Prediction of Sepsis Using LSTM Neural Network With Hyperparameter Optimization With a Genetic Algorithm

Petr Nejedly, Filip Plesinger, Ivo Viscor, Josef Halamek, Pavel Jurak

Institute of Scientific Instruments of the Czech Academy of Sciences, Brno, Czech Republic

Abstract

Here in this study, we propose an algorithm for the prediction of sepsis using Long Short-Term Memory (LSTM) neural networks. The algorithm was trained and validated on public dataset proposed by PhysioNet Challenge 2019. We have used a differential evolution genetic algorithm to find optimal training hyperparameters and probability threshold for the inference phase. The deployed version of the algorithm obtained normalized utility score 0.278 based on full challenge test set (team name: ISIBrno; rank 27 of 78). Presented method required 50 mins of single-core CPU processing time.

1. Introduction

In recent years, machine learning and deep learning techniques applied to signal processing have emerged with outstanding results. For example, application of convolutional neural networks classification of ECG was widely used in previous CINC challenges [1], [2]. In general, methods based on neural networks might be used not only for ECG but also for any other time-series. However, methods based on CNNs are not suitable for processing of signals with different signal lengths or non-uniform sampling rate. In this challenge [3], each data example might have a different number of observations with approximate sampling rate of one measurement per hour. For this reason, we choose to use neural networks with Long-Short-Term-Memory layers, which exhibit more suitable features for classification of time-series with nonuniform sampling rates. The LSTM neural networks have been proven to work for sequence classification among scientific fields like speech recognition, machine translation, sentiment analysis, etc. In the last few years, LSTMs has been widely studied to improve classification and analysis of electrophysiological signals like ECG and EEG and achieved impressive results. The architecture of LSTM allows for efficient propagation of gradients during the training phase, which helps to solve the vanishing gradients problem that is common in recurrent neural networks [4].

Here in this paper, we explore the power of LSTM neural network for early prediction of sepsis at the patients hospitalized in the Intensive Care Unit (ICU). The ability to correctly predict sepsis is assessed by challenge utility score that rewards true positive predictions and penalizes false positives and false negatives predictions, respectively.

2. Methods

The public training datasets consist of 40,336 measurements from two hospitals. Provided dataset consist of 40 features that might be used for model training: 8 vital signs (heart rate, oxygen saturation, systolic/diastolic/mean blood pressure, temperature, respiration rate, end-tidal carbon dioxide), biochemical laboratory measurements such as (HCO3, FiO2, pH, Calcium, Chloride, Creatinine, Bilirubin, Glucose, Lactate, Hematocrit, Platelets, etc…), and demographics (Age, Gender, ICU length of stay, etc…).

2.1 Handling Missing Values

The majority of values from biochemical laboratory measurements contains NaN (not a number) values. For this reason, the data preprocessing using two different methods converting NaN values to numerical types were tested. In this study, we have evaluated two methods: the first method converts Nan to "-1" and the second method use forward fill method that is widely used among sepsis detection studies [5]. Forward fill method converts NaN values to the last available meaningful value. In theory, this technique should improve the performance of methods that do not contain memory such as a basic feed-forward neural network. Since we have used LSTMs, we have shown that NaN conversion method does not have a significant effect on the model performance.

2.2 The LSTM Model

Preprocessed data are forwarded to the LSTM neural network with 100 neurons in the hidden layer and further transformed by a fully connected layer (100,2) and dropout layer (p=0.5). Finally, the softmax activation

Figure 1. Training and validation process pipeline
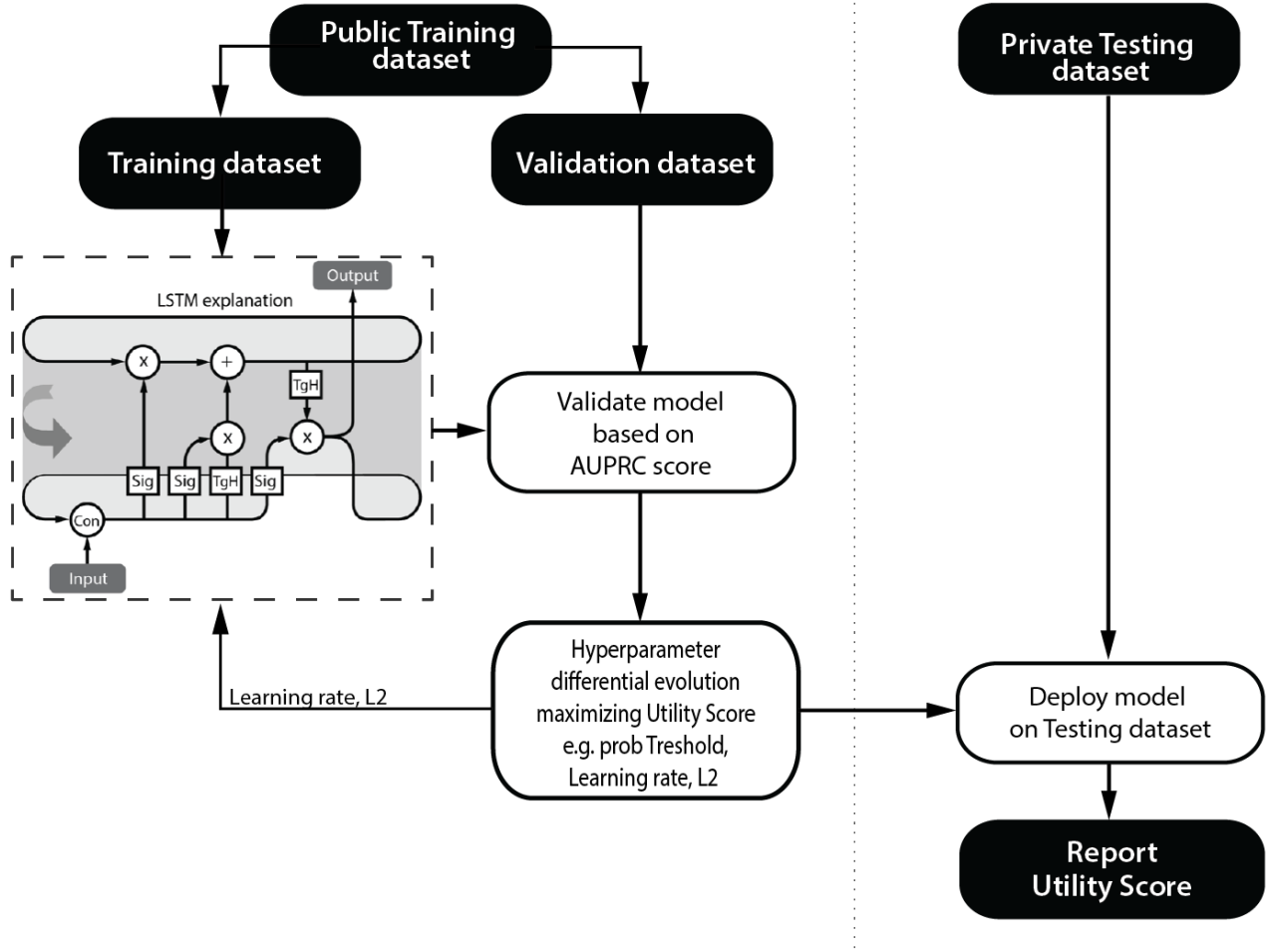
function transforms the output into probabilities for each classification class, i.e., sepsis and non-sepsis. For the training of the model, we have used Adam optimization technique [6] with learning rate 1e-4 and minibatch size of 1 example. In order to train the model, we have randomly split the dataset into training part (80%) and validation part (20%) for out-of-patient validation. The model was trained until the area under the precision-recall curve (AUPRC) from the validation part reached maximum value. The number of epochs varied (5-50) based on applied learning rate, dropout, and L2 regularization (Figure 1).

To evaluate the model performance during the validation, we have chosen to evaluate the AUPRC score. The advantage of AUPRC, in comparison with the challenge utility function, is invariance to the probability threshold. Furthermore, the ability of AUPRC to penalize False positive detections is critical during training on non-balanced data, while commonly used AUROC is not suitable for such a task. The final model with the highest AUPRC score was further evaluated with the challenge scoring function in order to evaluate the utility function.

The optimal probability threshold for the model output was evaluated by differential evolution [7] genetic algorithm maximizing utility score for the validation dataset.

Moreover, in the initial phases of the model development, we have also used the differential evolution optimization technique in order to set the proper learning rate for the LSTM network. By this method, we have found that the learning rate should be set in range 1e-4 to 1e-5 using Adam optimization technique.

The machine learning pipeline was programmed in Python 3.6 programming language while using standard data scientific libraries: Numpy, Pandas, Scipy, and Pytorch.

3.   Results

During the official phase of the challenge the algorithm was tested using subset of the full test set. Therefore, all decisions where based on these results. After the official phase, the algorithm performance was recomputed using the full test set (Tab.1). Results can be located under the team name "ISIBrno" in the official challenge leaderboard. The proposed method achieved rank 27 in the official phase of the challenge.

Table 1. Detailed scores for our best performing model using parts of hidden test set as well as the whole test set.

Model	Utility Score
Test set - hospital A	0.361
Test set - hospital B	0.327
Test set - hospital C	-0.182
Full test set	0.278

4.   Discussion

During our research tests, while using public dataset, we have found that model processing all provided features with forward fill nan handling achieved the best utility score. For this reason, we have choose to submit this model as our final entry. Final results evaluated on test set from hospital A and B shows that model achieved significantly higher utility score while comparing to the random predictor (utility score of random predictor is 0). On the other hand, results evaluated on test set from hospital C suggest that our model is not working properly while evaluated on this dataset. It is not possible to show, why model is not working on the test C, because this part of the dataset is not available for public usage. However, we can speculate that this drop in the performance might be connected to ICULOS feature, which is discussed below.

During our tests, we have found that feature ICULOS (ICU length of stay), have a significant effect on the model performance. While this feature is included in the training phase, the model exhibit a tendency to overtrain. Which means that model weights information from this feature more than other features. Practically this means that model in most cases switch to the sepsis state while ICULOC reaches the range of 40-50 hours, and keep high probability for subsequent input values (Figure 2).
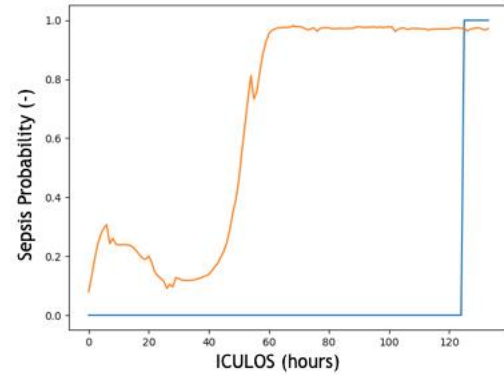

Figure 2. Example of LSTM probability output. The Graph shows a significant increase in predicted probability of sepsis between 40 – 60 hours of stay at ICU.

This has significant consequences for the model usability in real use-case. In general, this would mean, that if the patient stays in the ICU longer than a certain time, there is a significantly increased chance to have sepsis. This feature might be directly observed from the training set (Figure 3). Based on our findings from model training, excluding the ICULOS feature led to worse model performance. It's worth to mention that, all features are indirectly related to the ICULOS. This means, that if we provide measurements to the model approximately every hour, model might make decision based only on the number of iterations of incoming measurements.
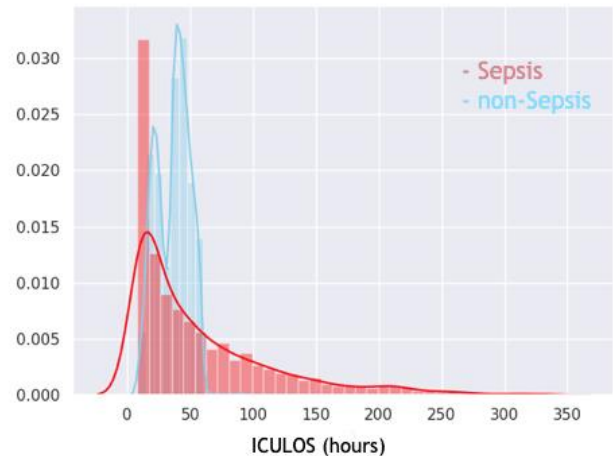

Figure 3. Probability distribution function showing the length of stay in ICU for public dataset. The picture depicts histogram of max(ICULOS) extracted from each training file.

In general, training of LSTM models requires high computational power and utilization of GPUs, which results from backpropagation-through time training algorithm. Furthermore, vanishing and exploding gradient problem should be addressed in order to achieve satisfactory results. Such problems might be solved by setting a proper learning rate and gradient clipping. The computational complexity of the proposed LSTM model for the inference requires significantly fewer computing operations in comparison with the training phase. Our model processed the challenge hidden testing dataset A in 50 mins, utilizing only CPU processing in the challenge docker environment in Google cloud submission system.

5.   Conclusion

This paper proposed the solution for the Computing in cardiology 2019 challenge utilizing the LSTM neural networks and hyperparameter optimization with a differential evolution genetic algorithm to find the hyperparameters for the inference. The method proposed here is theoretically able to optimize multiple hyperparameters of neural network, however in this paper we show only optimization of probability threshold output and learning rate. Our solution achieved 0.278 challenge utility score based on full test set, leading to rank 27 from the total of 78 teams. The proposed method achieved better results than random chance predictor (utility score for random change predictor is equal to 0).

Even though the utility score obtained by the proposed algorithm achieved a significantly better score than a random chance predictor. During the tests proposed in this study, we have observed that results are highly dependent on the feature that is describing the length of stay in the ICU. For this reason, we would like to stress out the need to further investigate effects of ICULOS, and properly conduct statistical testing. Given our method used for this dataset, is not clear whether and to which extent are other features relevant for the sepsis prediction. At the same time, it would be interesting to observe effects of hyperparameters that are used in the challenge utility score: i.e. weights for false positive, false negative and true positive detections, which might yield to completely different results.

References

[1]   F. Plesinger, P. Nejedly, I. Viscor, J. Halamek, and P. Jurak, "Parallel use of a convolutional neural network and bagged tree ensemble for the classification of Holter ECG," *Physiol. Meas.*, vol. 39, no. 9, 2018.

[2]   F. Plesinger, P. Nejedly, I. Viscor, J. Halamek, and P. Jurak, "Automatic detection of atrial fibrillation and other arrhythmias in Holter ECG recordings using rhythm features and neural networks," in *Comput Cardiol (Rennes IEEE)*, 2017, vol. 44

[3]   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. CriticalCare Medicine 2019;In press

[4]   Manaswi, Navin Kumar. "Basics of TensorFlow BT - deep learning with applications using Python: chatbots and face, object, and speech recognition with TensorFlow and Keras." edited by Navin Kumar Manaswi, 1–30. Berkeley, CA: Apress, 2018. https://doi.org/10.1007/978-1-4842-3516-4_1.

[5]   Q. Mao *et al.*, "Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU," *BMJ Open*, 2018.

[6]   Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization," 2014, 1–15. http://arxiv.org/abs/1412.6980.

[7]   J. A. Price, Kenneth, Storn, Rainer M., Lampinen, *Differential Evolution A Practical Approach to Global Optimization*. Berlin/Heidelberg: Springer-Verlag, 2005.

Address for correspondence:
Petr Nejedly & Filip Plesinger
Institute of Scientific Instruments of the CAS, v. v. i.
Czech Academy of Sciences
Královopolská 147, 612 64 Brno, Czech Republic

e-mail:
nejedly@isibrno.cz
fplesinger@isibrno.cz