Using a Word Analysis Method and GNNs to Classify Misinformation Related to 5G-Conspiracy and the COVI~D-19 Pandemic

Ferdinand Schaal[1], Jesper Phillips[1, 2]
[1]Simula Research Laboratory, Norway
[2]Bates College, USA
ferdisvea@gmail.com,jesperphillips@gmail.com

ABSTRACT

This paper addresses the FakeNews: Corona virus and 5G conspiracy task at MediaEval 2020. The task involves classifying misinformation that is related to conspiracy topics and the COVID-19 pandemic. The task is divided into two subtasks where we for each subtask are proposing a separate approach. The first subtask is a Natural Language Processing (NLP)-based detection task where we are proposing a simple text-based approach by looking at the frequency of words. The second subtask is a structural-based detection task where we are proposing a method using a Graph Neural Network (GNN) to perform classification by investigating spreading patterns.

1 INTRODUCTION

Digital wildfires are the rapid spread of online false information that poses the risk of both economical- and physical harm to society and individuals. In an effort to combat this, the FakeNews: Corona virus and 5G conspiracy task [7] aims to build a classifier that can identify fake news on Twitter. We are proposing two different approaches, one for each of the subtasks, to classify a data set of tweets that has been collected with the FACT framework [8]. The data set contains three classes; 5G Conspiracy, Other Conspiracy and Non Conspiracy.The task is split into two subtasks, each containing a labeled training set and unlabeled test set. The first subtask is a Natural Language Processing (NLP)-based detection task. The data in the NLP task consists off tweets and their content. For this subtask we are using a bag-of-words (BOW) method to detect fake news. The second subtask is a structure-based detection task. In this subtask the data are tweets represented by distribution graphs. A distribution graph is a subgraph of the Twitter graph and represents a specific tweet, where the central node is the author of the tweet and the rest of the nodes are users that have shared the tweet. The edges in the graphs are *friend/follower* relationships between the users. We also have additional node features that will be used in some of the experiments.

2 RELATED WORK

Fake news detection has been widely recognized in research in the recent past. There are numerous ways of tackling the challenging problem of fake news. Zhou et al. [12] divides the main methods used for fake news detection into four categories; *Knowledge-based*, *Style-Based*, *Propagation-Based* and *Source-Based*. Our two proposed methods for NLP-based detection and structure-based detection fall under the category of style-based and propagation-based methods, respectively. Our NLP-based approach is much similar to one of the proposed methods in Zhou et al. [11] that uses BOW to obtain the frequency of lexicons to detect fake news. Propagation-based methods typically uses Machine Learning (ML) methods such as support vector machine (SVM) [1], random forest [4] and recursive neural networks [6] to classify fake news. Unlike our approach, these methods rely on manual feature extraction. Manual feature extraction is not needed in our structure-based approach because the GNN operates directly on the distribution graphs.

3 APPROACH

3.1. NLP-based Fake News Detection

The groundwork for our NLP classifier is done by applying a BOW method on the training set. As classes contain differing amounts of tweets, we use a ratio of word count per tweet within a class to better compare between classes. By comparing such a ratio of two classes, we get insight into how frequently a word is used by one class compared to a different class. By applying this method to the 5G Conspiracy class in the training set we are able to interpret how often a word is used by a class. By defining the number of times, a word X is mentioned in a class for M_{Class} and the total amount of tweets in that class for N_{Class} we propose a method to score an individual word given by the following equation:

$$\text{Score of a word} = \frac{M_{5G}}{N_{5G}} \cdot \frac{N_{NonConsp}}{M_{NonConsp}} \qquad (1)$$

All words are assigned a score depending on the ratio of usage in the 5G-Conspiracy class compared to usage in the Non-Conspiracy class. A word like "and", which exists in most tweets regardless of class would score close to 1.00. While words used mostly by the 5G conspiracy class would score higher. By taking the geometric mean of the scores of each word within a tweet, we establish an overall score for a specific tweet. We then apply this method to every tweet in the test set and rank all the tweets based on their overall score.

To find a threshold score for classifying a 5G-Conspiracy tweet, we ran a preliminary test on the development data. We then calculated the sensitivity and 1− specificity of all thresholds of the results from the test on the development data. We found the sensitivity and 1−specificity to have the highest geometric mean at a threshold of 1.15. This threshold was then applied to the test set.

The binary classification was done by creating scores with the combined Other Conspiracy and Non-Conspiracy tweets as class

MediaEval'20, December 14-15 2020, Online

0, and 5G Conspiracy class as class 1 as requested by the task. We then created a multi class classification between 5G Conspiracy, Other Conspiracy and Non-Conspiracy as Class 1, 2, 3. To do so we ran our method twice: first using the binary method and secondly by classifying between other conspiracy and regular tweets. Thus, by combining these two runs, we are able to establish a multi-class classification of the 5G Conspiracy, Other Conspiracies, and Non-Conspiracy

3.2. *Structure-based Fake News Detection*

For the task of Structure-based Fake News Detection, Graph Neural Networks (GNNs) are used as a classifier. GNNs are generalizations of deep learning architectures such as Convolutional Neural Network (CNN) [5] and Recurrent Neural Network (RNN) [2] that allow neural networks to directly operate on graph structured data. There exist numerous kinds of GNNs within the categories of convolutional GNNs (ConvGNNs) and recurrent GNN (RecGNN). We have chosen to work with a spatial-based convolutional GNN because of its proven efficiency, generality and flexibility compared to the other GNNs [9]. Based on its promising results in previous work [10][3], our model will be the Graph Isomorphism Network (GIN) [10]. The reader is referred to Xu et al. [10] for a detailed description of GIN, but in short it uses a neighborhood aggregation scheme in order to perform node embeddings that allows for classification.

Three different classification tasks are conducted: one multi-class classification with node features, one multi-class classification without node features, and one binary classification task without node features. For the tasks where node features are not provided, we are using one-hot encoding of node degrees as input features. In the binary classification task, we are grouping Other Conspiracy and Non-Conspiracy together.

Hyperparameter optimization is conducted separately for each classification task with a 5-fold cross-validation, which splits the training set into a 80%/20% training/validation set for each fold.

To cope with the unbalanced sets, we oversample the minority classes in our training sets. The hyperparameters used in the cross validation are chosen based on the one used by Xu et al. [10] and are listed in Table 1.

Table 1: Tunable hyperparameters

Batch Size	Hidden Layers	Hidden Units
32, 64, 128	2, 3, 5	32, 64

The models are then retrained on another 80%/20% split of the training set using the best performing hyperparameter configuration. With drop-out and early stopping, we make sure that the model is not overfitting on the training set.

4 RESULTS AND ANALYSIS

4.1. *Results: NLP-based Fake News Detection*

For the NLP subtask, we achieved a score of 0.372 for multi-class classification and 0.385 for binary classification. It was expected that the score would be higher for the binary classification, as at its core, our classifier is a binary method, ran multiple times to make tertiary. We identify certain flaws in our method. Reviewing our results, we saw the method did better for the highest threshold of our three runs. We, therefore, suspect our thresholds may have been set too low. A different method of determining this threshold resulting in a higher threshold might have led to greater accuracy in the classification.

4.2. *Results: Structure-based Fake News Detection*

The best performing hyperparameters for all three experiments were batch size of 128, 32 hidden units, and three hidden layers when additional features were provided and two otherwise. This means that the shallower networks perform better than the deeper networks, which is likely due to the fact that the distribution graphs are relatively small in size and well connected. For the multi-class tasks our GIN model had a score of 0.1810 when using features and a score of 0.1375 without using features. The performance declined for the binary classification task, which received a score of 0.1122. Based on the results, using the additional features provided increases the performance significantly. Nevertheless, the performance is fairly low compared to the NLP subtask. Table 2 below reports the results of each experiment in both subtasks.

Table 2: Highest quantitative Results

Submitted Run	Metric Score
NLP Binary	0.385
NLP Multi-Class	0.372
Structure Binary	0.112
Structure Multi-Class	0.138
Structure Multi-Class with features	0.181

5 DISCUSSION AND OUTLOOK

In our NLP-based results we see that when the method fails, it often fails by not managing to interpret human characteristics. It has no method to identify characteristics like sarcasm and humor in tweets. A sarcastic tweet discussing the 5G-Conspiracy, yet not claiming it is true may still contain high scoring frequency words. Therefore such a tweet would be wrongly classified as 5G-Conspiracy. To counter this we would have had to incorporate methods more advanced than our simple word counting.

It is interesting that the binary classification for the structure-based approach performs worse compared to the multi-class classification. This might be because of similarities in the distribution graphs between the two conspiracy classes: 5G-Conspiracy and Other-Conspiracy, which makes it difficult to separate the two classes.

In the future, it would be interesting to see a hybrid method of the two proposed approaches. A combination of the NLP-based approach with the structure-based approach would most likely result in a much more robust model.

ACKNOWLEDGMENTS

This work was created in the context of the Norwegian SAMRISK-2 project "UMOD" (#272019) under the supervision of Johannes Langguth, Daniel Thilo Schroeder, and Konstantin Pogorelov.

REFERENCES

[1] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on Twitter. *Proceedings of the 20th International Conference Companion on World Wide Web, Www 2011* (2011), 675–684. https://doi.org/10.1145/1963405.1963500

[2] JL ELMAN. 1990. FINDING STRUCTURE IN TIME. *Cognitive Science* 14, 2 (1990), 179–211. https://doi.org/10.1207/s15516709cog1402_1

[3] Federico Errica, Marco Podda, Davide Bacciu, and Alessio Micheli. 2020. A Fair Comparison of Graph Neural Networks for Graph Classification. (2020).

[4] Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. 2013. Prominent features of rumor propagation in online social media. *Proceedings - Ieee International Conference on Data Mining, Icdm* (2013), 6729605, 1103–1108. https://doi.org/10.1109/ICDM.2013.61

[5] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio. 1999. Object recognition with gradient-based learning. *Shape, Contour and Grouping in Computer Vision* (1999), 319–45, 319–345.

[6] Jing Ma, Wei Gao, and Kam Fai Wong. 2018. Rumor detection on twitter with tree-structured recursive neural networks. *Acl 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (long Papers)* 1 (2018), 1980–1989. https://doi.org/10.18653/v1/p18-1184

[7] 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*.

[8] Daniel Thilo Schroeder, Konstantin Pogorelov, and Johannes Langguth. 2019. FACT: a Framework for Analysis and Capture of Twitter Graphs. In *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*. IEEE, 134–141.

[9] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S. Yu. 2019. A Comprehensive Survey on Graph Neural Networks [arXiv]. *Arxiv* (2019), 22 pp., 22 pp.

[10] Keyulu Xu, Stefanie Jegelka, Weihua Hu, and Jure Leskovec. 2019. How powerful are graph neural networks? *7th International Conference on Learning Representations, Iclr 2019* (2019).

[11] Xinyi Zhou, Atishay Jain, Vir V. Phoha, and Reza Zafarani. 2019. Fake News Early Detection: A Theory-driven Model [arXiv]. *Arxiv* (2019), 24 pp., 24 pp.

[12] Xinyi Zhou and Reza Zafarani. 2020. A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *Acm Computing Surveys* (2020), 37. https://doi.org/10.1145/3395046