6th International Conference on Nanotechnologies and Biomedical Engineering
Proceedings of ICNBME-2023, September 20–23, 2023, Chisinau, Moldova
Volume 2: Biomedical Engineering and New Technologies for Diagnosis, Treatment, and Rehabilitation

NLP Tools for Epileptic Seizure Prediction Using EEG Data: A Comparative Study of Three ML Models

Victor Iapascurta, Ion Fiodorov

https://doi.org/10.1007/978-3-031-42782-4_19

Abstract

Natural Language Processing (NLP) is an ever-evolving field of computer science that involves the development of algorithms that can process, analyze and understand human language. One of the most exciting areas of NLP is the creation of NLP language models with applications across almost every industry. However, most people only associate NLP with its traditional use in language translation, sentiment analysis, and chatbots. In reality, there are many less-common uses for NLP models that have the potential to transform businesses, improve customer experiences, and even save lives. In the healthcare industry, NLP models can be used to analyze unstructured medical data and help diagnose and treat patients more efficiently. For example, NLP can be used to analyze clinical notes, lab results, and other data combing through vast amounts of data to identify patterns and create targeted treatment plans. NLP-based medical diagnosis is still in its infancy, but it has the potential to revolutionize the healthcare industry in the coming years. This article explores a less common use of machine-learning language models built on transformed EEG data for epilepsy prediction using the Kolmogorov-Chaitin algorithmic complexity as the first step in generating text-like data, which are finally used for building machine learning models.

Keywords: natural language processing, machine learning, algorithmic complexity, epileptic seizure prediction.

6th International Conference on Nanotechnologies and Biomedical Engineering
Proceedings of ICNBME-2023, September 20–23, 2023, Chisinau, Moldova
Volume 2: Biomedical Engineering and New Technologies for Diagnosis, Treatment, and Rehabilitation

References

1. Morin, F., Bengio, Y.: Hierarchical probabilistic neural network language model. In: Proceedings of the 10th International Workshop on AI and Statistics. Proceedings of Machine Learning Research, vol. R5, pp. 246–252 (2005)
2. Mnih, A., Hinton, G.: A scalable hierarchical distributed language model. Adv. Neural Inf. Process. Syst. **21** (NISP 2008), Curran Associates Inc. 21, 1081–1088 (2009)
3. Harrison, C.J., Sidey-Gibbons, C.J.: Machine learning in medicine: a practical introduction to natural language processing. BMC Med. Res. Methodol. **21**, 158 (2021). https://doi.org/10.1186/s12874-021-01347-1
4. Kormilitzin, A. et al.: Med7: a transferable clinical natural language processing model for electronic health records. arXiv:2003.01271v2 [cs.CL] (2020)
5. Sutton, R.T. et al.: An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit. Medi. **3**, 17 (2020). https://doi.org/10.1038/s41746-020-0221-y
6. Sezgin, E., et al.: Extracting Medical Information from free-text and unstructured patientgenerated health data using natural language processing methods: a feasibility study with real-world data. JMIR Preprints, 24 (2022). https://preprints.jmir.org/preprint/43014
7. Funk, B., et al.: A framework for applying natural language processing in digital health interventions. J Med Internet Res **22**(2), e13855 (2020). https://doi.org/10.2196/13855
8. Beghi, E., et al.: Global, regional, and national burden of epilepsy, 1990–2016: a systematic analysis for the global burden of disease study 2016. Lancet Neurol. **18**(4), 357–375 (2019)
9. Wang, Z., Mengoni, P.: Seizure classification with selected frequency bands and EEG montages: a natural language processing approach. Brain Inform. **9**, 11 (2022). https://doi.org/10.1186/s40708-022-00159-3
10. Yew, A., et al.: Transforming epilepsy research: a systematic review on natural languageproce ssing applications. Epilepsia **00**, 1–14 (2022). https://doi.org/10.1111/epi.17474
11. Mikolov, T., et al.: Efficient Estimation of Word Representations in Vector Space. arXiv: 1301.3781v3 (2013)
12. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. Proceedings of Advances Neural Information Processing System Nevada, NV USA, pp. 3111–3119 (2013)
13. Xu, G., et al.: Sentiment analysis of comment text based on BiLSTM. IEEE Access **7**(2019), 51522–51532 (2019)
14. Devlin, J., Chang, M-W., Lee, K., Toutanova, K.: BERT: Pretraining of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805v2 (2019)
15. American Epilepsy Society Seizure PredictionChallenge (2014).www.kaggle.com/c/seizureprediction
16. Zenil, H.: Towards demystifying shannon entropy, lossless compression, and approaches to statistical machine learning. In: Proceedings of the International Society for Information Studies 2019 summit, University of California, Berkeley, vol. 47, p. 24 (2020). doi: https://doi.org/10.3390/proceedings2020047024
17. Zenil, H., et al.: The online algorithmic complexity calculator (OACC) v3.0 (2018). Algorithmic dynamics lab, science for life laboratory (SciLifeLab), unit of computational medicine. Center Mol. Med. Karolinska Inst. Stockholm, Sweden. www.algorithmicdynamics.net/software.html

6th International Conference on Nanotechnologies and Biomedical Engineering
Proceedings of ICNBME-2023, September 20–23, 2023, Chisinau, Moldova
Volume 2: Biomedical Engineering and New Technologies for Diagnosis, Treatment, and Rehabilitation

18. Iapascurta, V.: Combining algorithmic information dynamics concepts and machine learning for electroencephalography analysis:what can we get? Complex Syst. **31**(4), 389–413 (2022). https://doi.org/10.25088/ComplexSystems.31.4.389
19. H2O.ai. h2o: R Interface for H2O. R package version 3.32.1.1 (2020)