Early Prediction of Sepsis Using a Sliding Window-based AdaBoost Learning and Bayesian Regression

Meicheng Yang, Hongxiang Gao, Xingyao Wang, Fan Zhou, Jianan Di, Tiantian Wang, Jianqing Li, Chengyu Liu

State Key Laboratory of Bioelectronics, Jiangsu Key Lab of Remote Measurement and Control, School of Instrument Science and Engineering, Southeast University, Nanjing, China

Sepsis is a life-threatening condition and has caused extensive concern. The PhysioNet/Computing in Cardiology Challenge 2019 focuses on the early prediction of sepsis from multi-model clinical data, aiming to reduce the sepsis severe morbidity, mortality and medical costs.

To address this challenge, we proposed a sliding window-based AdaBoost learning and Bayesian regression method to predict the onset of sepsis up to six hours in advance. First, balanced records from non-sepsis group were randomly selected from the training data, generating a total of 4,600 records with equal number records of sepsis and non-sepsis. Then, a six-hour sliding window was applied to segment each record with a step of one hour, resulting in a total of 220,647 six-hour segments. If sepsis event fell within the window, this segment was labelled as 'sepsis', otherwise no. Features were extracted from all available patient co-variates, as well as from their differential time series. Gentle AdaBoost learning model was trained for determining a six-hour segment as 'sepsis' or not. The 0/1 results for the six-hour segments were merged using a Bayesian regression method to form the final onset prediction of sepsis event. Evaluations were performed using a standard five-fold cross-validation.

Results reported an area under receiver operating characteristic (AUROC) curve of 0.787 for sepsis 0/1 classification on the six-hour segments. The overall score of Utility defined by the challenge organizer was 0.416 from the five-fold cross-validation. In addition, the maximum running time was only 19% of quota, indicating the computing efficiency of our proposed method.