

Novelty Detection for the Early Prediction of Sepsis

Oliver Carr, Stefan Bostock, Nicolas Basty, John Prince, Navin Cooray, Kirubin Pillay, Maarten De Vos

Institute of Biomedical Engineering, University of Oxford
Oxford, United Kingdom

Aims: Sepsis is a condition which can cause multiple organ failure due to infections in the body. It is one of the leading causes of death in patients who are critically ill, yet is difficult to diagnose as patients often have other underlying diseases. We aim to detect the early onset of sepsis using physiological measures as part of the Physionet/Computing in Cardiology Challenge 2019.

Methods: Physiological measures were provided for each hour of each patient's stay in hospital. Feature selection was performed by selecting measures with the fewest number of missing entries. Features that were highly correlated were removed resulting in selecting: heart rate, pulse oximetry, temperature, systolic blood pressure, respiration rate, and demographic measures. Due to a large data imbalance, novelty detection was chosen to identify sepsis as a deviation in the feature vector (at each hour) using a model of normality. This model utilized all non-sepsis data which was not within 16 hours of the onset of sepsis. K-means was applied, fitting 100 clusters to the data, and a probability density distribution then centred on each cluster to determine the probability of each feature vector fitting the model of normality. A novelty score was calculated as the negative log of this probability, with a global threshold chosen to label each feature vector as sepsis or non-sepsis.

Results and Discussion: The results of a 10-fold stratified cross-validation achieved an F1 score of 0.13 and a score of 0.34 using the utility function. The utility score for patients who had sepsis was 0.49 (from a possible 0 to 1) and for non-sepsis patients was -0.16 (from a possible -2 to 0). The model has been submitted for testing on the hidden test set. Future work will focus on further optimizing model parameters and incorporating a time dependency.

Mean (standard deviation) of utility and F₁ scores for patients who developed sepsis (Sepsis), did not develop sepsis (Non-Sepsis), and the combined results (after 10-fold stratified cross validation).

Novelty Method Features used (Data for normality)	Cross-Validation ($n = 40,336$)				Hidden Test Set	
	Sepsis	Non-Sepsis	Combined		Combined	
	Utility	Utility	F ₁	Utility	F ₁	Utility
Vital Signs (normality model from non-sepsis patients)	-0.11 (0.02)	0.24 (0.01)	0.08 (0.01)	0.13 (0.02)	N/A	N/A
Vital Signs & Demographics (normality model from non-sepsis patients)	-0.24 (0.01)	0.50 (0.02)	0.10 (0.01)	0.26 (0.03)	Pending	Pending
Vital Signs & Demographics (normality model non-sepsis and pre sepsis)	0.49 (0.02)	-0.16 (0.01)	0.13 (0.01)	0.34 (0.02)	Pending	Pending