CIC at CheckThat! 2021: Fake News detection Using Machine Learning And Data Augmentation

Noman Ashraf[1], Sabur Butt[1], Grigori Sidorov[1] and Alexander Gelbukh[1]

[1](CIC, Instituto Politécnico Nacional, Mexico)

Abstract

Disinformation in the form of fake news, phoney press releases and hoaxes may be misleading, especially when they are not from their original sources and this fake news can cause significant harm to the people. In this paper, we report several machine learning classifiers on the CLEF2021 dataset for the tasks of news claim and topic classification using n-grams. We achieve an F_1 score of 38.92% on news claim classification (task 3a) and an F_1 score of 78.96% on topic classification (task 3b). In addition, we augmented the dataset for news claim classification and we observed that insertion of alternative words was not beneficial for the fake news classification task.

Keywords

fake news detection, fake news data augmentation, fake news topic classification, fake news claim classification,

1. Introduction

Increase in social media outlets has impacted many natural language problems such as emotion detection [1, 2], human behavior detection [3] question answering [4], threat detection [5], sexism detection [6], depression detection [7] etc. Easy and accessible dissemination of news in social media has resulted in a dire need for fake news identification and checks online. To ensure the credibility of news spreaders on social media, the research community needs to play its part in developing automatic methods of identification of false claims, disinformation and misinformation. Automatic detection of fake news aims to mitigate the time and human resources spent on identifying fake news and spreaders from the stream of continuously created data.

To tackle this problem, natural language processing (NLP) researchers have made many sophisticated attempts by creating specific tasks for detecting rumor [8, 9], fact checking [10, 11], deception [12, 13], article stance [14, 15, 16], satire [17, 18], check worthiness [10, 19, 20, 21, 22, 23], cherry picking [24, 25], clickbait [26, 27, 28] and hyperpartisan [29, 30] in English language. The tasks have been attempted using rules crafted by humans, machine learning (ML) models [31, 32] and deep learning (DL) methods [33, 34, 35].

In this paper, we have tackled two tasks of CLEF2021 fake news classification. The first task required multi-class classification of articles to determine if the claim made in the article is

CLEF 2021 – Conference and Labs of the Evaluation Forum, September 21–24, 2021, Bucharest, Romania
nomanashraf712@gmail.com (N. Ashraf); sabur@nlp.cic.ipn.mx (S. Butt); sidorov@cic.ipn.mx (G. Sidorov); gelbukh@gelbukh.com (A. Gelbukh)

true, false, partially false or other (lack of evidence to conclude). The second task required the classification of the topic of an article. The fake news article was required to be classified into five or more categories like election, health, conspiracy theory etc. The paper discusses the difference in results with various machine learning methods. We attempted to gauge the potential of machine learning methods on the described task. Both of these tasks were attempted and the results were presented in the competition.

2. Related Work

Faking a piece of news has been part of all eras of technology in the form of yellow journalism. However, since the advent of social media, the impact of the harm has grown many folds. It has hence been one of the most challenging problems for researchers to solve since the last decade as it is very difficult to distinguish fake text from real text. Theoretical fake news studies [12, 36, 37] has seen classification of fake news in the form of misinformation, disinformation, hysteria, falsehood, propaganda, clickbait and conspiracy theories. We have seen advances in the field in the recent decade that had a real-life impact.

There are various methods to differentiate fake news from real news such as bag-of-words (BOW) [38], n-grams [39], GloVe [40], term frequency—inverse document frequency (TF-IDF) [39] and contextual embeddings. Methods like bag-of-words do not include context and rely on word frequencies, albeit, researchers have also used semantic analyses [41] to determine truthfulness in a topic. We have also seen a well deep syntax approach [39] using probability context free grammar (PCFG) parsing trees. This approach uses rewritten uses of sentences to study differences in syntax structures in real and fake news. Another linguistic approach [14, 15, 16] is to consider the topic of the article and test its relevance with the content of the article. This is done by using linguistic features such as the length of the headlines, advertisements, text patterns, author attributes etc.

Various machine-learning methods have been used as well for fake news detection: support vector machine (SVM) [42], naïve bayes (NB) [32], logistic regression (LR) [43], k-nearest neighborhood (K-NN) [31], random forest (RF) [44] and decision trees (DT) [44]. These methods have displayed strength in classifying misinformation using various features. Since feature engineering is time-consuming, various neural network approaches such as long short-term memory (LSTM) with linguistic inquiry and word count (LIWC) features [35], recurrent neural networks (RNN) based models [45, 46, 34] for user engagement and convolutional neural network (CNN) based model [33, 47] with local features were applied to detect fake news.

3. Dataset

Dataset for task 3a consisted of 900 articles with four labels. The claim in the article is detected and classified as true, false, partially false or others. The “others” class identifies articles that cannot be proven as false, true or partially true. While the partially false articles are those that have weak evidence of the claim. In addition to this, task 3b uses the subset of task 3a articles but classifies the article in six categories namely education, health, crime, election, climate and economy. Table 1 and 2 show us the sample of the dataset for task 3a and 3b, while Table 3 shows the distribution of the dataset in both tasks according to their respective classes. The ellipsis in the text demonstrates the omission of the complete article in the Table 1 and 2.

Table 1
Samples of Task 3a

Public Id	Text	Title	Rating
5a228e0e	Distracted driving causes more deaths in Canada than impaired driving .It's why every province and territory has laws against driving while operating a cell phone. "Tell your passengers to stay off their phones while you are driving...	You Can Be Fined $1,500 If Your Passenger Is Using A Mobile Phone, Starting Next Week	false
0a450bd4	Her name is Taylor Zundel, and it sounds like she and her husband live in or near Salt Lake City. And she witnessed quite the irregularity when they showed up for early voting: Not just her husband, but at least one other voter, were told when they got there that records showed they had already voted...	Instagram Testimony: People Are Showing Up to Vote and Being Told They Already Voted	true

Table 2
Samples of Task 3b

Public Id	Title	Text	Domain
f6e07bea	Manchin Introduces Landmark Veterans Mental Health And Suicide Prevention Bill Washington, D.C. - U.S. Senator Joe Manchin (D-WV) introduced a landmark, bipartisan bill to improve Veterans' access to mental health care and make sure no Veteran's life is lost to suicide...	Manchin Introduces Landmark Veterans Mental Health And Suicide Prevention Bill	health
a3910250	Self-harm and violent attacks have hit record levels in prisons across England and Wales for the second time in a year, despite repeated warnings that jails are at crisis point and in desperate need of reform...	Self-harm and violent attacks hit record high in prisons across England and Wales for second time in a year	crime

4. Methodology

We used several machine learning algorithms such as logistic regression, multilayer perceptron, support vector machine and random forest. For RF and MLP classifiers, default parameters were used for all the experiments. We assign class weight parameter to "balance" for SVM and LR. In addition, "saga" kernel was used for LR. Stratified 5-fold validation is applied for the evaluation

Table 3
Data distribution of tasks

Classes	Size
false	461
true	135
partially false	216
other	76

(a) Task 3a

Classes	Size
education	28
health	126
crime	37
election	32
climate	46
economy	42

(b) Task 3b

of the results. While accuracy, precision, recall and F_1 are given for a thorough understanding of the results, the competition ranked the teams using F_1-macro. In NLP and opinion mining tasks [48] these classifiers performed best. We also considered the limitations of the task including an imbalanced dataset, especially for task 3a. The article contains grammatical errors, spelling errors and repetition of keywords. Repetition of keywords for fake news negatively influence the results of term frequency.

4.1. Pre-Processing

All pre-processing tasks were attempted using Ekphrasis [49] library. The normalization process included removing "url", "email" , "percent", "money", "phone", "user", "time", "date", and "number" instances from the text. The contraction was also unpacked for better context i.e. hasn't changed into has not. Since we often encounter elongated words in informal news articles, the elongated words were spell corrected to their base words.

4.2. Augmented Dataset

The data was augmented using word2vec embeddings adding a substitute of sentences. We used nlpaug library [50] in python, setting action type as insert and type as word2vec. Augmentation was done by inserting or replacing words in a sentence randomly leveraged by word2vec similarity search. For example, the sentence "The quick brown fox jumps over the lazy dog" was augmented to "The quick brown fox jumps Alzeari over the lazy Superintendents dog". The augmented dataset was used only for task 3a because the classes of task 3a were not balanced. As shown in Table 3 the "false" class has a significantly higher number of instances, hence, we applied augmentation for other classes. Table 5 shows the dataset statistics before and after augmentation.

Table 5
Dataset statistics before and after augmentation

Augmentation	Size	Train set	Development set	Test set
Before	900	80%	10%	10%
After	1335	80%	10%	10%

4.3. Features Extraction

The setup for all the algorithms is consistent throughout, with the only difference being the augmented dataset for task 3a. The logistic regression, multi-layer perceptron, random forest and support vector machine performed well in the experiments. For all the machine learning algorithms, word n-gram features including uni-gram, bi-gram and tri-gram were used. Final results were concluded using tri-gram features and term frequency—inverse document frequency for all experiments.

5. Results

The best performing results were submitted for both tasks. For task 3a the logistic regression model and for task 3b multi-layer perceptron model was submitted in the competition. Table 6 shows the results of the development set which shows logistic regression outperforming in task 3a with the support vector machine being the close second. Multi-layer perceptron performed the best in task 3a while support vector machine has the second best results. Table 7 shows how the machine learning model performed in comparison to the top 5 results presented in the competition. Our model achieved 5th place in the challenge in task 3b while in task 3a we ranked 10th. The best performing model in task 3a achieved the F1-macro of 83.70 and had a significant difference compared to our scores. While on task 3b machine learning models showed noteworthy results with 78.96 F1-macro.

Table 6
Results for task 3a and task 3b on the development set with n-gram features

Task	Model	Accuracy	Precision	Recall	F_1
Task 3a	LR	52.77	42.93	43.52	**41.96**
	MLP	53.88	41.07	37.30	37.17
	RF	56.44	42.24	32.86	31.30
	SVM	58.33	45.77	40.45	40.02
Task 3b	LR	79.22	78.35	71.96	72.97
	MLP	83.02	86.08	75.74	**78.35**
	RF	69.17	85.07	55.01	59.40
	SVM	77.96	87.56	69.03	73.64

6. Conclusion

In this paper, we analysed various machine learning algorithms to obtain the best F_1 for fake news claim classification and topic classification. Our results show that machine learning models with n-gram features are capable of competing albeit with limitations. The augmented dataset used for task 3a could not improve the results as the insertion of alternative words was not beneficial. Our model for task 3b achieved noteworthy results and we were placed fifth in the ranks for task 3b with 78.96 F1-macro.

Table 7
Comparison with top 5 results in the competition

Team name	F_1-macro
sushmakumari	83.76
Saud	51.42
kannanrrk	50.34
jmartinez595	46.80
hariharanrl	44.88
CIC	**38.92**

(a) Task 3a

Team name	F_1-macro
hariharanrl	88.13
sushmakumari	85.52
ninko	84.10
kannanrrk	81.78
CIC (5th ranked)	**78.96**

(b) Task 3b

Acknowledgments

The work was done with partial support from the Mexican Government through the grant A1-S-47854 of the CONACYT, Mexico and grants 20211784, 20211884, and 20211178 of the Secretaría de Investigación y Posgrado of the Instituto Politécnico Nacional, Mexico. The authors thank the CONACYT for the computing resources brought to them through the Plataforma de Aprendizaje Profundo para Tecnologías del Lenguaje of the Laboratorio de Supercómputo of the INAOE, Mexico.

References

[1] I. Ameer, N. Ashraf, G. Sidorov, H. G. Adorno, Multi-label emotion classification using content-based features in Twitter, Computación y Sistemas 24 (2021).
[2] L. Khan, A. Amjad, N. Ashraf, H.-T. Chang, A. Gelbukh, Urdu sentiment analysis with deep learning methods, IEEE Access 9 (2021) 97803–97812. doi:10.1109/ACCESS.2021.3093078.
[3] F. Bashir, N. Ashraf, A. Yaqoob, A. Rafiq, R. U. Mustafa, Human aggressiveness and reactions towards uncertain decisions, International Journal of ADVANCED AND APPLIED SCIENCES 6 (2019) 112–116.
[4] S. Butt, N. Ashraf, M. H. F. Siddiqui, G. Sidorov, A. Gelbukh, Transformer-based extractive social media question answering on TweetQA, Computación y Sistemas 25 (2021).
[5] N. Ashraf, R. Mustafa, G. Sidorov, A. Gelbukh, Individual vs. group violent threats classification in online discussions, in: Companion Proceedings of the Web Conference 2020, WWW '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 629–633.
[6] S. Butt, N. Ashraf, G. Sidorov, A. Gelbukh, Sexism identification using BERT and data augmentation - EXIST2021, in: International Conference of the Spanish Society for Natural Language Processing SEPLN 2021, IberLEF 2021, Spain, 2021.
[7] R. U. Mustafa, N. Ashraf, F. S. Ahmed, J. Ferzund, B. Shahzad, A. Gelbukh, A multiclass depression detection in social media based on sentiment analysis, in: 17th International Conference on Information Technology–New Generations (ITNG 2020), Springer International Publishing, Cham, 2020, pp. 659–662.

[8] A. Zubiaga, M. Liakata, R. Procter, Exploiting context for rumour detection in social media, in: International Conference on Social Informatics, Springer, 2017, pp. 109–123.
[9] L. Derczynski, K. Bontcheva, M. Liakata, R. Procter, G. W. S. Hoi, A. Zubiaga, Semeval-2017 task 8: Rumoureval: Determining rumour veracity and support for rumours, arXiv preprint arXiv:1704.05972 (2017).
[10] N. Hassan, F. Arslan, C. Li, M. Tremayne, Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, pp. 1803–1812.
[11] J. Thorne, A. Vlachos, C. Christodoulopoulos, A. Mittal, Fever: a large-scale dataset for fact extraction and verification, arXiv preprint arXiv:1803.05355 (2018).
[12] Rubin, V. L, Chen, Yimin, Conroy, N. K, Deception detection for news: Three types of fakes, Proceedings of the Association for Information Science and Technology 52 (2015) 1–4.
[13] S. Feng, R. Banerjee, Y. Choi, Syntactic stylometry for deception detection, in: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2012, pp. 171–175.
[14] A. K. Chaudhry, D. Baker, P. Thun-Hohenstein, Stance detection for the fake news challenge: Identifying textual relationships with deep neural nets, CS224n: Natural Language Processing with Deep Learning (2017).
[15] B. Ghanem, P. Rosso, F. Rangel, Stance detection in fake news a combined feature representation, in: Proceedings of the first workshop on fact extraction and VERification (FEVER), 2018, pp. 66–71.
[16] I. Augenstein, T. Rocktäschel, A. Vlachos, K. Bontcheva, Stance detection with bidirectional conditional encoding, arXiv preprint arXiv:1606.05464 (2016).
[17] C. Burfoot, T. Baldwin, Automatic satire detection: Are you having a laugh?, in: Proceedings of the ACL-IJCNLP 2009 conference short papers, 2009, pp. 161–164.
[18] A. N. Reganti, T. Maheshwari, U. Kumar, A. Das, R. Bajpai, Modeling satire in english text for automatic detection, in: 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), IEEE, 2016, pp. 970–977.
[19] C. Hansen, C. Hansen, S. Alstrup, J. Grue Simonsen, C. Lioma, Neural check-worthiness ranking with weak supervision: Finding sentences for fact-checking, in: Companion Proceedings of the 2019 World Wide Web Conference, 2019, pp. 994–1000.
[20] P. Nakov, G. Da San Martino, T. Elsayed, A. Barrón-Cedeño, R. Míguez, S. Shaar, F. Alam, F. Haouari, M. Hasanain, N. Babulkov, A. Nikolov, G. K. Shahi, J. M. Struß, T. Mandl, The CLEF-2021 CheckThat! lab on detecting check-worthy claims, previously fact-checked claims, and fake news, in: Proceedings of the 43rd European Conference on Information Retrieval, ECIR '21, Lucca, Italy, 2021, pp. 639–649.
[21] P. Nakov, G. Da San Martino, T. Elsayed, A. Barrón-Cedeño, R. Míguez, S. Shaar, F. Alam, F. Haouari, M. Hasanain, N. Babulkov, A. Nikolov, G. K. Shahi, J. M. Struß, T. Mandl, S. Modha, M. Kutlu, Y. S. Kartal, Overview of the CLEF-2021 CheckThat! Lab on Detecting Check-Worthy Claims, Previously Fact-Checked Claims, and Fake News, in: Proceedings of the 12th International Conference of the CLEF Association: Information Access Evaluation Meets Multiliguality, Multimodality, and Visualization, CLEF '2021, Bucharest, Romania (online), 2021.

[22] G. K. Shahi, J. M. Struß, T. Mandl, Overview of the CLEF-2021 CheckThat! lab task 3 on fake news detection, in: Working Notes of CLEF 2021—Conference and Labs of the Evaluation Forum, CLEF '2021, Bucharest, Romania (online), 2021.

[23] G. K. Shahi, J. M. Struß, T. Mandl, Task 3: Fake news detection at CLEF-2021 CheckThat!, 2021. URL: https://doi.org/10.5281/zenodo.4714517. doi:10.5281/zenodo.4714517.

[24] A. Asudeh, H. V. Jagadish, Y. Wu, C. Yu, On detecting cherry-picked trendlines, Proceedings of the VLDB Endowment 13 (2020) 939–952.

[25] V. F. Hendricks, M. Vestergaard, Alternative facts, misinformation, and fake news, in: Reality Lost, Springer, 2019, pp. 49–77.

[26] A. Chakraborty, B. Paranjape, S. Kakarla, N. Ganguly, Stop clickbait: Detecting and preventing clickbaits in online news media, in: 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), IEEE, 2016, pp. 9–16.

[27] Y. Chen, N. J. Conroy, V. L. Rubin, Misleading online content: recognizing clickbait as" false news", in: Proceedings of the 2015 ACM on workshop on multimodal deception detection, 2015, pp. 15–19.

[28] M. Potthast, S. Köpsel, B. Stein, M. Hagen, Clickbait detection, in: European Conference on Information Retrieval, Springer, 2016, pp. 810–817.

[29] J. Kiesel, M. Mestre, R. Shukla, E. Vincent, P. Adineh, D. Corney, B. Stein, M. Potthast, Semeval-2019 task 4: Hyperpartisan news detection, in: Proceedings of the 13th International Workshop on Semantic Evaluation, 2019, pp. 829–839.

[30] Y. Jiang, J. Petrak, X. Song, K. Bontcheva, D. Maynard, Team bertha von suttner at semeval-2019 task 4: Hyperpartisan news detection using elmo sentence representation convolutional network, in: Proceedings of the 13th International Workshop on Semantic Evaluation, 2019, pp. 840–844.

[31] G. L. Ciampaglia, P. Shiralkar, L. M. Rocha, J. Bollen, F. Menczer, A. Flammini, Computational fact checking from knowledge networks, PloS one 10 (2015) e0128193.

[32] S. Oraby, L. Reed, R. Compton, E. Riloff, M. Walker, S. Whittaker, And that's a fact: Distinguishing factual and emotional argumentation in online dialogue, arXiv preprint arXiv:1709.05295 (2017).

[33] Y. Yang, L. Zheng, J. Zhang, Q. Cui, Z. Li, P. S. Yu, Ti-cnn: Convolutional neural networks for fake news detection, arXiv preprint arXiv:1806.00749 (2018).

[34] J. Zhang, L. Cui, Y. Fu, F. B. Gouza, Fake news detection with deep diffusive network model, arXiv preprint arXiv:1805.08751 (2018).

[35] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, Y. Choi, Truth of varying shades: Analyzing language in fake news and political fact-checking, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 2931–2937.

[36] D. Fallis, A functional analysis of disinformation, iConference 2014 Proceedings (2014).

[37] E. C. Tandoc Jr, Z. W. Lim, R. Ling, Defining "fake news" a typology of scholarly definitions, Digital journalism 6 (2018) 137–153.

[38] G. Bhatt, A. Sharma, S. Sharma, A. Nagpal, B. Raman, A. Mittal, Combining neural, statistical and external features for fake news stance identification, in: Companion Proceedings of the The Web Conference 2018, 2018, pp. 1353–1357.

[39] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, R. Mihalcea, Automatic detection of fake news, arXiv preprint arXiv:1708.07104 (2017).

[40] J. Pennington, R. Socher, C. D. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 Conference on Rmpirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1532–1543.
[41] V. W. Feng, G. Hirst, Detecting deceptive opinions with profile compatibility, in: Proceedings of the Sixth International Joint Conference on Natural Language Processing, 2013, pp. 338–346.
[42] H. Zhang, Z. Fan, J. Zheng, Q. Liu, An improving deception detection method in computer-mediated communication, Journal of Networks 7 (2012) 1811.
[43] E. Tacchini, G. Ballarin, M. L. D. Vedova, S. Moret, L. de Alfaro, Some like it hoax: Automated fake news detection in social networks, 2017. arXiv:1704.07506.
[44] S. Gilda, Notice of violation of ieee publication principles: Evaluating machine learning algorithms for fake news detection, in: 2017 IEEE 15th Student Conference on Research and Development (SCOReD), 2017, pp. 110–115. doi:10.1109/SCORED.2017.8305411.
[45] J. Ma, W. Gao, K.-F. Wong, Rumor detection on Twitter with tree-structured recursive neural networks, Association for Computational Linguistics, 2018.
[46] N. Ruchansky, S. Seo, Y. Liu, Csi: A hybrid deep model for fake news detection, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017, pp. 797–806.
[47] Y. Liu, Y.-F. Wu, Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 32, 2018.
[48] G. Sidorov, S. Miranda-Jiménez, F. Viveros-Jiménez, A. Gelbukh, N. Castro-Sánchez, F. Velásquez, I. Díaz-Rangel, S. Suárez-Guerra, A. Trevino, J. Gordon, Empirical study of machine learning based approach for opinion mining in tweets, in: Mexican International Conference on Artificial Intelligence, Springer, 2012, pp. 1–14.
[49] C. Baziotis, N. Pelekis, C. Doulkeridis, Datastories at semeval-2017 task 4: Deep lstm with attention for message-level and topic-based sentiment analysis, in: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), Association for Computational Linguistics, Vancouver, Canada, 2017, pp. 747–754.
[50] E. Ma, Nlp augmentation, https://github.com/makcedward/nlpaug, 2019.