Plagiarism Detection with Paraphrase Recognizer Using Deep Learning

Yogesh Wadekar[1], Tushar Shendge[2], Manali Dhokale[3], Vaishnavi Ohol[4], Prof. Sagar Dhanake[5]

Students, Computer Engineering[1, 2, 3,4]

Assistant Professor, Computer Engineering[5]

D Y Patil Institute of Engineering and Technology, Ambi, Pune, Maharashtra, India

Abstract: *Plagiarism is a progressively widespread and growing issue within the educational field. Many plagiarism techniques square measure utilized by fraudsters, starting from a straightforward word replacement, phrase structure modification, to additional advanced techniques involving many varieties of transformation. Primarily human-based plagiarism detection is troublesome, not much accurate, and time-consuming method. In this paper, we tend to propose a plagiarism detection framework supported by 3 deep learning models: Doc2vec, Siamese Long Short-term Memory (SLSTM), and Convolutional Neural Network. Our system uses 3 layers: Preprocessing Layer together with word embedding, Learning Layers, and Detection Layer. To judge our system, we tend to dispense a study on plagiarism detection tools from the educational field and build a comparison supported a group of options. Compared to alternative works, our approach performs an honest accuracy of 97.26% and might notice differing kinds of plagiarism, permits to specify another dataset, and supports to check the document from an internet search. Plagiarism is a progressively widespread and growing issue within the educational field. Many plagiarism techniques square measure utilized by fraudsters, starting from a straightforward word replacement, phrase structure modification, to additional advanced techniques involving many varieties of transformation. Primarily human-based plagiarism detection is troublesome, not much accurate, and time-consuming method. In this paper, we tend to propose a plagiarism detection framework supported by 3 deep learning models: Doc2vec, Siamese Long Short-term Memory (SLSTM), and Convolutional Neural Network. Our system uses 3 layers: Preprocessing Layer together with word embedding, Learning Layers, and Detection Layer. To judge our system, we tend to dispense a study on plagiarism detection tools from the educational field and build a comparison supported a group of options. Compared to alternative works, our approach performs an honest accuracy of 97.26% and might notice differing kinds of plagiarism, permits to specify another dataset, and supports to check the document from an internet search.*

Keywords: Plagiarism detection, Plagiarism detection tools, Deep learning, Doc2vec, Stacked Long Short-Term Memory (SLSTM), Convolutional Neural Network (CNN), Siamese neural network

REFERENCES

[1] El Mostafa Hambi , Faouzia Benabbou, "A New Online Plagiarism Detection System based on Deep Learning", (IJACSA) International Journal of Advanced Computer Science and Applications Vol. 11, No. 9, 2020.

[2] Qiubo Huang, Xuezhi Song, Xuezhi Song, "Code Plagiarism Detection Method Based on Code Similarity and Student Behavior Characteristics", IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), 2020

[3] Ethan Hunt, Ritvik Janamsetty and team, "Machine Learning Models for Paraphrase Identification and its Applications on Plagiarism Detection", IEEE International Conference on Big Knowledge (ICBK) 2019

[4] Yahia Jazyah, "Open Learning, the Issue of Plagiarism - Efficient Algorithm", International Journal of Computers, Volume 3, 2018

[5] Chitra and Anupriya Rajkumar, "Plagiarism Detection Using Machine Learning- Based Paraphrase Recognizer", J. Intell. Syst. 2016; 25(3): 351–359

[6] Kim, Do-Guk Ko, Bonggyun. (2019). Investment Universe Construction Based on the Theme Keyword Search. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2939414.

[7] Janardhanan, Deepak Barrett, Enda. (2018). CPU Work load forecasting of Machines in Data Centers using LSTM Recurrent Neural Networks and ARIMA Models. 10.23919/ICITST.2017.8356346.

[8] Miralles, Luis Rosso, Dafne Jim´enez, Fernando Garc´ıa, Jos´e. (2017). A methodology based on Deep Learning for advert value calculation in CPM, CPC and CPA networks. Soft Computing. 21. 1-15. 10.1007/s00500-016-2468-4.

[9] Risquez, A., Dwyer, M. O.; Ledwith, A. (2011). "Thou shalt not plagiarize': from self-reported views to recognition and avoidance of plagiarism". Assessment Evaluation in Higher Education, vol. 2, no. 1, p. 34-43. http://doi.org/10.1080/02602938.2011.596926. 3 Ruip´erez, G.; Garc´ıa-Cabrero, J.C. (2016). Plagiarism and Academic Integrity in Germany. Comunicar, vol. 24, no. 48, p. 9-17. http://doi.org/10.3916/C48-2016-01.

[10] Suleiman, D., Awajan, A., Al-Madi, N. (2017). "Deep Learning Based Technique for Plagiarism Detection in Arabic Texts". 2017 International Conference on New Trends in Computing Sciences (ICTCS). doi:10.1109/ictcs.2017.42

[11] Liu, Y.-T., Zhang, H.-R., Chen, T.-W., Teng, W.-G. (2007). "Extending Web Search for Online Plagiarism Detection". 2007 IEEE International Conference on Information Reuse and Integration. doi:10.1109/iri.2007.4296615

[12] Sousa-silva, r. -"detecting trans lingual plagiarism and the backlash against translation plagiarists language and law" / linguagem e direito, vol. 1(1), 2014, p. 70-94.