Early Detection of Sepsis Using Feature Selection, Feature Extraction, and Neural Network Classification

Erik H Gilbertson, Khristian M Jones, Abigail M Stroh, Bradley M Whitaker

Electrical & Computer Engineering Department, Montana State University, Bozeman, MT, USA

Abstract

Introduction: This work represents an entry to the 2019 PhysioNET Computing in Cardiology Challenge.

Algorithm: Using the supplied biomedical data, we reduce the original 40 features to 10 principal components. One additional feature is generated from a quick Sequential Organ Failure Assessment (qSOFA) score. These 11 features are then fed into a deep neural network classifier implemented in Tensorflow. The features associated with each hour are analyzed independently. A sigmoid function is used for the activation functions, RMSprop for the optimizer functions, and categorical cross entropy for the loss function. We have found that this setup works best for our current method and leads to the highest accuracy with minimal loss.

Results: By testing our algorithm on a subset of the given dataset we achieved a validation score of 0.038. The official score received from the competition (under the team name "Whitaker's Warriors") was 0.022 placing us 65th out of the 78 scored entries.

Conclusions: Achieving a positive utility score of 0.022 shows our method of combining PCA with a quick SOFA score and classifying with a neural network is a promising approach. More work in the future could be done to increase the accuracy of the model by adding additional features to the input of the classifier and adjusting the parameters of the neural network.

1. Introduction

This work represents an entry to the 2019 PhysioNET Computing in Cardiology Challenge. The goal of the challenge was to accurately predict sepsis in clinical patients using 40 biological features recorded on an hourly basis. Details on the Challenge can be found at `https://physionet.org/challenge/2019/`. The database we used in conjunction with the Challenge is described in detail in [1]. The Challenge website and paper, and the citations therein, describe the scope of sepsis and outline the importance of early diagnosis and treatment [1–5].

For our solution we used a combination of feature learning using principal component analysis (PCA) and manual feature extraction on the data before it was input into a neural network classifier [6, 7]. This approach allowed our code to be lightweight because we reduced the dimensionality of the feature space. While our main goal was to achieve a high classification training accuracy, we also focused on minimizing model loss to prevent overfitting the classifier to the training data.

2. Algorithm

2.1. Preprocessing

To work with the data we first had to find a way to deal with the Not a Number (NaN) values that were inherent in the clinical data. These values are present when a certain biological feature was not measured during a given hour. We decided to replace the NaN entries in the first hour of data with zeros. When a value was recorded for one of the features, it would be used as a placeholder value for subsequent hours until a new value was recorded. This resulted in each hour having the most up-to-date measurements for the patients, even if the measurements were not taken at that particular time.

After removing the NaN entries, each feature was normalized to have a standard deviation of one and a mean of zero. The data was normalized for each patient every time an additional hour was added. This allowed us to have consistent data to input into our program, and to ensure that no feature was dominant in the analysis.

2.2. Feature Extraction

The goal of automatic feature extraction is to manipulate the data in such a way that the most salient information is represented in as few numbers as possible. This results in reducing the dimensionality of the data, which in turn reduces the complexity and training time of a classifier, while preserving the discriminating information [8].

For this work, we used PCA for feature extraction. PCA was originally developed in 1933, and continues to be
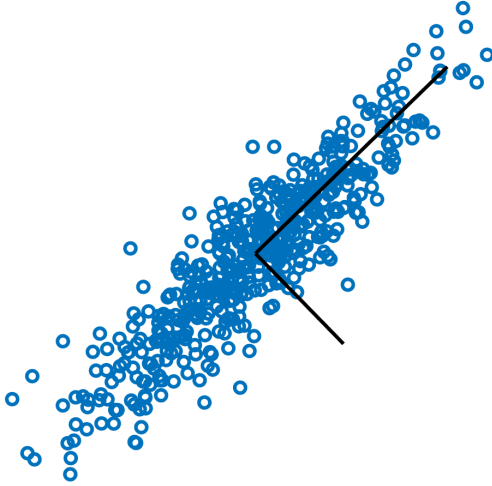
Figure 1. Visual representation of PCA in two dimensions. The overall dimension can be reduced by projecting data points on to the subspace spanned by a subset of the principal components. While some information is lost, much of the explained variance is preserved using just a few principal components.
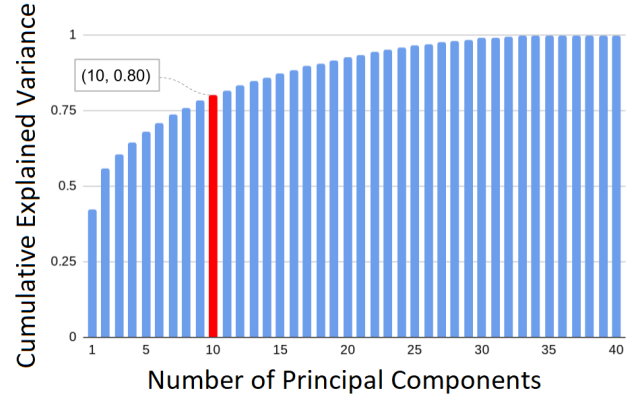

Figure 2. A bar chart displaying the accumulated explained variance for every principal component. The highlighted bar is the number of principal components used in our algorithm.

used in many signal processing and machine learning tasks [6, 9–11]. The goal of PCA is to reduce the dimensionality of data to a determined value while retaining as much of the variance as possible. For machine learning tasks, variance can represent information that is important for classification. Thus performing PCA removes any unnecessary information and allows us to keep the more essential parts.

A visual example of PCA in two dimensions is given in Fig. 1. Each circle indicates a data point represented by the two numbers of an (x,y) coordinate. The two solid lines indicate the principal components; the length of the line indicates the amount of variance explained by each component. The way to represent each point as a single number (reducing the dimensionality by a factor of two) while maintaining optimal variance would be to represent it as a scalar multiple of the largest principal component. This idea correlates to higher dimensions as well by using linear combinations of the largest principal components.

In this work, we used PCA to project each data point from its original 40-dimensional space down to a 10-dimensional space. In doing so, we retained 80% of the explained variance. A plot of the total explained variance for each number of principal components is shown in Fig. 2. It should be noted that the first two principal components contain the majority of the variance with a value of 56%. We decided to use 10 principal components because they contained 80% of the explained variance, and every subsequent principal component added 2% or less to that value.

It was interesting to note which of the original 40 features contributed the most to the principal components.

For the majority of the principal components, a mixture of the first seven features given from the dataset contributed the most variance. These features correspond to heart rate, pulse oximetry, temperature, systolic blood pressure, mean arterial pressure, diastolic blood pressure, and respiration rate. There were also notable anomalies from principal components three and six. Principal component three was heavily influenced by the fraction of inspired oxygen, pH levels, partial pressure of carbon dioxide from arterial blood, and oxygen saturation from arterial blood. Principal component six was heavily influenced by hematocrit, hemoglobin, leukocyte count, and platelet count. Another notable exception was how serum glucose levels had a notable effect on principal components seven, eight, and nine. While the majority of these features may not play a large role in clinical sepsis detection, the unsupervised PCA algorithm indicates they have the potential to contribute to sepsis detection.

2.3. Feature Selection

To complement the data gathered from the PCA process, we implemented a quick Sequential Organ Failure Assessment (qSOFA) [12] that allowed us to quickly determine whether a patient was at a high risk of mortality attributed to the onset of sepsis. The qSOFA score is not a diagnosis of sepsis, but rather a quick test to determine if a patient is at risk of sepsis. We include this indicator as a selected feature to be used as an input towards the classifier because of its causal behavior in attributing patient data and its overall importance to medical personnel.

A qSOFA score is calculated by examining a patient's systolic blood pressure, respiratory rate, and Glasgow coma scale score. A point is added to the patient's score if their respiratory rate is higher than 22 breaths per minute,

if the systolic blood pressure is less than 100 mmHg, or if their score on the Glasgow coma scale is less than 15. A score of 2 or more indicates a higher risk of mortality.

For this particular challenge, the Glasgow coma score is not given. Because of this we ignore this feature, and use a modifed qSOFA score as a binary indicator. In our algorithm, if a patient satisfies both the systolic blood pressure and respiratory rate conditionals at a certain hour, they are assigned a positive indicator (one) at that hour. If the patient does not satisfy both conditions, the indicator is left blank (zero). This value is then appended to the PCA coefficients and used as an input to the classifier.

2.4. Classification

The gathered features from the feature extraction and feature selection steps are then used within a classifier system to make a determination of sepsis. Classification is a guided process of using input features and mapping them to output discrete labels. The classifier we used is a neural network implemented in python using TensorFlow and Keras libraries. The TensorFlow and Keras libraries are used to create and synthesize a model with the selected inputs (features), matching sepsis outputs (labels), and the declaration of different activation layers.

Our model consists of the input layer (eleven nodes: ten PCA coefficients and the modified qSOFA score), two hidden layers (consisting of six nodes and four nodes), and the output layer (two nodes), as shown in Fig. 3. The output layer consists of two nodes: node one indicating a sepsis prediction six hours into the future and node two indicating that sepsis is not predicted. The four layers are all fully connected and each uses a logistic sigmoid function as the nonlinearity. The logistic sigmoid activation function is defined as

$$y = \frac{1}{1 + e^{-\alpha(x-\beta)}}, \tag{1}$$

where x is defined as the input from the preceding node. The parameter α indicates the steepness of the transition between -1 and 1 and the parameter β indicates the location of where the transition takes place. We used the default Tensorflow values for both parameters ($\alpha = \beta = 1$). The number of layers and hidden nodes were deliberately chosen as a primary means to improve computational speed and efficiency. A secondary goal in having a smaller network was to attempt to prevent overfitting of the inital training dataset, which contained only 4,000 patients. The architecture of the model was not changed when the much larger training dataset of 40,000 patients became available.

Initial development of this network used a softmax activation function as the nonlinearity output, but this was later changed to a sigmoid function. It was discovered that taking the softmax resulted in all patients being classified
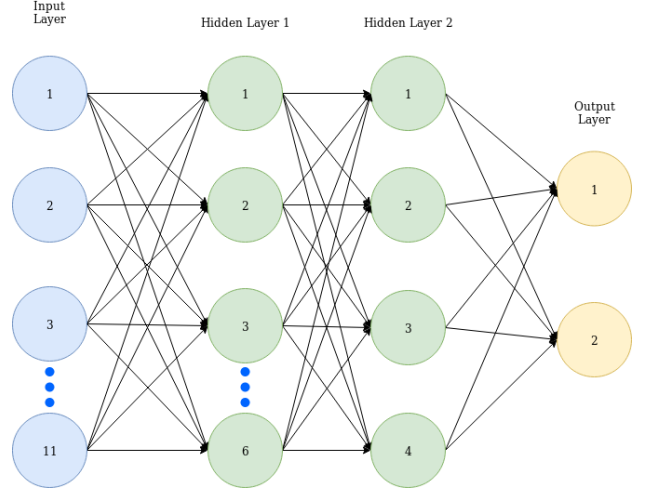

Figure 3. Visual representation of our neural network design. The input layer is made up of the our 11 features. The two hidden layers are shown and lead into the final output layer.

as non-sepsis. In comparison, the sigmoid function offered more variability, and was used for the remainder of development. Two output nodes, one representing a negative indicator of sepsis and another representing a positive indicator of sepsis are created. However, due to the binary nature of the outputs (if one was high, the other was low), only the positive indicator was used. The positive sepsis indicator would constantly remain low but would jump up several degrees of magnitude during what was believed to be potential indicators of sepsis. A user defined threshold of 0.1 was created and if the positive indicator for sepsis crossed that threshold, the classifier would mark that as a confirmation of sepsis.

In the training procedure, the 40,000 training patient data was split into a 90-10 test cross-validation set. We used Tensorflow's RMSprop optimizer to learn the weights and categorical cross entropy as the loss function. The RMSprop optimization algorithm is similar to gradient descent with momentum and worked best for our inputs [13]. We trained the network for a duration of 5 training sessions. During each session, the order of patients is changed as well as which patients are used in the 90-10 test/cross validation split.

3. Results

By testing our algorithm on an unknown subset of the given data set we were able to calculate a score based on the guidelines defined by the challenge organizers [1]. The score encourages the prediction of sepsis six hours prior to the clinical diagnosis of sepsis and provides positive points for predictions in the 0-12 hours before the onset of sepsis.

A negative score is achieved for predicting sepsis too early or too late, as well as for predicting sepsis for patients who do not develop the condition. Our developed algorithm achieved a validation score of 0.038. The official score given to our team, Whitaker's Warriors, was 0.022.

4. Conclusion

The methods described in this paper for using feature extraction and feature learning as preprocessors for a neural network classifier show promise in detecting sepsis from clinical data. By first preprocessing, then using PCA and a quick SOFA score to enhance the original data set we found we could extract enough information to detect sepsis in patients.

By attending the conference we were able to discuss and generate ideas for additional changes for our algorithm. Moving forward, our team will look into utilizing two neural networks: one to preform an hour-by-hour analysis, and another that uses a sliding window. By implementing a window, our algorithm would be able to utilize past information which should greatly increase the accuracy of our algorithm. Using a sliding window would also open up the opportunity to use the true SOFA score, the test that determines if a patient has sepsis or not by using past information. More work could be done to expand upon our feature extraction method by finding other useful features to add. We will also explore different neural network functions and architectures, as well as other classifiers that may be suitable for this situation. We used a small network architecture for efficiency in training and saving the model; however, a larger network may be more successful in accurately diagnosing sepsis.

Acknowledgments

References

[1] Reyna MA, Josef C, Jeter R, Shashikumar SP, Westover MB, Nemati S, Clifford GD, Sharma A. Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019 ;In Press.

[2] Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche JD, Coopersmith CM, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). Jama 2016;315(8):801–810.

[3] Paoli CJ, Reynolds MA, Sinha M, Gitlin M, Crouser E. Epidemiology and costs of sepsis in the united statesan analysis based on timing of diagnosis and severity level. Critical care medicine 2018;46(12):1889.

[4] Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, Suppes R, Feinstein D, Zanotti S, Taiberg L, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Critical care medicine 2006; 34(6):1589–1596.

[5] Seymour CW, Gesten F, Prescott HC, Friedrich ME, Iwashyna TJ, Phillips GS, Lemeshow S, Osborn T, Terry KM, Levy MM. Time to treatment and mortality during mandated emergency care for sepsis. New England Journal of Medicine 2017;376(23):2235–2244.

[6] Hotelling H. Analysis of a complex of statistical variables into principal components. Journal of educational psychology 1933;24(6):417.

[7] Hinton GE. Learning multiple layers of representation. Trends in cognitive sciences 2007;11(10):428–434. ISSN 1364-6613.

[8] Friedman JH. On bias, variance, 0/1—loss, and the curse-of-dimensionality. Data Mining and Knowledge Discovery Mar 1997;1(1):55–77. ISSN 1573-756X. URL https://doi.org/10.1023/A:1009778005914.

[9] Brunton BW, Brunton SL, Proctor JL, Kutz JN. Optimal sensor placement and enhanced sparsity for classification. arXiv preprint arXiv13104217 2013;.

[10] Nie F, Yuan J, Huang H. Optimal mean robust principal component analysis. In Xing EP, Jebara T (eds.), Proceedings of the 31st International Conference on Machine Learning, volume 32 of Proceedings of Machine Learning Research. Bejing, China: PMLR, 22–24 Jun 2014; 1062–1070. URL http://proceedings.mlr.press/v32/nieb14.html.

[11] Yi S, Lai Z, He Z, ming Cheung Y, Liu Y. Joint sparse principal component analysis. Pattern Recognition 2017;61:524 – 536. ISSN 0031-3203. URL http://www.sciencedirect.com/science/article/pii/S0031320316302370.

[12] Barbara P, Graziano C, Caputo W, Litvak I, Battinelli D, Hahn B. The quick sequential organ failure assessment (qSOFA) identifies septic patients in the out-of-hospital setting. The American Journal of Emergency Medicine 2018;36(6):1022 – 1026. ISSN 0735-6757. URL http://www.sciencedirect.com/science/article/pii/S073567571830086X.

[13] Ruder S. An overview of gradient descent optimization algorithms. arXiv preprint arXiv160904747v2 2016;.

Address for correspondence:
Bradley Whitaker (bradley.whitaker1@montana.edu)