Fake News Detection in Social Media Using Graph Neural Networks and NLP Techniques: A COVID-19 Use-Case

Abdullah Hamid[1*], Nasrullah Sheikh [2*], Naina Said[1*],

Kashif Ahmad[3*], Asma Gul[4], Laiq Hasan [1], Ala Al-Fuqaha[3]

[1] DCSE, University of Engineering and Technology, Peshawar, Pakistan, [2] IBM Research - Almaden
[3] Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa
University, Qatar Foundation, Doha, Qatar, [4] Department of Statistics, Shaheed Benazir Bhutto Women University,
Peshawar, Pakistan
{kahmad,aalfuqaha}@hbku.edu.qa,nasrullah.sheikh@ibm.com
{nainasaid,laiqhasan,Abdullahhamid}@uetpeshawar.edu.pk

ABSTRACT

The paper presents our solutions for the MediaEval 2020 task namely FakeNews: Corona Virus and 5G Conspiracy Multimedia Twitter-Data-Based Analysis. The task aims to analyze tweets related to COVID-19 and 5G conspiracy theories to detect misinformation spreaders. The task is composed of two sub-tasks namely (i) text-based, and (ii) structure-based fake news detection. For the first task, we propose six different solutions relying on Bag of Words (BoW) and BERT embedding. Three of the methods aim at binary classification task by differentiating in 5G conspiracy and the rest of the COVID-19 related tweets while the rest of them treat the task as ternary classification problem. In the ternary classification task, our BoW and BERT based methods obtained an F1-score of .606% and .566% on the development set, respectively. On the binary classification, the BoW and BERT based solutions obtained an average F1-score of .666% and .693%, respectively. On the other hand, for structure-based fake news detection, we rely on Graph Neural Networks (GNNs) achieving an average ROC of .95% on the development set.

1 INTRODUCTION

In the modern world, social media is playing its part in several ways, for instance in news dissemination and information sharing, social media outlets, such as Twitter, Facebook, and Instagram, have been proved very effective [1, 6, 7, 9]. However, it also comes with several challenges, such as collecting information from several sources, detecting and filtering misinformation [4, 5, 11]. Similar to other events and pandemics, being one of the deadly pandemics in the history, COVID-19 has been the subject of discussion over social media since its emergence. Without any surprise, a lot of misinformation about the pandemic are circulated over social networks. In order to identify misinformation spreaders and filter fake news about COVID-19 and 5G conspiracy, a task namely "FakeNews: Corona Virus and 5G Conspiracy Multimedia Twitter-Data-Based Analysis" has been proposed in the benchmark MediaEval 2020 competition [8].

This paper provides a detailed description of the methods proposed by team DCSE_UETP for the fake news detection task. The

task consists of two parts, namely (i) text-based misinformation detection (TMD), and (ii) structure-based misinformation detection (SMD). The first task (TMD) is based on textual analysis of COVID-19 related information shared on Twitter during January 2020 and 15th of July 2020, and aims to detect different types of conspiracy theories about COVID-19 and its vaccines, such as that "the 5G weakens the immune system and thus caused the current corona-virus pandemic etc., [8]. In the SMD task, the participants are provided with a set of graphs, each representing a sub-graph of Twitter, and corresponds to a single tweet where the vertices of the graphs represent accounts. Similar to TMD, in this task, the participants need to detect and differentiate between 5G and other COVID-19 conspiracy theories.

2 PROPOSED APPROACH

2.1 Methodology for TMD Task

For the text-based analysis, we employed two different methods including a (i) Bag of Words (BoW), and a (ii) BERT model-based solution [3]. Before proceeding with the proposed methods, it is to be noted that the dataset provided for the text-based analysis is not balanced where one of the classes namely *non-conspiracy* contains a very high number of samples while the rest are composed of relatively fewer samples. In total, the majority class contains 4412, while the other two classes, namely *5G conspiracies*, and other conspiracies, are composed of 1263 and 785 samples, respectively. In order to balance the dataset, we rely on an ensemble of different re-sampled datasets, where N models are built/trained by dividing the class with a higher number of samples into n-differing parts as illustrated in Figure 1. After training N models, the results of the models are combined using two different late fusion methods including a majority voting method, and summation of the posterior probabilities. In the majority voting, since we have four models, in the case of tie we consider the accumulative probabilities/scores to assign a label to a test sample.

Before deploying BoW and BERT, text has been cleaned by removing punctuation's keys, such as commas, full-stops, emojis, URLs, and stop words. Once the text is pre-processed, we proceed with the tokenization and creation of BoW vocabulary, which is followed by generation of the feature vector for each sentence. A Naives Bayes classifier is then trained on the extracted features. On the other hand, a logistic regression model is trained on word embeddings generated via BERT.
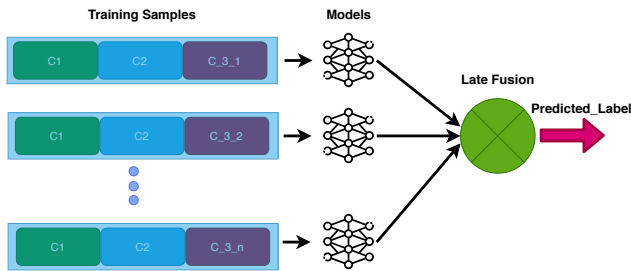
Figure 1: An illustration of the data balancing techniques used in the work.

2.2 Methodology for SMD Task

Graphs representation learning using Graph Neural Networks (GNNs) have been shown to be effective in various domains such as social networks, biological networks, and financial networks. GNNs aggregate the neighborhood representation within k hops and then apply a pooling such as *SUM, MEAN, MAX* to obtain the final representation of the node. Furthermore, GNN's can be used to learn the representation of a simple graph structures [2, 10, 12], which then can be used to classify the graphs. For graph classification, these methods learn the representation of nodes, followed by graph *READOUT* method, which is aggregating the node features obtained after the final iteration of GNN.

We model this problem as a graph classification task. Following Keyule et al.[12], we train our model using three classes of the graphs *5G Conspiracy, non-conspiracy, other-conspiracy*, and learn the representation of the graphs.

3 RESULTS AND ANALYSIS

3.1 Evaluation Metric

For the evaluation of the proposed methods, we used two different metrics, namely (i) Micro F1-Score, and (ii) AUC (Area Under The Curve) ROC (Receiver Operating Characteristics) curve. AUC ROC is the official evaluation metric on the task, and all the test results are reported in terms of AUC ROC. On the other hand, F1-score is used for the evaluation of the methods on the development set.

3.2 Runs Description in TMD Task

For TMD, we submitted six different runs mainly relying on two approaches, namely BERT and BoW, under two late fusion schemes. Three of the runs are based on binary classification while the three deal the task as ternary classification problem. It is to be noted that the fusion schemes are used to combine the scores/output of the four individual models trained as result of the data balancing method as described earlier.

The first three runs are based on the ternary classification task, where run 1 and run 2 are based on BoW with majority voting and accumulative classification scores of the individual models. The third and final ternary run is based on BERT features, where a logistic regression model is trained on word embeddings generated by BERT. As can be seen in Table 1a, overall, better results are obtained with BoW approach under the majority voting scheme.

The last three runs are based on the binary classification task, where the first two (i.e., Run 4 and Run 5) are based on BoW with majority voting and accumulative classification based fusion methods while the final one (i.e., Run 6) is based on BERT with accumulative score based fusion scheme. As expected, the performance of all the methods is significantly higher on the binary classification task compared to ternary classification task.

Similar trend has been also observed on the test set, where overall better results are obtained with BoW under majority voting scheme.

Table 1: Evaluation of our proposed approaches for (a) TMD and (b) SMD tasks on both development and test sets. *In TMD, for the test set the official metric AUC (Area Under The Curve) ROC (Receiver Operating Characteristics) curve is used while the results on the development set are reported in terms of Micro F1-Score. On the other hand, SMD is evaluated in terms of AUC ROC.*

(a) TMD

Run	Dev. Set (F-Score)	Test Set (AUC ROC)
Run 1	0.6066	0.3815
Run 2	0.5666	0.3588
Run 3	0.5333	0.3002
Run 4	0.6933	0.3944
Run 5	0.6666	0.3803
Run 6	0.6533	0.3447

(b) SMD

Run	Dev. Set
Run 1	.9500

3.3 Runs Description in SMD Task

For training the model, we divide the dataset into train/valid/valid (80/10/10). We used the grid search to obtain the best hyperparameters. The model has four MLP layers, and use *MAX* and *MEAN* operations for neighbor pooling and graph pooling respectively. The model is trained on 1000 epochs with a learning rate of 0.01, and dropout 0.3 is applied on the final layer output. The final embedding size is 128. We evaluate our model on AUC-ROC and the result of the test set is given in Table 1(b). The results show that the model has discriminative power to learn to classify the graph structures. Furthermore, it shows that the diffusion of information depending on the type of information being spread forms a diffusion pattern.

4 CONCLUSIONS AND FUTURE WORK

The challenge is composed of two tasks, one aiming to analyze and detect COVID-19 related fake news using tweets' text while the other aims to analyze network structure for the possible detection of the fake news. For the first task, we mainly relied on two state-of-the-art methods namely BoW and BERT embeddings under different fusion schemes. Overall better results are obtained with BoW under the majority voting scheme. For the SMD task, we rely on GNNs to differentiate among different conspiracy theories on COVID-19. In the current implementations, both textual and structural information are used independently, in the future we aim to enrich the structural information with the textual information for better detection of fake news.

REFERENCES

[1] Kashif Ahmad, Konstantin Pogorelov, Michael Riegler, Nicola Conci, and Pal Halvorsen. 2019. Social media and satellites: Disaster event

detection, linking and summarization. *MULTIMEDIA TOOLS AND APPLICATIONS* 78, 3 (2019), 2837–2875.

[2] Cătălina Cangea, Petar Veličković, Nikola Jovanović, Thomas Kipf, and Pietro Liò. 2018. Towards Sparse Hierarchical Graph Classifiers. (2018). arXiv:stat.ML/1811.01287

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[4] Siva Charan Reddy Gangireddy, Cheng Long, and Tanmoy Chakraborty. 2020. Unsupervised Fake News Detection: A Graph-based Approach. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media*. 75–83.

[5] Yi Han, Shanika Karunasekera, and Christopher Leckie. 2020. Graph Neural Networks with Continual Learning for Fake News Detection from Social Media. *arXiv preprint arXiv:2007.03316* (2020).

[6] Muhammad Imran, Prasenjit Mitra, and Carlos Castillo. 2016. Twitter as a lifeline: Human-annotated twitter corpora for NLP of crisis-related messages. *arXiv preprint arXiv:1605.05894* (2016).

[7] Chuang Liu, Xiu-Xiu Zhan, Zi-Ke Zhang, Gui-Quan Sun, and Pak Ming Hui. 2015. How events determine spreading patterns: information transmission via internal and external influences on social networks.

New Journal of Physics 17, 11 (2015), 113045.

[8] Konstantin Pogorelov, Daniel Thilo Schroeder, Luk Burchard, Johannes Moe, Stefan Brenner, Petra Filkukova, and Johannes Langguth. 2020. FakeNews: Corona Virus and 5G Conspiracy Task at MediaEval 2020. In *MediaEval 2020 Workshop*.

[9] Naina Said, Kashif Ahmad, Michael Riegler, Konstantin Pogorelov, Laiq Hassan, Nasir Ahmad, and Nicola Conci. 2019. Natural disasters detection in social media and satellite imagery: a survey. *Multimedia Tools and Applications* 78, 22 (2019), 31267–31302.

[10] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2018. How Powerful are Graph Neural Networks? *CoRR* abs/1810.00826 (2018). arXiv:1810.00826 http://arxiv.org/abs/1810.00826

[11] Shuo Yang, Kai Shu, Suhang Wang, Renjie Gu, Fan Wu, and Huan Liu. 2019. Unsupervised fake news detection on social media: A generative approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 5644–5651.

[12] Rex Ying, Jiaxuan You, Christopher Morris, Xiang Ren, William L. Hamilton, and Jure Leskovec. 2018. Hierarchical Graph Representation Learning with Differentiable Pooling. *CoRR* abs/1806.08804 (2018). arXiv:1806.08804 http://arxiv.org/abs/1806.08804