

Incorporating Lexical-Semantic Knowledge into Convolutional Neural Network Framework for Pediatric Disease Diagnosis

Authors : Xiaocong Liu, Huazhen Wang, Ting He, Xiaozheng Li, Weihang Zhang, Jian Chen

Abstract : The utilization of electronic medical record (EMR) data to establish the disease diagnosis model has become an important research content of biomedical informatics. Deep learning can automatically extract features from the massive data, which brings about breakthroughs in the study of EMR data. The challenge is that deep learning lacks semantic knowledge, which leads to impracticability in medical science. This research proposes a method of incorporating lexical-semantic knowledge from abundant entities into a convolutional neural network (CNN) framework for pediatric disease diagnosis. Firstly, medical terms are vectorized into Lexical Semantic Vectors (LSV), which are concatenated with the embedded word vectors of word2vec to enrich the feature representation. Secondly, the semantic distribution of medical terms serves as Semantic Decision Guide (SDG) for the optimization of deep learning models. The study evaluate the performance of LSV-SDG-CNN model on four kinds of Chinese EMR datasets. Additionally, CNN, LSV-CNN, and SDG-CNN are designed as baseline models for comparison. The experimental results show that LSV-SDG-CNN model outperforms baseline models on four kinds of Chinese EMR datasets. The best configuration of the model yielded an F1 score of 86.20%. The results clearly demonstrate that CNN has been effectively guided and optimized by lexical-semantic knowledge, and LSV-SDG-CNN model improves the disease classification accuracy with a clear margin.

Keywords : convolutional neural network, electronic medical record, feature representation, lexical semantics, semantic decision

Conference Title : ICBB 2021 : International Conference on Bioinformatics and Biomedicine

Conference Location : Istanbul, Türkiye

Conference Dates : June 28-29, 2021