Randomly Under Sampled Boosted Tree for Predicting Sepsis From Intensive Care Unit Databases

Peter Doggart[1], Megan Rutherford[1]

[1]B-Secur, Belfast, Northern Ireland

Abstract

Bacterial infection can result in sepsis; a toxic immune response by the body. Although the rate of mortality due to sepsis has fallen within the UK, overall rates remain higher than in Europe. Early detection of sepsis has been linked to elevated successful outcomes. This work focuses on the use of a Random Under Sample (RUS) Boosted Tree for classifying sepsis from intensive care unit databases.

Full training set (40,336 subjects) achieved sensitivity and specificity of 53.4% and 83.6% respectively (AUC:77.79%) with 5 fold cross-validation. In the unofficial phase of the challenge the model achieved a normalized utility score of 0.267. The model did not achieve a score on the full test set (team name B-Secur).

The results show that the model is capable of detecting sepsis in patients. However, there is more work to be done in order to improve performance. Future work will investigate the use of fixed time windows rather than individual hourly measurements to increase prediction performance.

1.   Introduction

Sepsis is defined as "life threatening organ dysfunction caused by dysregulated host response to infection".[1] It is estimated that approximately 5 million sepsis – related deaths occur annually around the world.[2] To reduce the global burden of sepsis, prevention and or early intervention are of importance. Mortality rates associated with Sepsis have been shown to increase with each hour of delayed treatment.[3]

Early goal directed therapy (EGDT) and 'Sepsis Six' bundles are care packages that have been developed to reduce the risk of mortality following the onset of sepsis.[4] They include actions such as oxygen and intravenous antibiotics administration. These bundles favor early intervention as illustrated by their recommended time frames for completion. For example, the 'Sepsis Six' approach must be completed within one hour of sepsis onset. Compliance with bundled care packages has been shown to reduce mortality by 24% when compared to standard practices.[5]

Clinical diagnosis of sepsis is based on the presence of an infection and a raised Sequential Organ Failure Assessment Score (SOFA).[4] However the tests required to compute the SOFA score may not be continuously attainable due to resource / staffing limitations and or due to a patient being sufficiently unwell. An opportunity exists to improve upon early sepsis detection based upon the data available to clinicians through machine learning. Early detection will optimize the use of sepsis care bundles and subsequently reduce the incidences of mortality resulting from Sepsis.

2.   Method

Vital sign measurements (e.g. heart rate, temperature, respiration etc.) were typically found to be continuously available within the training data. Conversely, laboratory results within the training data set were not found to be continuously available i.e. no hourly updates. Subsequently, the training data set contains a significant proportion of missing values. To alleviate the burden of missing data, our model utilizes a simple forward filling procedure, in which the current value, if missing, is replaced by the previous value; the current value remains unchanged if there is no previous entries. With respect to measures of blood pressure only (mean arterial pressure, MAP, systolic blood pressure , SBP, and diastolic blood pressure, DBP), after forward filling, missing measures where estimated using Equation 1 if at least two of the three measures were present at a given time interval.

$$MAP = \frac{SBP + 2DBP}{3} \qquad (1)$$

With respect to the model developed, variables of interest initially included all vital sign measurements, laboratory measurements and the ICU length of stay. Using all of the data from all of the patients within the training data set (after forward filling), present measures for each variable (e.g. heart rate from all subjects) were clustered into 7 discrete bins. Clustering was carried out using K-means clustering. K-means clustering was chosen over equal bin sizes as it allows non-linear spread of values to be represented more robustly. Seven clusters were chosen

as a tradeoff between subsequent classifier performance and interpretability of clusters (e.g. cluster 1 contains the lowest observed heart rates and cluster 7 contains the highest observed heart rates). An additional cluster (eight clusters in total) was utilised to categorise missing data. In the context of clinical care, missing values often provide additional information about a patients or doctors' intuition i.e. the frequency at which laboratory tests were carried out and can improve prediction performance.

The training data set is unbalanced; there are many more non-septic than septic subjects. This can lead to bias when building machine learning models in favour of the biggest class; in this instance non-septic subjects. To prevent a model that is biased towards labelling people as non – septic, Random Undersampling Boosting (RUSBoost) and a decision tree ensemble classifier were employed. RUSBoost is a boosting algorithm specifically designed to be effective in classifying unbalanced data. In this work, the RUSBoost algorithm takes the total number of septic members (n_{sepsis}) within the training data as the basic unit for sampling. For each training iteration of the ensemble decision tree, the non-sepsis class is under sampled by randomly taking only n_{sepsis} observations. Each decision tree or weak learner is therefore trained on data which contains the same number of observations for both classes, removing classification bias. Each subsequent weak learner is trained to try to correct the misclassifications of its predecessor, leading to increased overall classification accuracy. A decision tree ensemble with 200 learners was utilised.

The model (Figure 1) was trained / assessed using 5-fold cross validation on the provided data. The model was set up for classification such that it was only dependant on current measures and not historical values. Typically, cross validation would be done once all the data had been concatenated into one global dataset. However, due to the use of forward filling, this represents a potential for subsequent test data sets to be contaminated with training data. To prevent this, all data files were initially sorted into two categories; septic and non-septic. Cross validation was then performed on the indexes of the septic and non-septic files independently. This insured that data from one patient could not be within a training and a test data set. Each of the corresponding folds for septic and non-septic patients were then combined. The data used in building / testing the model was supplied by the 2019 PhysioNet Challenge.[6] The optimum threshold for use with the model for sepsis detection was found via use of a receiver operating characteristic curve (ROC, area under the curve; AUC). Performance of the model was assessed via the PhysioNet utility score and a confusion matrix (sensitivity, specificity).

3. Results

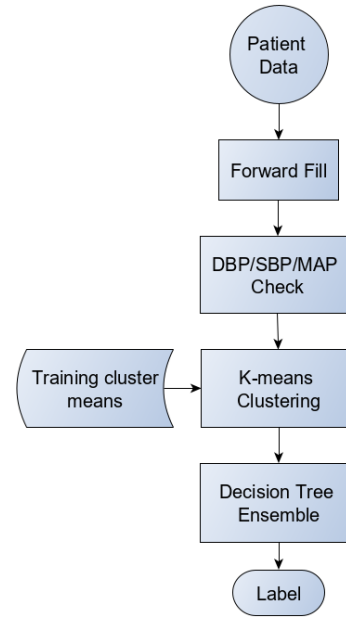With respect to the ROC curve, the optimum sepsis

Figure 1. Flow diagram representing the process of classification with new patient data.
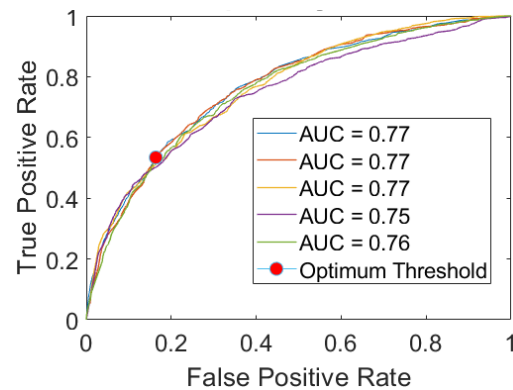

Figure 2. ROC representing the optimum sepsis threshold across each of the test folds.

Figure 3. Confusion matrix representing the mean classifications made by the Randomly Under Sampled Boosted Tree across all test folds

threshold across all folds was found to provide an AUC of 77.8%, Figure 2. Overall, as illustrated in Figure 2, the performance of the model across test folds was found to be stable. The model was found to have an average sensitivity of 53.4% and an average specificity of 83.6% across test folds, Figure 3. The mean utility score during cross validation was 0.267.

4.   Conclusion

The results show that the model is capable of detecting sepsis in patients. However, there is substantial room for improvement. One of the key challenges with this data was the lack of additional information supplied about additional care that the patients were undergoing. For example, are measures of respiration rate natural or those of a ventilator machine. The care / different treatments which patients are undergoing may adversely impact the ability to detect sepsis without the ability to use them as controlled variables. Future work will investigate the use of fixed time windows rather than individual hourly measurements to increase prediction performance.

References

[1] M. Singer et al., "The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)," *JAMA*, vol. 315, no. 8, pp. 801-810, Feb. 2016.

[2] C. Fleischmann et al., "Assessment of Global Incidence and Mortality of Hospital-Treated Sepsis. Current Estimates and Limitations," *Am. J. Respir. Crit. Care Med.*, vol. 193, no. 3, pp. 259-272, Feb. 2016.

[3] C. W. Seymour et al., "Time to Treatment and Mortality During Mandated Emergency Care for Sepsis," *N. Engl. J. Med.*, vol. 376, no. 23, pp.2235-2244, May. 2017.

[4] A Keeley et al., "The Recognition and Management of Sepsis and Septic Shock: A Guide for Non-Intensivists," *Postgrad. Med. J.*, vol. 93, no. 1104, pp. 626-634, Jul. 2017.

[5] A. V. Barochia et al., "Bundled Care for Septic Shock: An Analysis of Clinical Trials," *Crit. Care. Med.*, vol. 38, no. 2, pp. 668-678, Feb. 2010.

[6] M. Reyna et al., "Early Prediction of Sepsis from Clinical Data: the PhysioNet/Computing in Cardiology Challenge 2019". Critical Care Medicine 2019; In press.

Address for correspondence:

Peter Doggart.
B-Secur, Catalyst Inc, 20 Queens Rd, Belfast, Northern Ireland, BT3 9DT.
peter.doggart@b-secur.com