

COVID_ICU_BERT: A Fine-Tuned Language Model for COVID-19 Intensive Care Unit Clinical Notes

Authors : Shahad Nagoor, Lucy Hederman, Kevin Koidl, Annalina Caputo

Abstract : Doctors' notes reflect their impressions, attitudes, clinical sense, and opinions about patients' conditions and progress, and other information that is essential for doctors' daily clinical decisions. Despite their value, clinical notes are insufficiently researched within the language processing community. Automatically extracting information from unstructured text data is known to be a difficult task as opposed to dealing with structured information such as vital physiological signs, images, and laboratory results. The aim of this research is to investigate how Natural Language Processing (NLP) techniques and machine learning techniques applied to clinician notes can assist in doctors' decision-making in Intensive Care Unit (ICU) for coronavirus disease 2019 (COVID-19) patients. The hypothesis is that clinical outcomes like survival or mortality can be useful in influencing the judgement of clinical sentiment in ICU clinical notes. This paper introduces two contributions: first, we introduce COVID_ICU_BERT, a fine-tuned version of clinical transformer models that can reliably predict clinical sentiment for notes of COVID patients in the ICU. We train the model on clinical notes for COVID-19 patients, a type of notes that were not previously seen by clinicalBERT, and Bio_Discharge_Summary_BERT. The model, which was based on clinicalBERT achieves higher predictive accuracy (Acc 93.33%, AUC 0.98, and precision 0.96). Second, we perform data augmentation using clinical contextual word embedding that is based on a pre-trained clinical model to balance the samples in each class in the data (survived vs. deceased patients). Data augmentation improves the accuracy of prediction slightly (Acc 96.67%, AUC 0.98, and precision 0.92).

Keywords : BERT fine-tuning, clinical sentiment, COVID-19, data augmentation

Conference Title : ICBNLP 2023 : International Conference on Biomedical Natural Language Processing

Conference Location : London, United Kingdom

Conference Dates : January 23-24, 2023