

Awakened at CheckThat! 2022: Fake News Detection using BiLSTM and Sentence Transformer

Ciprian-Octavian Truică^{1,2}, Elena-Simona Apostol^{1,2} and Adrian Paschke³

¹Department of Information Technology, Uppsala University, Lägerhyddsvägen 1, Uppsala, 75105, Sweden

²Computer Science and Engineering Department, Faculty of Automatic Control and Computers, University Politehnica of Bucharest, Splaiul Independenței 313, Bucharest, 060042, Romania

³Fraunhofer Institute for Open Communication Systems, Berlin, 10589, Germany

Abstract

In recent years, online social networks and online news venues have become some of the main news and event-related information spreading mediums. Although using these mediums has facilitated the speed of accessing information, it also created a new phenomenon used for propaganda and disinformation: fake news. As fake news has detrimental consequences to society, new technologies need to be developed in order to stop their harmful effects. In this paper, we propose two Bidirectional Long Short-Term Memory (BiLSTM) architectures with sentence transformers to solve two tasks: (1) a multi-class mono-lingual fake news detection task (i.e., mono-lingual task); and (2) a multi-class cross-lingual fake news detection task (i.e., cross-lingual task). For the mono-lingual task, we train and test a BiLSTM with BART sentence transformers model on an English dataset and obtain an accuracy of ~ 0.53 and an F1-Score of ~ 0.32 . For the cross-lingual task, we train a BiLSTM with XLM sentence transformers model on an English dataset and test the model using transfer learning on a German dataset. For this task, we obtain an accuracy of ~ 0.28 and an F1-Score of ~ 0.19 .

Keywords

Fake News Detection, Neural Networks, Sentence Transformers, Transfer Learning

1. Introduction

With the digital age, new mass media paradigms for information distribution have been adopted by the general public. The current paradigms have shifted from the journalistic rigorous imposed by editors to personalized social media where anyone can spread event related news. This new approach aggravates the risk of fake news [1, 2], which has detrimental consequences to society by facilitating the spread of misinformation in the form of fake news, propaganda, conspiracy theories, political bias, etc. The practices of spreading misinformation online by malicious actors need to be tackled from different points of view, i.e., from a journalistic and fact-checking perspective to a more technological-based one.

In this paper, we address the problem of detecting fake news from a technological perspective by using the CheckThat! 2022: Fake News Detection Challenge datasets. To tackle the problem


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


✉ ciprian-octavian.truica@it.uu.se; ciprian.truica@upb.ro (C. Truică);

elena-simona.apostol@it.uu.se; elena.apostol@upb.ro (E. Apostol); adrian.paschke@fokus.fraunhofer.de

(A. Paschke)

ORCID 0000-0001-7292-4462 (C. Truică); 0000-0001-6397-4951 (E. Apostol); 0000-0003-3156-9040 (A. Paschke)

 © 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)

we propose two neural network with sentence transformer models for (1) multi-class mono-lingual fake news detection; and (2) multi-class cross-lingual fake news detection. We use transfer learning in order to train a model on an English dataset and test it on German text.

In this work, we aim to answer the following two research questions:

- (Q_1) Does a simple neural network with sentence transformers offer good results for multi-class mono-lingual fake news detection?
- (Q_2) Can cross-lingual sentence transforms be used through transfer learning in multi-class cross-lingual fake news detection?

To answer question (Q_1), we propose the use of a Bidirectional Long Short-Term Memory (BiLSTM) neural network with BART sentence transformers. While to answer question (Q_2), we train a BiLSTM neural network with XLM sentence transformers on English textual data and use transfer learning to solve a cross-lingual fake news detection task by testing the model on German textual data.

This paper is structured as follows. In Section 2, we discuss some of the current literature on fake news detection. In Section 3, we present our approach for mono-lingual and cross-lingual fake news detection. In Section 4, we present the datasets, experimental setup, and results. Finally, in Section 5, we summarize our findings and hint at future work.

2. Related Work

The task of fake news detection has been examined from various perspectives and using various models from traditional Machine Learning to more elaborate yet more powerful Neural Network based models. Extensive work in the field of fake news detection has led to many solutions focusing on either model or data-driven approaches. Among the traditional Machine Learning models (e.g., Support Vector Machine, Logistic Regression, Decision Trees, AdaBoost, Naïve Bayes), the model that performs very good in many cases is Multinomial Naïve Bayes [3].

Several current solutions use complex Deep Neural Network architectures for this task. Many solutions for multi-class mono-lingual fake news detection show promising results when using Convolutional Neural Network (CNN) based architectures. FNDNet [4] is such an architecture that obtains good results in comparison even with recurrent networks, i.e., LSTM. OPCNN-FAKE [5] is an optimized CNN based solution that uses a hyperopt optimization technique to adapt the values of parameters for each component layer in order to achieve high performance. Other Deep Learning solutions focus on recurrent networks, e.g., (Bi)GRU, (Bi)GRU, (Bi)LSTM, obtaining the best results when also using attention mechanisms [1]. BiLSTM based solutions (e.g., Samantaray and Kumar [6], Trueman et al. [7]) are very promising as this type of recurrent network is able to capture both past and future information.

In multi-class classification, the employed embedding model is very important. As shown in Ilie et al. [1], many Deep Learning models have an increase in accuracy when using custom trained word embeddings versus pre-trained ones. Other models use advanced pre-trained transformers instead of the more classical word embeddings. Different transformer models can be applied for fake news detection, e.g., BERT (Bidirectional Encoder Representations from Transformers) [8], RoBERTa (A Robustly Optimized BERT pre-training Approach) [9], BART

(Bidirectional and Autoregressive Transformer) [10]. As such, MisRoBÆRTa [2] is a complex architecture that combines BART and RoBERTa for a multi-class classification task. Another solution to the multi-class classification problem is proposed by Liu et al. [11] that offers a two-stage BERT-based model.

3. Methodology

In this section, we present the methodology used for fake news detection. For encoding the text, we used two sentence transformer [12] approaches. (1) For multi-class fake news detection of news articles in English, we use BART (Bidirectional and Auto-Regressive Transformers) [10] sentence transformers. (2) For cross-lingual news articles, we use XLM (Cross-Lingual Language Model) [13] sentence transformers. We employed a BiLSTM (Bidirectional Long Short-Terms Memory) as the classification model.

3.1. Sentence Transformers

Sentence transformers are a modification of the pre-trained BERT networks that use siamese and triplet network structures to derive semantically meaningful sentence embeddings that can be compared using cosine-similarity [12]. We construct BART [10] and XLM [13] sentence transformers for the mono-lingual and cross-lingual classification tasks, respectively.

XLM [13] is a Transformer architecture that uses two approaches during pre-training depending on the type of data. For mono-lingual data, it uses an unsupervised modeling technique such as Casual Language Modeling (CLM) or Masked Language Modeling (MLM). For cross-lingual data, XLM employs a supervised modeling technique that combines MLM with Translation Language Modeling (TLM).

BART [10] is a generalized BERT that uses a transformer-based neural machine translation architecture. The architecture uses a left-to-right decoder (as in GPT [14] architecture) and a standard Sequence-to-Sequence bidirectional encoder (as in BERT [8]).

3.2. Classification Models

For classification, we propose a deep neural network architecture that contains the following layers: (1) Input layer; (2) BiLSTM layer; and (3) Dense layer.

The input layer instantiates the neural network. It is used to produce a symbolic tensor-like object that has the size of the sentence transformer.

LSTM (Long Short-Term Memory) [15] is a recurrent neural network that process *past* information using two state components: (1) a hidden layer for the short-term memory; and (2) an internal cell state for long-term memory. The BiLSTM layer encapsulates both *past* and *future* information through the use of two hidden states. The forward hidden state processes the *past* information using a forward LSTM, while the backwards hidden state process the *future* information provided by employing a backward LSTM. To encode both the *past* and *future*, the BiLSTM concatenate into on hidden state the forward and backward hidden state at every time-step. The number of units for this layer can be determined experimentally using ablation and hyperparameter testing (see [2] for more details). For the LSTM cell, we use the classic

implementation presented in [15, 16]. For this cell, the recurrent activation function is *sigmoid*, the kernel weights are initialized using the Glorot uniform linear transformation [17], and the bias vector is initialized with zeros.

The Dense Layer is a fully connected Perceptron layer used for classification. The number of units in this layer is equal to the number of classes. The activation function for this layer is the *sigmoid*.

4. Experimental Results

In this section, we present the experimental results of the proposed models for the CheckThat! 2022 Fake News Detection task for both the mono-lingual and cross-lingual challenges.

4.1. Dataset

The CheckThat! 2022 Task 3 [18, 19] consists of two subtasks as follows: (1) multi-class mono-lingual fake news detection of news articles (English) [20, 21] (i.e., mono-lingual task); and (2) multi-class cross-lingual fake news detection task (German) (i.e., cross-lingual task). The steps used in the data collection are defined in Shahi [22].

For the mono-lingual task, the English training data is the same as from the CheckThat! 2021 version [23]. The number of classes for this task is four: *false*, *partially false*, *other*, and *true*. The number of labels has been defined after a thorough study of 83 classes was conducted by fact-checkers [24]. The dataset contains an English training, development, and testing set.

For the cross-lingual task, a new test dataset in German is introduced. The main focus of this task is to use transfer learning to detect fake news content in low resource languages. Thus, the training for this task is done on the English training and development datasets, and then it is tested on the German testing set. For this task, we use the same labels as for the mono-lingual task.

As we used the same training data for both the mono-lingual and cross-lingual tasks, we concatenated the English training and development sets to train the model. Table 1 shows the label distribution for the training dataset. We observe that the dataset is highly imbalanced. Table 2 presents the label distribution for the test dataset.

Table 1
Train dataset statistics

Label	No. Documents	Percentage
false	578	45.73%
partially false	358	28.32%
true	211	16.69%
other	117	9.26%
Total	1 264	100.00%

Table 2

Test dataset statistics

Label	No. Documents English	No. Documents German
false	315	191
partially false	56	97
true	210	243
other	31	55
Total	612	586

4.2. Experimental Setup

For the sentence transformer, we used the pre-trained BART (*facebook/bart-large*) and XLM (*sentence-transformers/stsb-xlm-r-multilingual*) from HuggingFace Transformer [25]. We train the sentence transformers using the SentenceTransformers Python 3 package [12].

For classification, the BiLSTM layer uses 100 LSTM units configured as in [15]. The dense layer contains 4 units (equal to the number of classes), and the sigmoid function as activation. We used the ADAM optimizer and a 64 batch size. The model is trained for 100 epochs. To prevent overfitting, we used an early stopping mechanism that monitors the Accuracy during training. We use Keras with TensorFlow as backend for implementing the neural model.

The implementation is available online on GitHub at the following url: <https://github.com/elena-apostol/AwakenedCheckThat2022>.

4.3. Results

Table 3 presents the overall results for both the mono-lingual (i.e., both train and test sets use English texts) and cross-lingual (i.e., the train set is in English and the test set is in German). We observe that, by training the BiLSTM with BART sentence embeddings on English, we obtain an accuracy of ~ 0.53 for the mono-lingual task and an accuracy of ~ 0.28 for the cross-lingual task. With these results, the Awakened team obtained the 3rd and the 5th place in the competition for the mono-lingual task and cross-lingual task, respectively. As a general observation, the results are highly influenced by the dataset’s size and class imbalance.

Table 3

Overall results for the two tasks

Task	Dataset	Sentence Transformer	Accuracy	F1-Score
Mono-lingual	English	BART	0.531045	0.323094
Cross-lingual	German	XML	0.283276	0.185991

For the mono-lingual task, the low performance of the model is directly impacted by two dataset related aspects: (1) the dataset size is small, being inadequate for a neural network approach; and (2) the dataset is highly imbalanced, miss-classification being a real challenge. When analyzing the evaluation metrics per class (Table 4), these two aspects are even more emphasized by the results obtained per class. For the *false* label, the model obtains ~ 0.67

precision and ~ 0.83 recall. Thus, the interpretation of these results shows that the models manage to correctly determine fake news. For the *true* label, the model obtains ~ 0.76 precision and ~ 0.21 recall. This shows that the model also manages to discriminate well between true news and the other 3 types of texts. Also, these results show that the contextual, semantic, and syntactic information encoded by the sentence transformer for the *true* and *false* labels are very specific to these classes. Thus, the textual dissimilarities between *fake news* and *real news* are more prominent. The precision and recall for the *partially true* (~ 0.13 precision and ~ 0.32 recall) and *other* (~ 0.04 precision and ~ 0.03 recall) classes are very small. These results indicate that these textual data are more similar to the other two classes. Thus, the model does not manage to discriminate correctly between these two labels and the *true* and *false* ones.

Table 4

Detailed results for the multi-class mono-lingual task using English

Class	Precision	Recall	F1-score
false	0.673521	0.831746	0.744318
partially false	0.129496	0.321428	0.184615
true	0.758620	0.209523	0.328358
other	0.038461	0.032258	0.035087

For the cross-lingual task, we observe that the model manages to obtain an accuracy of ~ 0.28 and an F1-Score of ~ 0.19 (Table 3). These results are also impacted by the transfer learning algorithm besides the dataset’s size and the imbalanced labels. Based on these observations, we can conclude that the multi-lingual sentence transformers do not manage to correctly find similarities between the English and German texts that are labeled with the same class. When analyzing the per class results, we observe that the *true* labeled German documents are predicted with a high precision (~ 0.59), but the recall for these labeled documents is ~ 0.05 . The interpretation of these values for the true label is that the model manages to determine the true positives more accurately than false negatives. In other words, for the *true* labeled documents, the model manages to return more relevant results to this label than irrelevant ones but does not manage to return most of the relevant results for this label. For the *false* labeled German documents, the interpretation of the results is, as expected, in reverse as for the *true* labeled documents. Thus, with a precision of ~ 0.35 and a recall of 0.65 , the model manages to return most of the relevant results for this label but does not manage to return all relevant results to this label. The model does not manage to classify any of the German documents in the test set labeled with *other* and it has a very low F1-Score for the prediction of documents labeled with *partially true*.

5. Conclusions

In this paper, we trained two BiLSTM neural networks with sentence transformers for data encoding models to detect the veracity of fake news. The first model is trained and tested on an English news article dataset for multi-class mono-lingual fake news detection. This model encodes the textual data using BART sentence transformer. The second model is trained on

Table 5

Detailed results for the multi-class cross-lingual task using German

Class	Precision	Recall	F1-score
false	0.345303	0.654450	0.452079
partially false	0.145833	0.288659	0.193771
true	0.590909	0.053497	0.098113
other	0.000000	0.000000	0.000000

the same English dataset and tested on a German dataset for multi-class cross-lingual fake news detection. The model encodes the textual data using XLM sentence transformer and takes advantage of transfer learning to solve the task of cross-lingual fake news detection. We use the first model to answer our first research question (Q_1). With an accuracy of ~ 0.53 and a F1-Score of ~ 0.32 , we conclude that it is worth investigating more the use of simple neural networks with sentence transformers for mono-lingual fake news detection task. The second model is used to answer our second research question (Q_2). With an accuracy of ~ 0.28 and a F1-Score of ~ 0.19 , we conclude that the BiLSTM XML sentence transform model does not manage to correctly find similarities between the English and German texts. Although, the use of cross-lingual transformers and transfer learning for multi-class classification in theory could prove useful, for the multi-class cross-lingual fake news detection task at hand, they perform poorly. For both mono-lingual and cross-lingual tasks, we observed that: (1) the dataset size needs to be large to be adequate for a neural network approach; and (2) the dataset needs a balanced label distribution to mitigate against miss-classification.

In future work, we aim to use transformer embeddings instead of sentence transformers. We also plan to test other cross-lingual transformers for transfer learning in a larger study to determine if the conclusions obtained on this small dataset generalize or are obtained by this data-driven approach.

Acknowledgments

The research presented in this paper was supported in part by the German Academic Exchange Service (DAAD) through the projects "AWAKEN: content-Aware and netWork-Aware faKE News mitigation" (grant no. 91809005) "Deep-Learning Anomaly Detection for Human and Automated Users Behavior" (grant no. 91809358), in part by the German Federal Ministry of Education and Research (BMBF) project "PANQURA - a technology platform for more information transparency in times of crisis" under Grant 03COV03F, in part by the European Union project "FAST-LISA - Fighting hAte Speech Through a Legal, ICT and Sociolinguistic approach" under Grant 101049342, and in part by the EU CEF project "NORDIS - Nordic observatory for digital media and information DISorder" under Grant number2394203).

References

- [1] V.-I. Ilie, C.-O. Truică, E.-S. Apostol, A. Paschke, Context-Aware Misinformation Detection: A Benchmark of Deep Learning Architectures Using Word Embeddings, *IEEE Access* 9 (2021) 162122–162146. doi:10.1109/ACCESS.2021.3132502.
- [2] C.-O. Truică, E.-S. Apostol, MisRoBERTa: Transformers versus Misinformation, *Mathematics* 10 (2022) 1–25(569). doi:10.3390/math10040569.
- [3] M. Singh, M. W. Bhatt, H. S. Bedi, U. Mishra, Performance of Bernoulli’s Naïve Bayes classifier in the detection of fake news, *Materials Today: Proceedings* (2020).
- [4] R. K. Kaliyar, A. Goswami, P. Narang, S. Sinha, FNDNet – A deep convolutional neural network for fake news detection, *Cognitive Systems Research* 61 (2020) 32–44. doi:10.1016/j.cogsys.2019.12.005.
- [5] H. Saleh, A. Alharbi, S. H. Alsamhi, OPCNN-FAKE: Optimized convolutional neural network for fake news detection, *IEEE Access* 9 (2021) 129471–129489.
- [6] S. Samantaray, A. Kumar, Bi-directional Long Short-Term Memory Network for Fake News Detection from Social Media, in: *Intelligent and Cloud Computing*, Springer, 2022, pp. 463–470.
- [7] T. E. Trueman, A. Kumar, P. Narayanasamy, J. Vidya, Attention-based C-BiLSTM for fake news detection, *Applied Soft Computing* 110 (2021) 107600. doi:10.1016/j.asoc.2021.107600.
- [8] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in: *Conference of the North American Chapter of the Association for Computational Linguistics, ACL*, 2019, pp. 4171–4186. doi:10.18653/v1/N19-1423.
- [9] 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. arXiv:1907.11692.
- [10] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, L. Zettlemoyer, BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension, in: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics*, 2020, pp. 7871–7880. doi:10.18653/v1/2020.acl-main.703.
- [11] C. Liu, X. Wu, M. Yu, G. Li, J. Jiang, W. Huang, X. Lu, A Two-Stage Model Based on BERT for Short Fake News Detection, in: *International Conference on Knowledge Science, Engineering and Management*, Springer, 2019, pp. 172–183. doi:10.1007/978-3-030-29563-9_17.
- [12] N. Reimers, I. Gurevych, Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, in: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Association for Computational Linguistics, 2019, pp. 3982–3992. doi:10.18653/v1/d19-1410.
- [13] A. Conneau, G. Lample, Cross-lingual Language Model Pretraining, in: *Advances in Neural Information Processing Systems*, volume 32, 2019, pp. 1–11. URL: <https://proceedings.neurips.cc/paper/2019/file/c04c19c2c2474dbf5f7ac4372c5b9af1-Paper.pdf>.

- [14] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al., Language models are unsupervised multitask learners, OpenAI blog 1 (2019) 9.
- [15] S. Hochreiter, J. Schmidhuber, Long Short-Term Memory, *Neural Computation* 9 (1997) 1735–1780. doi:10.1162/neco.1997.9.8.1735.
- [16] F. A. Gers, J. Schmidhuber, F. Cummins, Learning to forget: Continual prediction with LSTM, *Neural Computation* 12 (2000) 2451–2471. doi:10.1162/089976600300015015.
- [17] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Y. W. Teh, M. Titterton (Eds.), *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, PMLR, Chia Laguna Resort, Sardinia, Italy, 2010, pp. 249–256. URL: <https://proceedings.mlr.press/v9/glorot10a.html>.
- [18] P. Nakov, A. Barrón-Cedeño, G. D. S. Martino, F. Alam, J. M. Struß, T. Mandl, R. Míguez, T. Caselli, M. Kutlu, W. Zaghrouani, 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: *Lecture Notes in Computer Science*, Springer International Publishing, 2022, pp. 416–428. doi:10.1007/978-3-030-99739-7_52.
- [19] P. Nakov, A. Barrón-Cedeño, G. Da San Martino, F. Alam, J. M. Struß, T. Mandl, R. Míguez, T. Caselli, M. Kutlu, W. Zaghrouani, 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.
- [20] G. K. Shahi, D. Nandini, FakeCovid – A Multilingual Cross-domain Fact Check News Dataset for COVID-19, in: *Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media, ICWSM, 2020*, pp. 1–9. URL: http://workshop-proceedings.icwsm.org/pdf/2020_14.pdf. doi:10.36190/2020.14.
- [21] 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.
- [22] G. K. Shahi, AMUSED: An Annotation Framework of Multi-modal Social Media Data, *arXiv preprint arXiv:2010.00502* (2020).
- [23] G. K. Shahi, J. M. Struß, T. Mandl, Overview of the CLEF-2021 CheckThat! lab task 3 on fake news detection, *Working Notes of CLEF* (2021).
- [24] 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.
- [25] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. L. Scao, S. Gugger, M. Drame, Q. Lhoest, A. Rush, Huggingface’s transformers: State-of-the-art natural language processing, in: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Association for Computational Linguistics, 2020, pp. 38–45. doi:10.18653/v1/2020.emnlp-demos.6.