

iCompass at CheckThat! 2022: Combining Deep Language Models for Fake News Detection

Bilel Taboubi¹, Mohamed Aziz Ben Nessir¹ and Hatem Haddad¹

¹*iCompass, Emeraude Palace, Rue du Lac Windermère, Les Berges du Lac, Tunis 1053*

Abstract

Users of social media tend to explore different platforms to obtain news and find information about different events and activities. Furthermore they read, share, publish news with no prior knowledge of the certainty of being 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 part of CLEF2022 - CheckThat! Lab Fighting the COVID-19 Infodemic and Fake News Detection. Task 3 was carried out using two BERT base uncased and data preprocessing with stop-words removal, lemmatization. We achieve an F1 score of 0.339 on news classification on English dataset.

Keywords

Categorical Classification, fake news detection, BERT, RoBERTa

1. Introduction

Social media platforms have grown to unimaginable heights with a vast amount of information exponentially increasing. This information flow increase allows social media 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 amount of information being shared is uncontrollable and cannot be totally 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! [1]. The task required multi-class 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) [2]. The paper discusses the results obtained on the English dataset with pre-trained transformers models and pre-processing techniques applied.

CLEF 2022: Conference and Labs of the Evaluation Forum, September 5–8, 2022, Bologna, Italy

✉ bileltaboubi20@gmail.com (B. Taboubi); mohamedaziz.bennessir@etudiant-isi.utm.tn (M. A. B. Nessir); haddad.hatem@gmail.com (H. Haddad)

🆔 0000-0003-3599-7229 (H. Haddad)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

2. Tasks Definition

Task 3 is a multi-class classification problem. Given the text and the title 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 [3] defined the labels for the categories 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 and social media platforms became a main part in our daily life, our main source for information as well as for misinformation. With the huge increase of false information social media needs to curb the spreading of misinformation through their platform. As a result, fake news detection got wide attention in the NLP research community. In [4], authors conducted an exploratory study of COVID-19 misinformation on Twitter. They created two dataset, the first one contains 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. The non-English tweets got translated into English language with the use of google translator API. The second dataset contains a 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 [5] Das et al. proposed an Ensemble model for COVID-19 fake news detection for the Constraint COVID19 shared task [6]. They combined pre-trained models with a heuristic algorithm based on the username handle and link-domain in tweets. In [7], authors created a multilingual dataset 'The FakeCovid' collected from 92 different fact-checking websites with 5182 articles circulated in 105 countries where 40.8% of articles are written in the English language, the dataset was manually annotated in in three languages English, Hindi and German due to limited knowledge about other different languages. 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 [8], 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 [9], authors presented overview of the CLEF-2021 CheckThat! Lab on Task 3 fake news detection where

they described Task 3A, which is about determining whether a claim is true, partially true, false, or other, and Task 3B, which is about classifying an article to a topical domain (health, crime, climate, election, and education). Thus they 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 additional data from different fact-checking websites. For Task 3B there were 20 teams and the best system was made by NITK_NLP [10] achieving 0.88 marco F1 score with an ensemble of three transformers models.

4. Data Preparation

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

Table 1
Samples of Task 3 dataset

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 2
Data distribution of task 3

rating	occurence
True	211
False	578
Partially False	358
Other	117

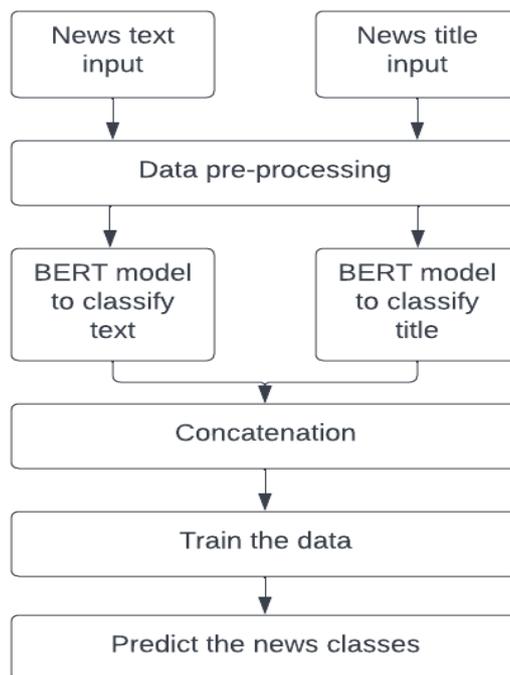
For the data pre-processing we applied various techniques, such as applying lowercase, lemmatization, English stopwords removal such as “are”, “the”, “is” and etc, punctuation removal using NLTK [11] library. The dataset contained null values for texts and titles. In order to make it more manageable null values for titles were replaced by their texts and null values for texts were replaced with their titles.

5. Approach

This paper introduced two concatenated parallel BERT models for classifying whether the news are real or fake. The process of predicting the news different categories true, false, partially false or other is done by using the following architecture:

- Title input layer, Text input layer.
- BERT model for text input followed by a gated recurrent unit with 128 units and 0.3 probability for dropout and a dropout layer with 0.1 probability.
- BERT model for title input following by global max pooling layer and dropout layer with 0.1 probability.
- Concatenation layer to concatenate the output of the BERT models.
- Dense layer with softmax activation function and four units.

Figure 1: News classes prediction steps



As presented in figure 1 the model takes as input each of text and title after getting pre-processed. The inputs layers passes the data to BERT models one for the text input and the other for title input. Before predicting the classes, the output of each BERT model get concatenated.

6. Pre-trained Models

In order to achieve the best results different pre-trained models were used, combined and fine-tuned with different hyperparameters for the multi-class classification task.

BERT base Uncased

BERT [12] is a trained Transformer Encoder stack that uses bidirectional self-attention. The BERT's model architecture is composed by multiple 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 [13] 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).

7. Results & Discussion

Pre-trained models BERT base uncased and RoBERTa were trained and finetuned 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.

Table 3

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

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

RoBERTa was pre-trained on a bigger vocabulary than BERT base uncased but it got outperformed and that is due to the limited resources available to train the models, the batch size and sequence length was limited where we was unable to exceed 10 batch size and 128 sequence length while training RoBERTa model with the proposed architecture.

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 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 task 3 ranking leaderboard with 0.339 macro F1 measure among 25 participants as shown in table 4.

Table 4
Top 3 on Task 3 English leaderboard

Team	Accuracy	F1-Score
iCompass	0.547	0.339
nlpiruned	0.541	0.332
awakened	0.531	0.323

The low macro F1 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.

Table 5
Classification report on the test set

iCompass	precision	recall	F1-Score
false	0.636	0.832	0.721
other	0.105	0.065	0.079
partially false	0.145	0.214	0.173
true	0.602	0.281	0.383

8. Conclusion

In this paper, we analysed pre-trained models BERT base uncased and RoBERTa. In order to obtain the best macro F1 for fake news classification on the English dataset different pre-processing techniques were used such as stopwords removal, lemmatization, etc. with the purpose of reducing irrelevant words from the text and title for training. Our model attained 0.339 macro F1 measure which is unsatisfactory and that is due to the data low distribution specially for the categories ‘other’ and ‘partially false’. In future, we will explore augmentation and resampling strategies to create a large balanced dataset for training and validating our proposed model and try to overcome our limitations.

References

- [1] J. Köhler, G. K. Shahi, J. M. Struß, M. Wiegand, M. Siegel, T. Mandl, Overview of the CLEF-2022 CheckThat! lab task 3 on fake news detection, in: Working Notes of CLEF 2022—Conference and Labs of the Evaluation Forum, CLEF '2022, Bologna, Italy, 2022.
- [2] P. Nakov, A. Barrón-Cedeño, G. Da San Martino, F. Alam, J. M. Struß, T. Mandl, R. Míguez, T. Caselli, M. Kutlu, W. Zaghoulani, C. Li, S. Shaar, G. K. Shahi, H. Mubarak, A. Nikolov, N. Babulkov, Y. S. Kartal, J. Beltrán, The clef-2022 checkthat! lab on fighting the covid-19 infodemic and fake news detection, in: M. Hagen, S. Verberne, C. Macdonald, C. Seifert, K. Balog, K. Nørsvåg, V. Setty (Eds.), *Advances in Information Retrieval*, Springer International Publishing, Cham, 2022, pp. 416–428.
- [3] P. Nakov, A. Barrón-Cedeño, G. Da San Martino, F. Alam, J. M. Struß, T. Mandl, R. Míguez, T. Caselli, M. Kutlu, W. Zaghoulani, C. Li, S. Shaar, G. K. Shahi, H. Mubarak, A. Nikolov, N. Babulkov, Y. S. Kartal, J. Beltrán, M. Wiegand, M. Siegel, J. Köhler, Overview of the CLEF-2022 CheckThat! lab on fighting the COVID-19 infodemic and fake news detection, in: *Proceedings of the 13th International Conference of the CLEF Association: Information Access Evaluation meets Multilinguality, Multimodality, and Visualization*, CLEF '2022, Bologna, Italy, 2022.
- [4] G. K. Shahi, A. Dirkson, T. A. Majchrzak, An exploratory study of covid-19 misinformation on twitter, *Online Social Networks and Media* 22 (2021) 100104. URL: <https://www.sciencedirect.com/science/article/pii/S2468696420300458>. doi:<https://doi.org/10.1016/j.osnem.2020.100104>.
- [5] S. D. Das, A. Basak, S. Dutta, A heuristic-driven ensemble framework for covid-19 fake news detection, 2021. URL: <https://arxiv.org/abs/2101.03545>. doi:10.48550/ARXIV.2101.03545.
- [6] P. Patwa, S. Sharma, S. Pykl, V. Guptha, G. Kumari, M. S. Akhtar, A. Ekbal, A. Das, T. Chakraborty, Fighting an infodemic: COVID-19 fake news dataset, in: *Combating Online Hostile Posts in Regional Languages during Emergency Situation*, Springer International Publishing, 2021, pp. 21–29. URL: https://doi.org/10.1007/978-3-030-73696-5_3. doi:10.1007/978-3-030-73696-5_3.
- [7] G. K. Shahi, D. Nandini, FakeCovid- A Multilingual Cross-domain Fact Check News Dataset for COVID-19, ICWSM, 2020. URL: <https://doi.org/10.36190/2020.14>. doi:10.36190/2020.14.
- [8] G. K. Shahi, Amused: An annotation framework of multi-modal social media data, 2020. URL: <https://arxiv.org/abs/2010.00502>. doi:10.48550/ARXIV.2010.00502.
- [9] J. M. S. Gautam Kishore Shahi, T. Mandl, Overview of the clef-2021 checkthat! lab: Task 3 on fake news detection (2021). URL: <http://ceur-ws.org/Vol-2936/paper-30.pdf>.
- [10] A. K. M. Hariharan RamakrishnaIyer LekshmiAmmal, Overview of the clef-2021 checkthat! lab: Task 3 on fake news detection (2021). URL: <http://ceur-ws.org/Vol-2936/paper-49.pdf>.
- [11] S. Bird, E. Klein, E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit*, " O'Reilly Media, Inc.", 2009.
- [12] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2018. URL: <https://arxiv.org/abs/1810.04805>. doi:10.48550/ARXIV.1810.04805.

- [13] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, 2019. URL: <https://arxiv.org/abs/1907.11692>. doi:10.48550/ARXIV.1907.11692.