iCompass@CLEF2022 CheckThat! Lab: Combining Deep Language Models for Fake News Detection*

Bilel Taboubi[1], Mohamed Aziz Ben Nessir[1], and Hatem Haddad[1][0000−0003−3599−7229]

iCompass, 49 rue de Marseille, Tunis, Tunisia
{bileltaboubi20,bennessir.mohamedaziz,haddad.hatem}@gmail.com

Abstract. Users of social media read, share, publish news with no prior knowledge if the news are real or fake. This necessitates the development of an automated system for fake news detection. In this paper we report a system and its output as a part of CLEF2022 - CheckThat! Lab Fighting the COVID-19 Infodemic and Fake News Detection. The task 3 was carried out using a variety of techniques. We achieve an F1 score of 34% on news classification for English dialect.

Keywords: Categorical Classification · fake news detection · BERT · RoBERTa

1 Introduction

Social media platforms has grown to unimaginable heights with a vast amount of information exponentially increasing which allows these platforms to be a host for plenty of unwanted, untruthful and misleading information that can be made and shared by anyone. As a result a category of people took advantage of it and started disseminating false information about people or entities, making negative impacts to individuals, business and society.The uncontrollable amount of information being shared can no more be covered by manually fact checking sites as a result an automated system to detect whether an information is real or fake is in need. In this paper, we have tackled Task 3: Fake News Detection CLEF2022-CheckThat!. The task required multi-class categorical classification of articles to determine the article claim is true, false, partially false or other (lack of evidence). This task is offered as a mono-lingual task in English and as cross-lingual task for English and German (English training data, German test data). The paper discusses the results obtained on the English test data with pre-trained transformers models and pre-processing techniques applied on the English train data for training.

* Supported by iCompass.

2 Tasks Definition:

Task 3 is a supervised categorical multi-class classification problem. Given the text of a news article, determine whether the main claim made in the article is true, partially true, false, or other.This task is offered as a mono-lingual task in English and as cross-lingual task for English and German.

CheckThat!2022 lab organizers defined the labels for the categories are as follows:

False - The claim made in an article is untrue.

Partially False - The given claim have weak evidence of the claim and cannot be considered as 100% true or false,

True - The claim is totally true.

Other- Articles that cannot be proven as false, true or partially true.

3 Literature Review

Internet became a main part in our daily life, our main source for information and news which can be fake or real. As a result, Fake news detection got wide attention in the NLP research community. In [1], authors conducted an exploratory study the propagation, authors and content of misinformation on tweets with COVID-19 content creating two dataset, the first one with 1500 tweets relating to 1274 false and 226 partially false claims collected from fact-checked claims related to COVID-19 by professional fact-checking organisations with different languages that got translated into English with google translator API. The second dataset containing a background corpus of 163,096 English tweets with purpose of understanding the misinformation around COVID-19. This study showed that false claims propagate faster than any other fake news category and even verified twitter accounts of celebrities and organizations are taking part in misinformation spread. In [2], authors created a multilingual dataset 'The FakeCovid' collected from 92 different fact-checking website with 5182 articles circulated in 105 countries where 40.8% of articles articles are in the English language, The dataset were manually annotated in in three language English, Hindi and German due a the limited languages knowledges. They applied BERT based without finetuning and with preprocessing techniques on the data such as abbreviations and contractions of words, spelling correction to achieve F1-score of 0.76 on English dataset. In [3], authors presented a semi-automatic framework 'AMUSED' to collect data from different networking sites such as Twitter, YouTube, and Reddit in different languages with the following steps, identify domain and data sources, scrap the web and detect language, extract social media links and crawl data from them, label the crawled data, human verification and finally merge the social media crawled data with the details from the news articles. They made a use case of COVID-19 Misinformation with the framework to collect 8,077 fact-checked news articles from 105 countries in 40 languages.In [4], authors presented overview of the CLEF-2021 CheckThat! Lab on Task 3 fake news detection where he described the task3A which is about

determining whether a claim is true, partially true, false, or other, and task3B which is about classifying an article to a topical domain (health, crime, climate, election, and education).Thus he described the provided data for each task and their collection and annotation steps, the participants team and their solution. There were 27 teams for Task 3A, The best performing system for task 3A was obtained by NoFake team and achieved a macro F1-score of 0.84 and was ahead of the rest by a rather large margin, they applied BERT base and trained it with an additional data from different fact-checking websites. For task 3B there were 20 teams and the best system was made by NITK_NLP [5] achieving 0.88 marco F1 score with an ensemble of three transformers models.

4 Data Description

The dataset provided dataset contains about 1264 articles in English (title and text) with the respective label (true, partially true, false, or other) divided into training and development sets, 900 rows for train set and 394 rows for the dev set. Table 1 show a sample of the dataset for task 3 and table 2 shows the distribution of the dataset according to their respective classes.

public_id	**text**	**title**	**rating**
1145ea7c	U.S. military officials worked to ensure President Trump wouldn't see the warship that bears the name of the late senator, a frequent target ...	The Texas State Senate – Senator Paul Bettencourt: District 7	true
2d06d27c	A 2,500-strong border and coastguard corps could see armed personnel sent to Greece. The island of Lesbos has been deluged with migrants The European Union's ...	EU army to protect borders	false

Table 1. Samples of Task 3 dataset

rating	**occurence**
True	211
False	578
Partially False	358
Other	117

Table 2. Data distribution of task 3

5 Data preparation

For the data pre-processing we applied various techniques, such as applying lowercase, lemmatization returning words to root form, English stopwords removal such as “are”, “the”, “is” and etc, punctuation removal using NLTK library. The dataset contained null values for texts and titles. In order to make it more manageable null values for titles were replaced by their texts and null values for texts were replaced with their titles.

6 Proposed Methodology

This paper introduced two concatenated parallel BERT models for classifying *whether the news are real or fake. The process of predicting whether the news is true, false, partially false or other type by using this model 1

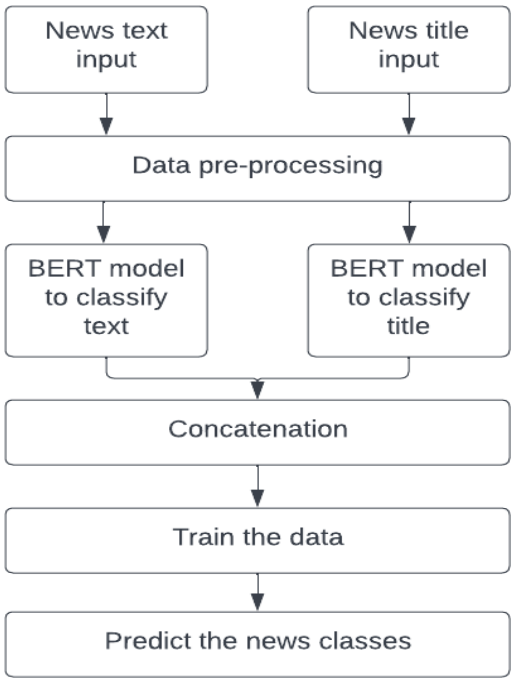

Fig. 1. News classes prediction steps

7 Pre-trained Models

Different pre-trained models were used in order to achieve the best results when fine-tuning it in a multi-task fashion.

BERT base Uncased

BERT [6] is a trained Transformer Encoder stack that uses bidirectional self-attention. The BERT's model architecture have many encoder layers (also called Transformer Blocks) twelve in the Base version. Thus, it has larger feedforward networks (768 hidden units) and 12 attention heads.The model is trained on unlabeled data over different pre-training tasks. For finetuning, the BERT model is initialized with the pre-trained parameters and can be used directly.The model initial parameters change by training it by labeled data from the downstream tasks such as Masked LM, Next Sentence Prediction.

RoBERTa

The self-supervised transformer model RoBERTa [7] was trained on a enormous corpus of English data containing five English-language corpora of varying sizes and domains, totaling over 160GB of uncompressed text. Self-supervised means it was pre-trained on raw texts with no human annotation, and then utilized an automated way to generate inputs and labels from those texts. RoBERTa model achieves state-of-the-art results on GLUE (The General Language Understanding Evaluation), RACE (The ReAding Comprehension from Examinations) and SQuAD (The Stanford Question Answering Dataset).

8 Results

Pre-trained models BERT base uncased and RoBERTa were trained and fine-tuned with the following architecture: the model is a multi-input, a concatenation between 2 sub model just before the classification layer where the first is taking text input followed by embedding layer which will contain a BERT model , Gated recurrent network layer with 128 units and 0.3 dropout rate, global max pooling and a dropout layer. The second sub model consisted of an input layer, Embedding layer which will contain a second BERT model, Global max pooling layer and a dropout layer. The average training time of a model is around 8 minutes. Best results achieved by each pre-trained model is presented in the table 3 where they got trained on the train set, tested with dev set.

RoBERTa was pre-trained on a bigger vocabulary than BERT base uncased but still getting outperformed and that is due the limited resources available to train the models, we couldn't increase the batch size and sequence length where we was able to train RoBERTa for only 10 batch size and 128 sequence length.

The submitted model was BERT base uncased, trained with a 10 epochs, 2e-5 learning rate for Adam optimizer, a sequence length of 128, 22 batch size

Type	F1	Accuracy	Precision	Recall
BERT base uncased	0.513	0.511	0.555	0.511
RoBERTa	0.227	0.237	0.220	0.237

Table 3. Task 1A Pre-trained models results on test set.

and categorical cross entropy loss function.The model achieved F1_score 0.513 on the dev set.

Our model for task 3 achieved interesting results on English test set and we were placed first in the ranks for task 3 with 34% F1-weighted among 25 participants as shown in table 4.

Team	Accuracy	F1-Score
iCompass	0.5473856209150327	0.33913726061970056
nlpiruned	0.5408496732026143	0.3324961059439111
awakened	0.5310457516339869	0.323094873671759

Table 4. Top 3 on Task 3 English leaderboard

The low F1_weighted score can be explained with the categories 'other' and 'partially false', since these classes presented low precision and recall scores as shown in the table. 5

iCompass	precision	recall	F1-Score
false	0.6359223300970874	0.8317460317460318	0.720770288858322
other	0.10526315789473684	0.06451612903225806	0.07999999999999999
partially false	0.14457831325301204	0.21428571428571427	0.1726618705035971
true	0.6020408163265306	0.28095238095238095	0.38311688311688313

Table 5. Classification report on the test set

9 Conclusion

In this paper, we analysed pre-trained models BERT base uncased and RoBERTa in order to obtain the best F1 for fake news classification on the English dataset, stopwards removal, lemmatization, etc. are used as data preprocessing tasks for eliminating less important words for training. Our model attained 34% F1_weighted which is not unsatisfactory and that is due the data low distribution specially

for the categories 'other' and 'partially false'.In future, we will explore augmentation and resembling strategies to create a large balanced dataset for training and validating our proposed model and try to overcome our limitations.

References

1. Gautam KishoreShahia, AnneDirksonb, Tim A.Majchrzakc: An exploratory study of COVID-19 misinformation on Twitter. Online Social Networks and Media, Volume 22, pages 100104 (2021) https://doi.org/https://doi.org/10.1016/j.osnem.2020.100104.
2. Gautam Kishore Shahi and Durgesh Nandini: FakeCovid- A Multilingual Cross-domain Fact Check News Dataset for COVID-19, Book title Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media (2020) https://doi.org/https://arxiv.org/abs/2006.11343v1
3. Shahi, Gautam Kishore: AMUSED: An Annotation Framework of Multi-modal Social Media Data, (2020), https://doi.org/https://arxiv.org/abs/2010.00502
4. Shahi, Gautam Kishore and Struß, Julia Maria and Mandl, Thomas: Overview of the CLEF-2021 CheckThat! lab task 3 on fake news detection, Working Notes of CLEF (2021), https://doi.org/https://arxiv.org/abs/2010.00502
5. Hariharan RamakrishnaIyer LekshmiAmmal, Anand Kumar Madasamy: NITK_NLP at CheckThat! 2021: Ensemble Transformer Model for Fake News Classification (2021), https://doi.org/http://ceur-ws.org/Vol-2936/paper-49.pdf
6. Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (2018), https://doi.org/ arXiv:1810.04805
7. Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov: RoBERTa: A Robustly Optimized BERT Pretraining Approach (2019), https://doi.org/arXiv:1907.11692