	Bangladesh Army University of Science and Technology (BAUST)			
	Department of Computer Science and Engineering			
	Assignment #1, Winter 2023		Level-4	Term-I
	Course Code: CSE 4131	Course Title: Artificial Neural Networks and Fuzzy Systems		
	Submission Date:	CO Number: CO2	Full Marks: 15	
ID: 200101057		Name: ABU SHADAT SHAIKAT		

Fake News Detection on Social Media

1. Introduction

Social media has seen a rise in the spread of fake news, which can mislead the public and give some parties unfair financial, political, or psychological advantages. The use of machine learning and data mining methods to identify misleading information has gained popularity as a research topic. The significant data sparsity issue, the absence of user comments, privacy problems, and the inability to offer explanations make it tough to recognize fake news on social media. The Graph-aware Co-Attention Network (GCAN) is a novel model suggested in this paper to identify bogus news in a more realistic social media environment. GCAN can determine whether a source tweet story is bogus based just on its brief text content, the persons who have retweeted it, and user profiles. Under three, GCAN can identify fake information

2. Literature Review

2.1 A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media

Proposed Method: The proposed method in the FakeNewsNet paper is the creation of a comprehensive data repository that includes news content, social context, and dynamic information for studying fake news on social media. The authors collected two datasets, PolitiFact and Gossip Cop, from 2016 to 2017 and extracted news content, social context, and dynamic information such as user comments, user engagement, and network information. The data was preprocessed and stored in a format that can be easily used for machine learning tasks, such as fake news detection.

The authors also proposed several research directions that can be explored using the FakeNewsNet repository, such as analyzing the effectiveness of different features for fake news detection, exploring the impact of social context on fake news, and studying the diffusion patterns of fake news on social media. The authors hope that the FakeNewsNet repository will serve as a valuable resource for researchers and practitioners to study and combat fake news on social media.

Result: The authors utilized the PolitiFact and Gossip Cop datasets from the Fake Newsnet repository to perform fake news detection task. They used 80% of data for training and 20% for testing and deployed several state-of-the-art baselines for fake news detection. The baselines included news content-based methods, social context-based methods, and a combination of news content and social context. Evaluation metrics included accuracy, precision, recall, and F1 score. The results showed that the SAF model provided the best performance in terms of accuracy, recall, and F1 score for news content-based methods, while logistic regression had better precision. SAF /A provided similar results to SAF /S but had higher precision. SAF provided better accuracy than both SAF /S and SAF /A for both datasets, indicating that social engagements can help fake news detection in addition to news articles. The compared baselines models provided reasonably good performance results for fake news detection, with accuracy mostly around 65% on PolitiFact.

Limitations: While the Fake Newsnet repository is a valuable resource for researchers studying fake news detection and related topics, there are some limitations to consider. One limitation is that the repository only includes data from a limited number of sources, which may not be representative of the larger landscape of fake news. Additionally, the data collection strategy used in this study may introduce some biases into the repository, as it relies on web search results and may not capture all instances of fake news. Furthermore, the preliminary exploration study presented in this paper focuses on a relatively small set of features, and more work is needed to fully understand the potential of the repository for supporting research in this area. Finally, while the authors propose several promising avenues for future work, it remains to be seen how effectively the repository will be integrated with front-end software and other tools for studying fake news.

Overall, while Fake Newsnet is a valuable resource for researchers, it is important to consider these limitations when interpreting the results of studies that rely on this repository.

2.2 Combating Hostility: Covid-19 Fake News and Hostile Post Detection in Social Media

Proposed Method: The proposed methods for two different tasks are described in the given content. For the binary classification task of detecting fake or real tweets, SVM with linear and RBF kernels are used along with tf-idf feature extraction technique. Word embedding features are also applied to SVM. CNN, BiLSTM, and a combined CNN+BiLSTM network are implemented for the DL approach. For the multi-label multi-class problem of identifying hostile dimensions, label powerset (LP) along with SVM is used for the ML approach. BiLSTM network with Word2Vec embedding technique is employed for the DL approach. Different hyperparameters are varied and experimented with for both approaches. Overall, SVM shows superior performance for the first task, and BiLSTM with Word2Vec embedding performs well for the second task.

Result: The article discusses the evaluation results of two tasks - Task A and Task B. In Task A, four models were evaluated based on the weighted f1 score, with the SVM model using the tf-idf approach achieving the highest f1 score of 94.39%. In contrast, the best model in Task B was determined based on coarse-grained (CG) and fine-grained (FG) f1-scores, and the results showed that the combination of SVM and Word2Vec achieved the highest scores. However, the best performing model in Task A had an f1 score of 98.69%, which was significantly better than the best model in Task B.

Method	A	P	R	F
CNN + BiLSTM	92.01	92.01	92.01	92.01
SVM + TF-idf	94.35	94.42	94.39	94.39
SVM + Word2Vec (ED=200)	92.66	92.67	92.66	92.66
SVM + Word2Vec (ED=150)	92.94	92.94	92.94	92.94
Best	98.69	98.69	98.69	98.69

Table: Evaluation results of task-A on test set. Here A, P, R, F denotes accuracy, precision, recall, weighted f1 score respectively and ED indicates embedding dimension

The article presents the evaluation results of task-B on the test set using different models and features. The best performing model for CG and FG classification achieved a f1 score of 86.03% and 50.66%, respectively, using LPSVM with n-gram range (1, 3). For the defamation and fake class, BiLSTM + Word2Vec obtained the highest f1 score of 28.65% and 52.06%, respectively, while LPSVM + Ngram (1, 2) and LPSVM + Ngram (1, 3) achieved the highest f1 score of 64.97% (for fake class) and 57.91% (for offensive class). However, the best performing model lags behind the best results obtained in task-B by more than 10% for CG and 14% for FG.

Method	CG	Defame	Fake	Hate	Offense	FG
BiLSTM + Word2Vec	83.37	28.65	63.63	52.06	55.72	52.80
LPSVM + Unigram	84.10	25.81	61.30	44.39	53.59	49.12
LPSVM + Ngram (1,2)	85.31	27.59	64.97	47.21	51.72	50.98
LPSVM + Ngram (1,3)	86.03	21.74	63.33	46.67	57.91	50.66
Best	97.15	45.52	82.44	59.78	62.44	64.40

Table: Evaluation results of Task-B on the test set. All values presented in f1 score and CG, FG denotes coarse-grained, and fine-grained f1 scores

Limitations: The article discusses the limitations and challenges encountered while developing a system for the CONSTRAINT 2021 shared task, which aims to detect fake news and hostility in social media. The system achieved impressive results, but some misclassified examples suggest areas for improvement.

One limitation is the presence of influential words, such as "covid19", "coronavirus", "modi", and "vaccine", that occur frequently in both real and fake tweets, making it difficult to differentiate between them. The fake news class also presents a challenge, as it often claims unverified facts using credible sources such as FDA and WHO. Identifying hostility is another challenge, as some posts inherently express hostility, and separating defame, hate, and offensive posts is often confusing due to overlapping characteristics.

The article also notes that machine learning models performed better than deep learning models in the shared task, possibly due to a lack of training examples in some classes. However, using large pre-trained language models, such as BERT, improved accuracy significantly. The article suggests that future work could explore ensemble techniques and attention layers to improve the performance of CNN and BiLSTM models. Increasing the number of posts in hostile classes could also improve model performance.

In conclusion, the article highlights the challenges of detecting fake news and hostility in social media and provides insights into the limitations of current models. While the developed system achieved impressive results, there is still room for improvement, and future research could explore new techniques to overcome the identified limitations.

2.3 Fake News Detection on Social Media: A Data Mining Perspective

Proposed Method: The article proposes a data mining perspective for detecting fake news on social media. It suggests four research directions: data-oriented, feature-oriented, model-oriented, and application-oriented. The article emphasizes the need for a comprehensive and large-scale fake news benchmark dataset, as well as more advanced techniques for feature extraction and model building. Additionally, it suggests that there is a need to develop practical applications for detecting fake news that can be easily used by the general public.

Result: The article presents an analysis of fake news detection from a data mining perspective. It highlights four categories of research directions: Data-oriented, Feature-oriented, Model-oriented, and Application-oriented.

The Data-oriented category focuses on different kinds of data characteristics such as datasets, temporal, and psychological factors. The article argues that there is no existing benchmark dataset that includes resources to extract all relevant features. Therefore, a comprehensive and large-scale fake news benchmark dataset is needed. Additionally, the article suggests exploring early fake news detection and understanding the psychological factors that lead to fake news dissemination, such as the echo chamber effect.

The Feature-oriented category aims to determine effective features for detecting fake news from multiple data sources such as news content and social context. The article explores the use of linguistic-based and visual-based techniques for feature extraction from text and images, respectively. Moreover, it suggests that more advanced visual-based features are required to differentiate real and fake visual content, and more research is needed to extract user-specific and post-based features.

The Model-oriented category opens the door to building more effective and practical models for fake news detection. The article suggests that more research can be done to build more complex and effective models, such as aggregation methods, probabilistic methods, ensemble methods, or projection methods, to better utilize extracted features.

The Application-oriented category focuses on developing applications to detect fake news, such as browser extensions, social media plugins, or mobile apps. The article suggests exploring different ways to present the results of fake news detection to users and exploring the use of gamification to engage users in the detection process.

Overall, the article suggests that there are many research directions that can be explored to improve fake news detection. It emphasizes the need for a comprehensive and large-scale fake news benchmark dataset, as well as more advanced techniques for feature extraction and model building. Additionally, it suggests that there is a need to develop practical applications for detecting fake news that can be easily used by the general public.

Limitations: As the field of fake news detection continues to grow, it is important to acknowledge the limitations of current research and the challenges that lie ahead. While significant progress has been made in developing techniques to identify fake news, there are still several obstacles that need to be overcome to ensure the accuracy and effectiveness of these methods.

One of the primary limitations of current research is the lack of standardized datasets. As mentioned in previous studies, there is currently no benchmark dataset that includes all relevant features for detecting fake news. This makes it difficult for researchers to compare and evaluate the performance of different techniques, as they may be using different datasets with varying levels of complexity and relevance.

Furthermore, there is a significant amount of variability in the types of fake news that exist. Fake news can take on many different forms, from fabricated stories to manipulated images and videos. This variability makes it challenging to develop a one-size-fits-all approach to detecting fake news, as different techniques may be better suited for different types of fake news.

Another challenge facing researchers is the rapidly evolving nature of social media platforms. Social media sites are constantly changing their algorithms, policies, and features, which can impact the effectiveness of detection techniques. Additionally, the sheer volume of data on social media platforms can make it difficult to accurately identify and track the spread of fake news.

Another limitation of current research is the lack of attention given to the intent behind fake news. While existing techniques are effective at identifying the authenticity of news content, they do not necessarily take into account the motivations behind the creation and dissemination of fake news. Understanding the intent behind fake news can help researchers develop more targeted and effective detection techniques.

Finally, it is important to acknowledge the potential ethical implications of fake news detection techniques. There is a risk that these techniques could be used to unfairly target and censor certain groups or individuals, or to perpetuate existing biases and prejudices. It is important for researchers to be mindful of these concerns and to develop methods that are fair, transparent, and unbiased.

In conclusion, while significant progress has been made in developing techniques to detect fake news, there are still several limitations and challenges that need to be addressed. Standardized datasets, attention to the evolving nature of social media platforms, consideration of the intent behind fake news, and ethical concerns are all important areas for future research. By acknowledging these limitations and working to address them, researchers can continue to develop more accurate and effective techniques for detecting fake news.

2.4 Fake News Detection on Social Media using Geometric Deep Learning

Proposed Method: The proposed method for fake news detection on social media using geometric deep learning involves using a graph neural network to model the social network structure and relationships between users. The model uses both textual and visual features to detect fake news and incorporates a propagation mechanism to account for the spread of fake news within the network. Experimental results show that the proposed method outperforms several baselines in terms of accuracy and F1 score.

Result: The study aimed to detect fake news in two different settings: URL-wise and cascade-wise. In the URL-wise setting, the true/fake label of a news story's URL was predicted from all the Twitter cascades it generated, with each URL resulting in an average of 141 cascades. In the cascade-wise setting, the label associated with a URL was predicted based on only one cascade from that URL. The study assumed that all the cascades associated with a URL inherited the label of the URL. The authors found their assumption to be true in most cases but noted that future research could analyze comments accompanying tweets/retweets to confirm this.

Limitation: The article presents a promising approach for fake news detection on Twitter using geometric deep learning. The proposed method integrates various data related to user profiles and activity, social network structure, news spreading patterns, and content. The advantage of using deep learning is its ability to learn task-specific features automatically, which improves the accuracy of the model. The study showed that the model has high accuracy and robust behavior in different settings, demonstrating the potential of geometric deep learning for fake news detection.

However, there are limitations that need further exploration. One of them is the potential language and geography independence of the model, which needs experimental validation. Additionally, the study of adversarial attacks is crucial to examine the resilience of the model and to enhance its interpretability. Adversarial attacks could also provide insights into the decision-making process of the graph neural network. The authors conjecture that graph-based approaches are difficult to manipulate in practice, making their method particularly attractive.

Furthermore, the authors intend to explore additional applications of their model beyond fake news detection. For example, news topic classification and virality prediction are possible areas of study. The study opens up several avenues for future research and highlights the potential of geometric deep learning in social network data analysis.

3. Result and Analysis

1. The authors evaluated several state-of-the-art baselines for fake news detection using PolitiFact and Gossip Cop datasets. They used news content-based methods, social context-based methods, and a combination of news content and social context. The results showed that the SAF model provided the best performance in terms of accuracy, recall, and F1 score for news content-based methods, while logistic regression had better precision. The compared baselines models provided reasonably good performance results for fake news detection, with accuracy mostly around 65% on PolitiFact.
2. The article discusses the evaluation results of two tasks, Task A and Task B. Task A evaluated four models based on the weighted f1 score, with the SVM model using the tf-idf approach achieving the highest f1 score of 94.39%. In contrast, the best model in Task B was determined based on coarse-grained (CG) and fine-grained (FG) f1-scores, and the results showed that the combination of SVM and Word2Vec achieved the highest scores. However, the best performing model in Task A had an f1 score of 98.69%, which was significantly better than the best model in Task B.
3. The article highlights four categories of research directions for fake news detection: Data-oriented, Feature-oriented, Model-oriented, and Application-oriented. The Data-oriented category emphasizes the need for a comprehensive and large-scale fake news benchmark dataset and explores early fake news detection and understanding the psychological factors that lead to fake news dissemination. The Feature-oriented category aims to determine effective features for detecting fake news from multiple data sources such as news content and social context. The Model-oriented category suggests exploring more effective and practical models for fake news detection, such as aggregation methods, probabilistic methods, ensemble methods, or projection methods. The Application-oriented category focuses on developing applications to detect fake news, such as browser extensions, social media plugins, or mobile apps.
4. The study aimed to detect fake news in two different settings: URL-wise and cascade-wise. In the URL-wise setting, the true/fake label of a news story's URL was predicted from all the Twitter cascades it generated. In the cascade-wise setting, the label associated with a URL was predicted based on only one cascade from that URL. The authors found their assumption to be true in most cases but noted that future research could analyze comments accompanying tweets/retweets to confirm this.

4. Conclusion

Overall, these four papers focus on different aspects of fake news detection and mitigation using various techniques and approaches. The first paper proposes a deep learning approach based on geometric deep learning to detect fake news on social media platforms, which achieved high accuracy and robustness. The second paper provides a review of existing literature on the characterization and detection of fake news and discusses promising future directions for research. The third paper describes the system and results of the CONSTRAINT 2021 shared task, which aimed to detect hostile posts in online communities, and highlights the need for future work on improving the performance of models using ensemble techniques and increasing the number of posts in hostile classes. Finally, the fourth paper presents a comprehensive repository called FakeNewsNet for collecting relevant data from different sources, which has the potential to facilitate various research directions such as fake news detection, mitigation, evolution, and malicious account detection. In conclusion, these papers contribute valuable insights and techniques to the growing field of fake news detection and mitigation, and highlight the importance of continued research in this area to address the negative impacts of fake news on individuals and society.

References

[1] Baccianella, Esuli, and Sebastiani 2010] Baccianella, S.; Esuli, A.; and Sebastiani, F. 2010. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In Lrec, volume 10, 2200–2204.

[2] [Castillo et al. 2014] Castillo, C.; El-Haddad, M.; Pfeffer, J.; and Stempeck, M. 2014. Characterizing the life cycle of online news stories using social media reactions. In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing, 211–223. ACM.

[3] [Chen, Conroy, and Rubin 2015] Chen, Y.; Conroy, N. J.; and Rubin, V. L. 2015. Misleading online content: Recognizing clickbait as false news. In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection, 15– 19. ACM.

[4] [Del Vicario et al. 2016] Del Vicario, M.; Vivaldo, G.; Bessi, A.; Zollo, F.; Scala, A.; Caldarelli, G.; and Quattrociocchi, W. 2016. Echo chambers: Emotional contagion and group polarization on facebook. Scientific reports 6:37825.

[5] Farajtabar et al. 2017] Farajtabar, M.; Yang, J.; Ye, X.; Xu, H.; Trivedi, R.; Khalil, E.; Li, S.; Song, L.; and Zha, H. 2017. Fake news mitigation via point process based intervention. arXiv preprint arXiv:1703.07823

[6] [Gilbert 2014] Gilbert, C. H. E. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. Pdf

[7] Manning et al. 2014] Manning, C.; Surdeanu, M.; Bauer, J.; Finkel, J.; Bethard, S.; and McClosky, D. 2014. The stanford corenlp natural language processing toolkit. In Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations, 55–60

[8] Zeitel-Bank, N., Tat, U., et al.: Social media and its effects on individuals and social systems. Journal Management, Knowledge, And Learning (2014)

[9] Wiegand, M., Siegel, M., Ruppenhofer, J.: Overview of the germeval 2018 shared task on the identification of offensive language (2018)

[10] Sadia Afroz, Michael Brennan, and Rachel Greenstadt. Detecting hoaxes, frauds, and deception in writing style online. In ISSP'12

[11] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. Technical report, National Bureau of Economic Research, 2017.

[12] Meital Balmas. When fake news becomes real: Combined exposure to multiple news sources and political attitudes of inefficacy, alienation, and cynicism. Communication Research, 41(3):430–454, 2014

[13] Alessandro Bessi and Emilio Ferrara. Social bots distort the 2016 us presidential election online discussion. First Monday, 21(11), 2016.

[14] Carlos Castillo, Mohammed El-Haddad, J¨urgen Pfeffer, and Matt Stempeck. Characterizing the life cycle of online news stories using social media reactions. In CSCW'14.