

EHR(Electronic Health Record), Clinical Assertion modeling of COVID patients with NLP techniques

Samin Islam, Nazmus Sakib Saad,

Md. Sabbir Hossain, Annajiat Alim Rasel
Department of Computer Science and Engineering,
BRAC University Dhaka, Bangladesh

Abstract

Applications of NLP(Natural Language Processing) in the healthcare system is a fast-growing scenario to analyze, interpret and extract enormous patient information. EHRs are the collections of a patient's medical history which is poorly annotated and highly unstructured data, and NLP techniques (OCR- Optical Character Recognition, NER - Named Entity Recognition, Clinical Assertion Detection modeling etc) can be incorporated to introduce automation in health care. The global pandemic COVID-19 has shaken up the entire world with the anonymity of the Novel Coronavirus. Scientists across the globe have been studying this catastrophe, gauging different aspects (nature, social factors, bio-chemical, comorbidities, risk factors of medical history) that might increase the death toll due to this infectious disease. These numerous published articles and journals can be traced down for literature review in an efficient way with the help of automated text-mining in NLP. Starting with converting the data into structured data, extract meaningful medical jargon, word embeddings extract information. A BiLSTMCNN based approach has been used in (Kocaman and Talby 2020) on the COVID-19 Open Research Dataset (Wang et al. 2020) (CORD-19), where the researcher has built a text-mining system keeping three things in mind; recognizing different entities, assertion status detection and using pre-trained NER models. As the rapid development in the world of NLP has been conspicuous; we, following (Kocaman and Talby 2020) are going to introduce BERT (Bidirectional Encoder Representations from Transformers), a pre-trained language model on the CORD-19(Wang et al. 2020) dataset for efficiency. We hope our analysis on this large scale of unstructured data will give a better result with higher accuracy and faster computation, even though this model is quite expensive.

Al-Aiad, A.; Duwairi, R.; Fraihat, M. 1 Oct 2018. Survey: deep learning concepts and techniques for electronic health record. In 2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA) (pp. 1-5). IEEE.

Alimova, I. and Tutubalina, E. 1 Mar 2020. Multiple features for clinical relation extraction: a machine learning approach. Journal of biomedical informatics.;103:103382.

Assale, M.; Dui, LG.; Cina, A.; Seveso, A.; Cabitza, F. 17 Apr 2019. The revival of the notes field: leveraging the unstructured content in electronic health records. Frontiers in medicine.;6:66.

Doğan, RI.; Leaman, R.; Lu, Z. 1 Feb 2014. NCBI disease corpus: a resource for disease name recognition and concept normalization. Journal of biomedical informatics.;47:1-0.

Kocaman, V. and Talby, D. 7 Dec 2020. Improving clinical document understanding on COVID-19 research with spark NLP.

Li, I.; Pan, J.; Goldwasser, J.; Verma, N.; Wong, WP.; Nuzumlal, MY.; Rosand, B.; Li, Y.; Zhang, M.; Chang, D.; Taylor, RA. 7 Jul 2021. Neural Natural Language Processing for Unstructured Data in Electronic Health Records: a Review..

Wang, LL.; Lo, K.; Chandrasekhar, Y.; Reas, R.; Yang, J.; Eide, D.; Funk, K.; Kinney, R.; Liu, Z.; Merrill, W.; Mooney, P. 9 Jul 2020. Cord-19: The covid-19 open research dataset.

References

Ahamed, S. and Samad M. 5 Apr 2020. Information mining for covid-19 research from a large volume of scientific literature.

Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.