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 preliminary test score received from the Challenge was 0.040.

Conclusions: Achieving a positive utility score of 0.040 shows our method of combining PCA and 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 diagnose and detect sepsis in clinical patients using 40 biological features that were 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 components 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 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, the data was normalized to have a standard deviation of one and a mean of zero. This allowed us to have consistent data to input into our program, so each hour and each feature could be treated equally in our 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 used in many signal processing and machine learning tasks [6, 9–11]. The goal of PCA is to reduce the dimensionality

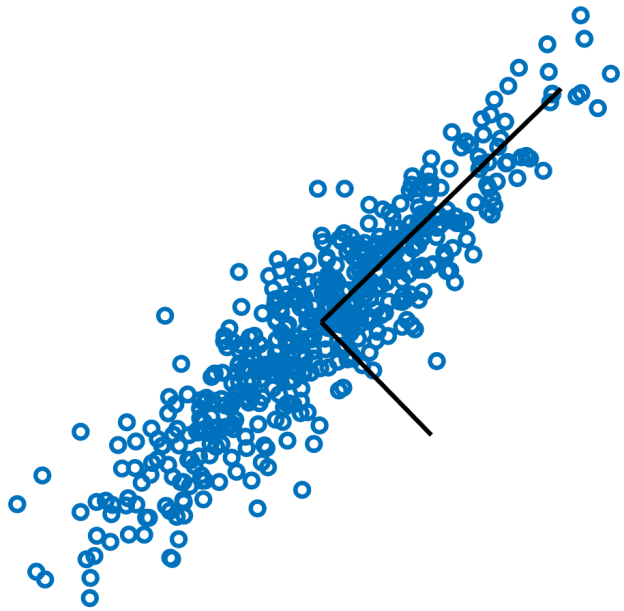

Figure 1. Visual representation of PCA in two dimensions. The overall dimension can be reduced by projecting data points on to the subspace spanned by a subset of the principal components. While some information is lost, much of the explained variance is preserved using just a few principal components.

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 76.7% of the explained variance.

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

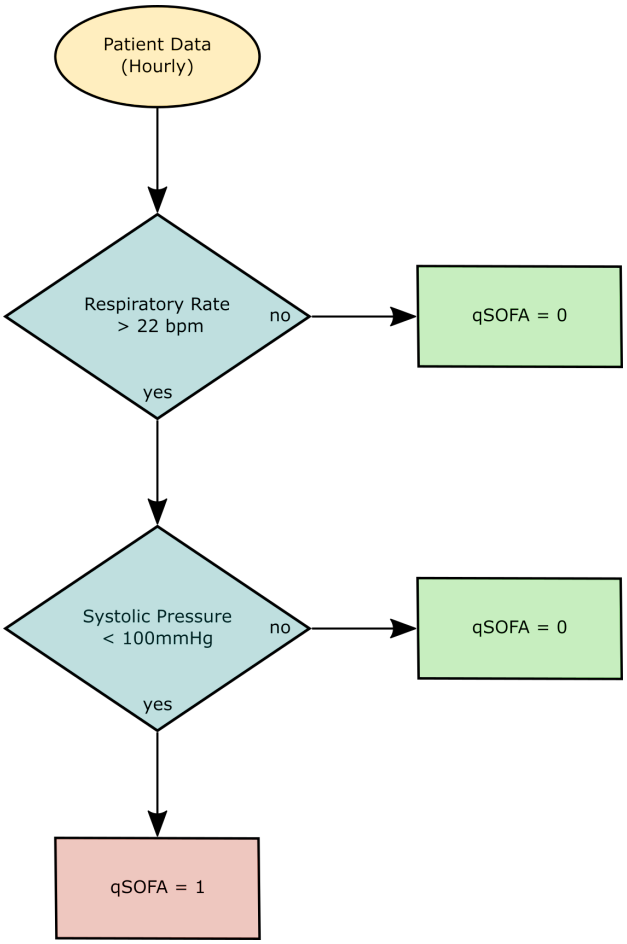

Figure 2. Block diagram of the simple algorithm that determines qSOFA score.

whether a patient was at high risk of mortality failure attributed to onset of sepsis. The qSOFA score is not a diagnosis of sepsis, but rather a binary value that predicts mortality. 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.

The qSOFA score is calculated by examining a patient's systolic blood pressure and respiratory rate at every hour. If a patients respiratory rate is higher than 22 breaths per minute and the systolic blood pressure is less than 100 mmHg, the patient has a higher risk of mortality. Therefore, if a patient satisfies both conditionals at a certain hour, they are assigned a positive indicator at that hour. Fig. 2 presents this algorithm as a block diagram.

2.4. Classification

The gathered features from the feature extraction and feature selection steps are then used within a classifier sys-

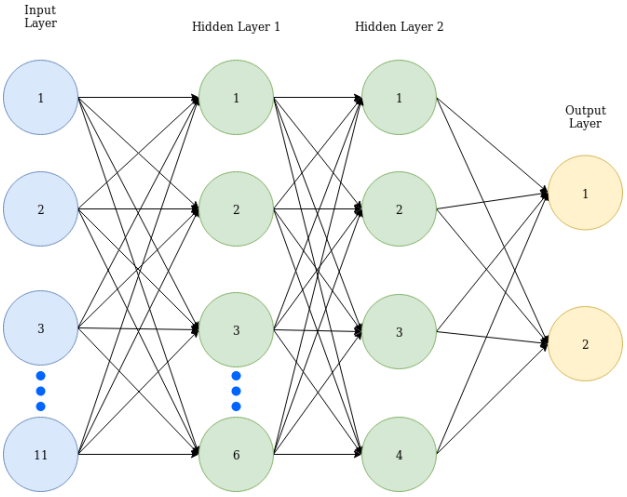

Figure 3. Visual representation of our neural network design. The input layer is made up of the our 11 features. The two hidden layers are shown and lead into the final output layer which is either a 0 or 1 for sepsis or no sepsis.

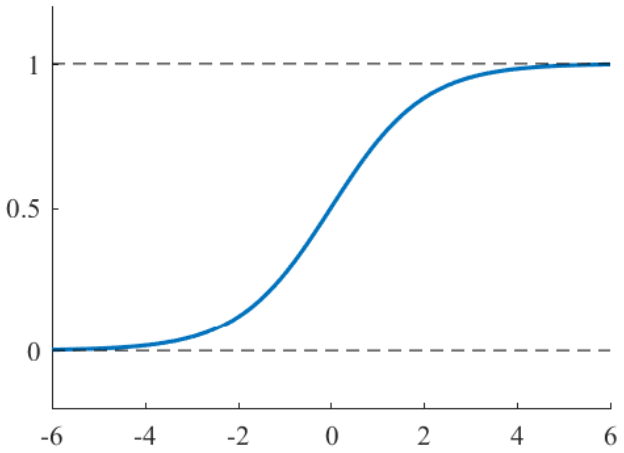

Figure 4. Example of a sigmoid function. The signal moves smoothly from 0 to 1, allowing it to act as a binary neuron indicator function. The smooth transition is helpful for training neural networks. The slope and location of the transition are parameters that are tuned in the training process.

tem 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), two hidden layers (six nodes and four nodes), and the output layer. The output layer consists of two nodes: one indicating a sepsis prediction six hours into the future and the other indicating that sepsis is not predicted. The four layers are all fully connected and each uses a sigmoid function as the nonlinearity. The architecture of the neural network is illustrated in Fig. 3.

A sigmoid function is shown in the plot it Fig. 4. A sigmoid activation function is useful in neural networks because the function transitions smoothly and quickly from 0 to 1 [13]. This is an optimal for our case because it allows us to determine the likelihood of either option (septic and non septic) for patients. The option with the higher likelihood was selected as the algorithm's output for that particular hour.

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 [14]. 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 achieving 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 achieves a validation score of 0.038. The preliminary test score calculated by applying our algorithm to a subset of unseen challenge data was 0.040.

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.

Moving forward, 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 clas-

sifiers 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. In addition, our algorithm only classifies patient data one hour at a time. Future work will focus on how to consider the history of a patient's biological markers in sepsis prediction.

Acknowledgments

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number P20GM103474. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The authors also acknowledge the Undergraduate Scholars Program at Montana State University for their support.

Finally, the authors thank the Computing in Cardiology Conference and PhysioNET Challenge organizers for the work they performed to host this competition.

References

[1] Reyna M, Josef C, Jeter R, Shashikumar S, Moody B, Sharma A, Nemati S, Clifford G. Early prediction of sepsis from clinical data – the Physionet Computing in Cardiology Challenge 2019. In 2019 Computing in Cardiology Conference (CinC). 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] Sutskever I, Hinton GE. Deep, narrow sigmoid belief networks are universal approximators. Neural Computation 2008;20(11):2629–2636. URL https://doi.org/10.1162/neco.2008.12-07-661. PMID: 18533819.

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

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