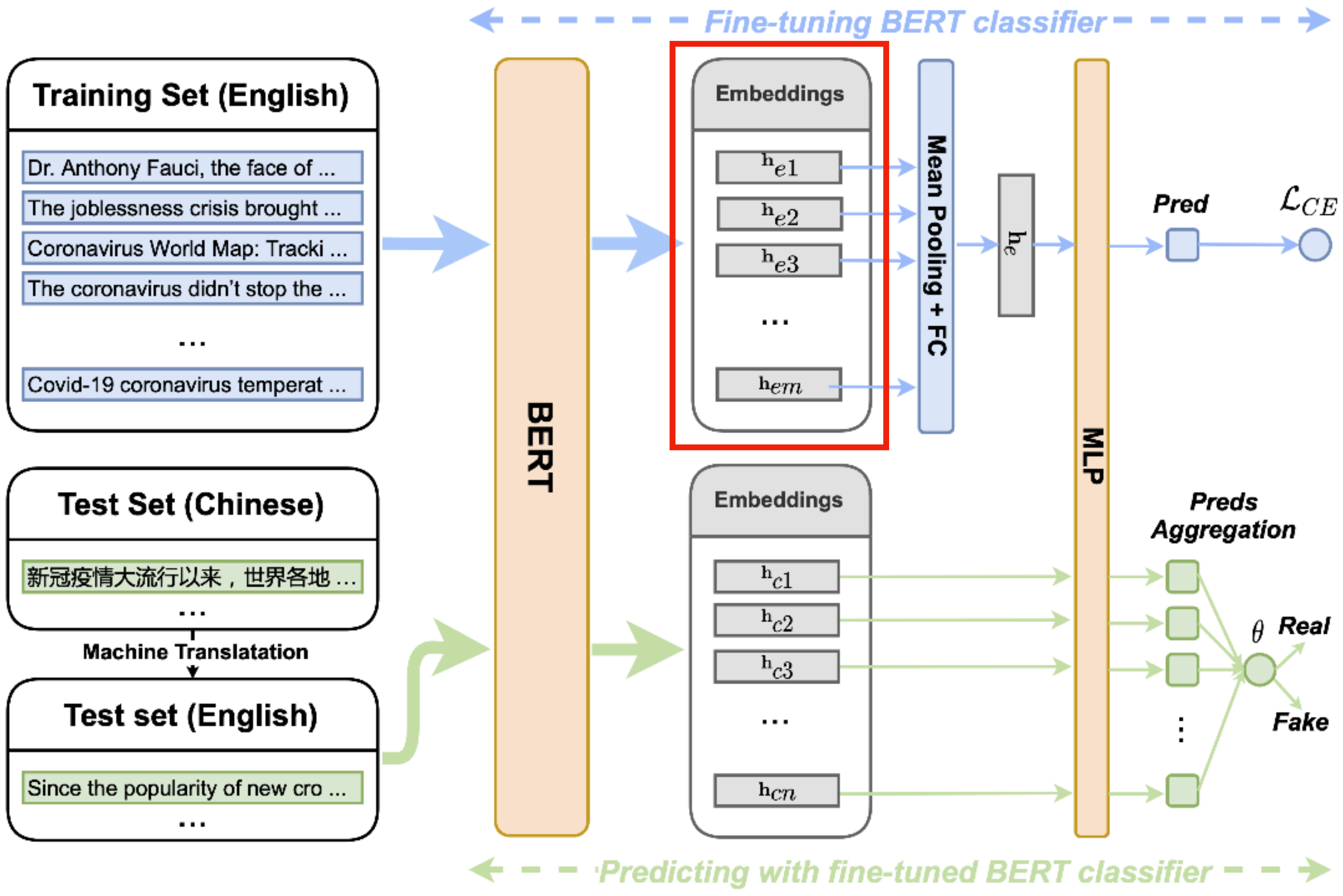
Methodology

Model Training

- To tackle this problem, decide to break the long body text into sub-text groups.

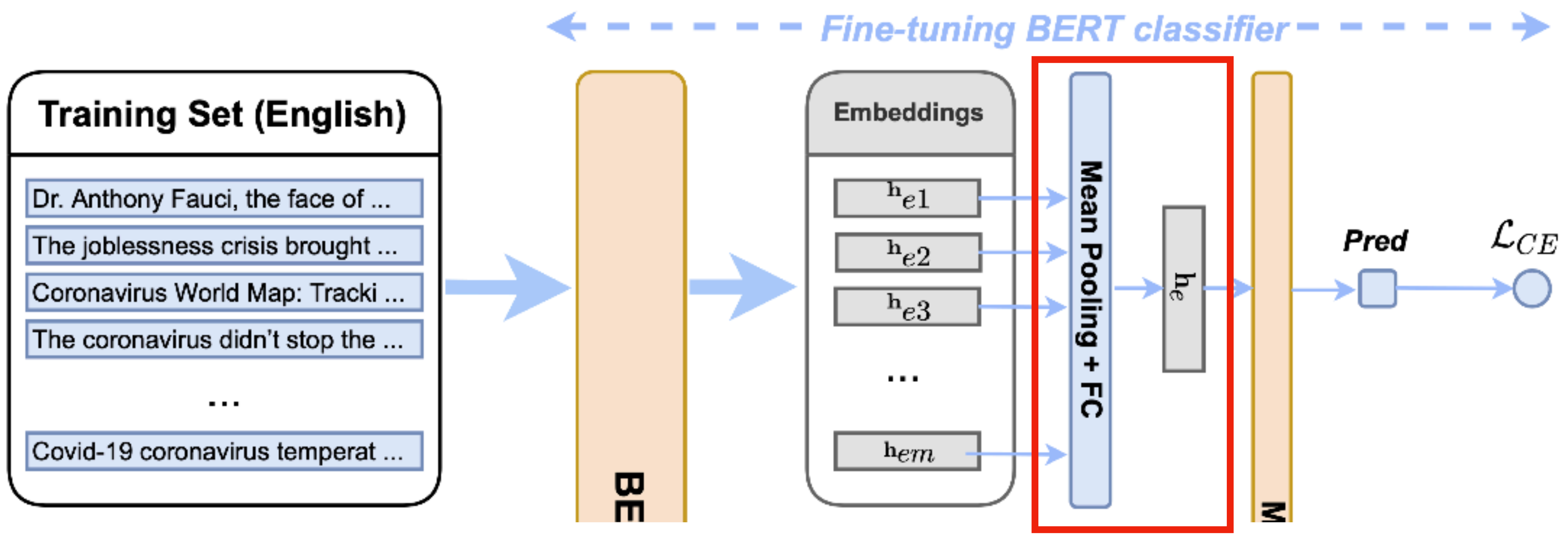
- Specifically, given the tokenized body text T_e for a piece of news e, break T_e into a set of m token groups $TG_e = \{t_{e_1}, \cdots, t_{e_m}\}$ sequentially where $m = \left\lceil \frac{|T_e|}{500} \right\rceil$.

- It's to say that one sub-text size is 500.



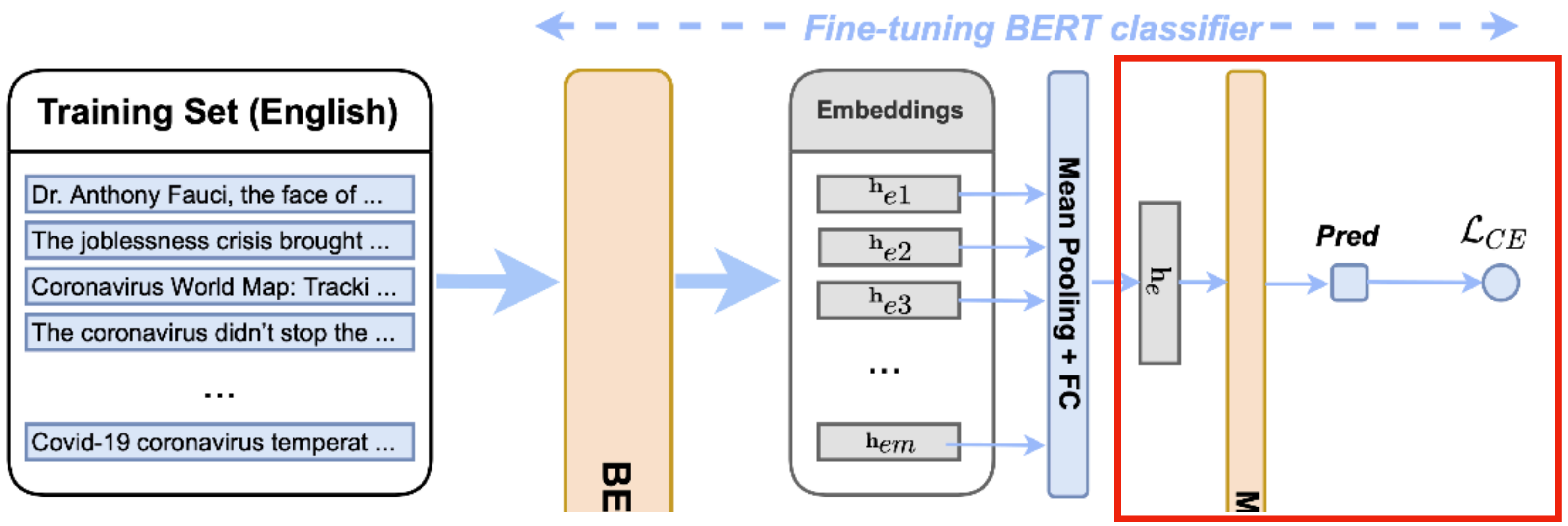
Methodology

Model Training



- Represent the final news embedding $\mathbf{h}_e = \mathrm{FC}\left(\frac{\sum_{i=1}^{m} \mathrm{BERT}\left(t_{ei}\right)}{m}\right)$
- Each token group $t_{ei} \in TG_e$ (sub-text) is encoded by BERT separately.
- A mean pooling layer and a fully-connected layer are applied over all sub-texts to yield the final 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.

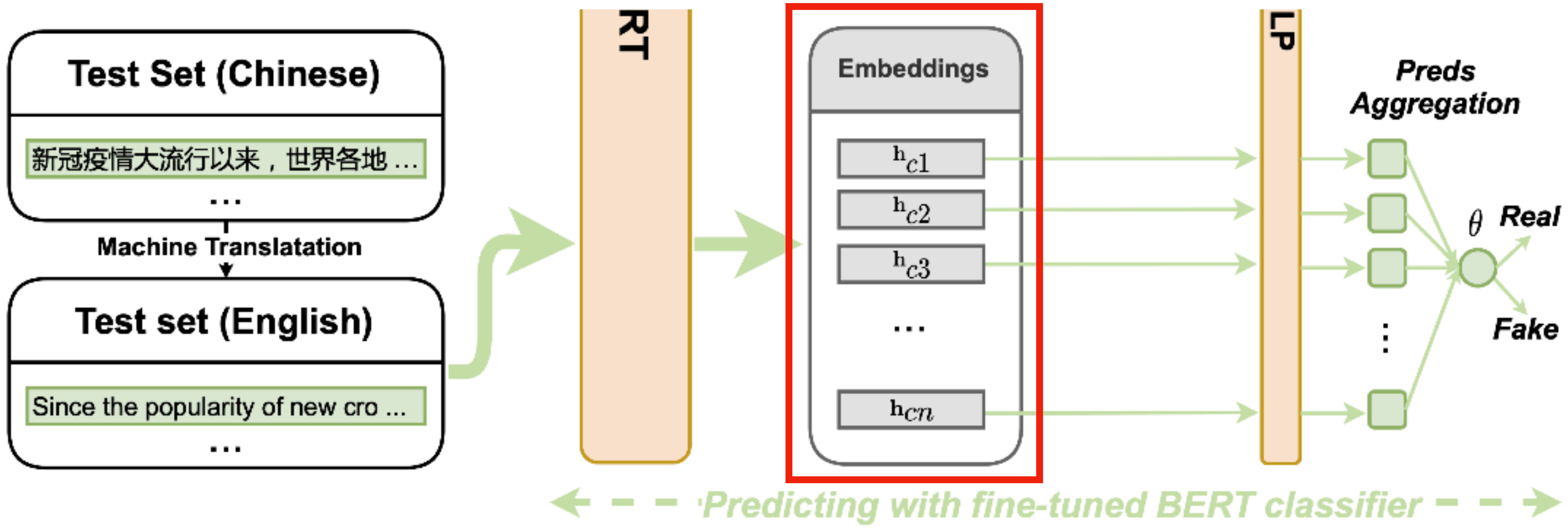
Methodology
Model Training



- Since data only has fake and real news, adopt the binary cross-entropy loss function to update the classifier C:
- $$\mathscr{L} = \sum_{e \in \mathscr{N}_e} -\log(y_e \cdot \mathrm{ReLU}(\mathrm{MLP}(\mathbf{h}_e)))$$
- The loss function is optimized using SGD, and it's equivalent to fine-tuning the pre-trained BERT encoder.

Methodology

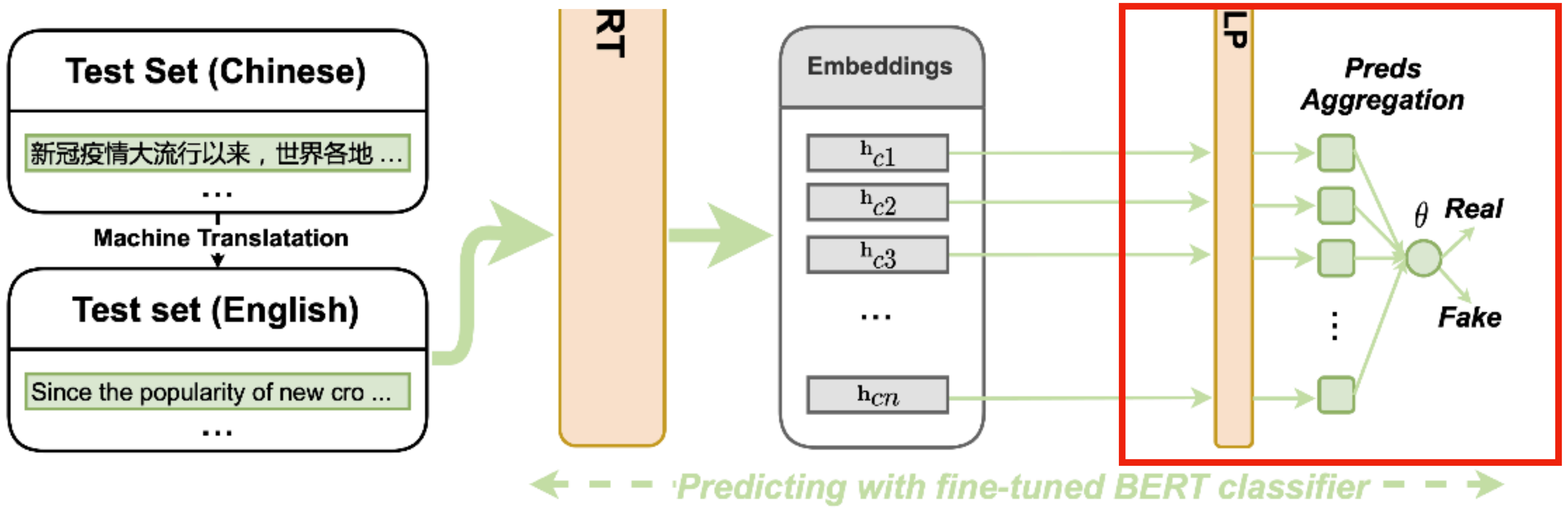
News Verification



- First translate the test data into English with Google Translator API to align the input data language of C.
- Compared baselines that encode the Chinese directly, and their performance is worse than proposed approach.
- The sub-texts set for news c is $TG_c = \{t_{c_1}, \cdots, t_{c_n}\}$ sequentially where $n = \left\lceil \frac{|T_e|}{100} \right\rceil$.
- Note that 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.

Methodology

News Verification



- The prediction result p_c for the test news c is obtained via the following equation:
- $$p_c = \begin{cases} 1, & \text{if } \dfrac{\sum_{j=1}^{n} |C(t_{cj})|}{n} \geq \theta \\ 0, & \text{if } \dfrac{\sum_{j=1}^{n} |C(t_{cj})|}{n} < \theta \end{cases}$$
- Aggregate the prediction results of all sub-texts for c and θ is a classification threshold empirically set to 0.8.

Experiments

Dataset

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

- Training dataset consists of all English COVID-19 news from three datasets:
    - ReCOVery, FakeCovid, and COAID.
- Evaluate the performance of proposed CrossFake on the Chinese COVID-19 news dataset collected in the paper.

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/

Experiments

Baselines - Monolingual

- CSI: employs an LSTM to encode the news content to detect fake news.
- SAFE: uses TextCNN to encode news textual information.
- exBAKE: utilizes the vanilla BERT as the English text encoder.
- Train the models above on the English training data and evaluate them on the translated test data.

Experiments

Baselines - Cross-lingual

- CLEF: leverages 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.
  - Only encodes news articles since our dataset does not include other data types.

Experiments

Results

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*	$75.00_{3.94}$	$71.45_{5.14}$	$70.47_{7.56}$	$70.67_{5.12}$

- 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.

Experiments
Results

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*	$75.00_{3.94}$	$71.45_{5.14}$	$70.47_{7.56}$	$70.67_{5.12}$

- Performance of CrossFake-avg benefits from the knowledge captured from long news articles.
  - 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.
- Observe that aggravating sub-text predictions could help alleviate the bias induced by a classifier with a higher reference for fake news.
  - exBAKE-sub has better accuracy and precision than exBAKE.

Experiments
Results

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*	$75.00_{3.94}$	$71.45_{5.14}$	$70.47_{7.56}$	$70.67_{5.12}$

- CNN-based model outperforms the RNN-based model.
- SAFE outperforms CSI significantly in terms of accuracy.
    - 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.
    - SAFE adopts TextCNN, extract the local critical information related to fact-checking.

Experiments
Results

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*	$75.00_{3.94}$	$71.45_{5.14}$	$70.47_{7.56}$	$70.67_{5.12}$

- Pre-trained multi-lingual models are ineffective.
- EMET & CLEF can encode Chinese news without translation.
- However, both of them perform poorly on the test set.

Experiments

Results

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*	$75.00_{3.94}$	$71.45_{5.14}$	$70.47_{7.56}$	$70.67_{5.12}$

- 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.
  - Might lack of the domain knowledge for the news events, and it's difficult to map the terms like "COVID-19" and their corresponding cross-lingual words to a similar space.
  - It suggests the the pre-trained language models should be kept up-to-date to handle the emerging events.

Experiments
Results

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*	$75.00_{3.94}$	$71.45_{5.14}$	$70.47_{7.56}$	$70.67_{5.12}$

- 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 dataset.
- Beside slicing text like CrossFake, more effective approaches are demanded to encode the long news text.

Discussion and Limitation

Translation Quality

- Machine translation quality is a bottleneck of cross-lingual tasks, especially in emerging events.
- Coronavirus is mistranslated as “new crown virus”, misleading the fake news classifier.
- Moreover, have attempted to translate all English training data to Chinese and train a Chinese fake news classifier.
  - Result of performance is bad since low translation quality harms the training data quality.

Discussion and Limitation

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, the annotated misinformation appear after a lengthy introduction, which exceeds the maximum sequence length most language models can process.
- 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.

Discussion and Limitation

Information Location

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

Conclusion and Future Work

- Make the first attempt to detect COVID-19 fake news under a cross-lingual setting.
- Collect and annotate a Chinese COVID-19 news dataset and proposed an end-to-end fake news detector CrossFake.
  - Trained on English news and could detect most of the collected Chinese fake news after translation.
- Experimental results demonstrate the advantage of encoding more news content and limitation of pre-trained multi-lingual encoders.
- Moreover, the event-centric analysis based on our data is another research direction.

Comments

of CrossFake

- A simple cross-lingual fake news detection task.
- Train on English corpus and classification on translated Chinese news.
- Brute slice text to sub-text groups.
- Mentioned SAFE, but only use the textual information.