Sepsis Detection in Sparse Clinical Data Using Long Short-Term Memory Network with Dice Loss

Tomas Vicar[1], Petra Novotna[1], Jakub Hejc[1], Marina Ronzhina[1], Radovan Smisek [1]

[1] Department of Biomedical Engineering, Faculty of Electrical Engineering and Communications, Brno University of Technology, Brno, Czech Republic

Abstract

This paper aims to present a methodology for sepsis prediction from clinical time-series data. Sepsis is one of the most threatening states which could occur while treating a patient at the intensive care unit. Therefore its prediction could significantly improve the quality of the patient treatment.

In this work, we address the problem of sepsis prediction with Long Short-Term Memory (LSTM) network with specialized deep architecture with residual connections. The output of the network is sepsis prediction score at each point in time.

Feature normalization into the fixed range of values is applied including replacing missing values with numerical representation from outside the normalized range. Therefore, the LSTM network is able to include missing values in the learning process. Also, the rarity of sepsis occurrence in the provided dataset is a challenging problem. This problem is addressed by the application of dice loss providing automatically weighted classes by the occurrence of the feature.

The proposed method leads to 0.281 normalized utility score on the full test set as the best official PhysioNet/Computing in Cardiology (CinC) Challenge 2019 entry of ECGuru10 team.

1. Introduction

Sepsis is a longterm chronic critical illness often associated with prolonged inflammation, immune suppression, organ injury and lean tissue wasting [1]. Organ dysfunction can be identified as an acute change in total SOFA score ≥ 2 points consequent to the infection [2]. It can lead to a life-threatening organ dysfunction caused by a dysregulated host response to infection. Definitions of sepsis and septic shock [2] differentiated sepsis from uncomplicated infection. Sepsis is the primary cause of death from infection, especially if not recognized and treated promptly. Its recognition mandates urgent attention

Early prediction of such a state at intensive care units (ICUs) could be crucial for patient treatment.

Machine learning methods are a field of interest in critical care and sepsis detection [3]. Variety of machine learning methods have been used including support vector machines [3], hidden Markov models [4], XGBoost [5], random forest [6], etc.

Currently, recurrent reural networks (RNN) and especially LSTM networks [7] are a very powerful tool, which is wildly used in many tasks, where input is a sequence with variable length, including signal classification [8], natural language processing [9], etc. In [10] authors modeled clinical time-series with multi-output Gaussian processes and fed latent variables into a RNN to classify the patient as septic or non-septic. In [11] authors combined static information extracted from information about patient with fully connected (FC) network and dynamic information extracted from time-series with a combination of convolutional neural network and LSTM. They used this for sepsis prediction, where the whole network is end-to-end differentiable.

LSTM network deals promptly with several problems that occurred while performing the analysis. Sepsis prediction task is a prediction from the signal of variable length with the use of only previous time points. RNN are perfectly suitable for this type of task. It also effectively captures the underlying temporal structure of time-series.

In this work, we address the problem of sepsis prediction with LSTM network with specialized deep architecture with residual connections. Proposed solution deals with two main obstacles: a high portion of missing values in the provided dataset and highly imbalanced dataset.

2. Methods

2.1. Missing values problem

Missing values are one of the most challenging problems considering clinical time-series data. There are several standard approaches of how to handle missing values. The most simple one is replacement with constant value

e.g. zero or average. Another possibility is to use the latest known value. However, standard interpolation methods cannot be used, due to their non-causality (future value cannot be used for calculation of current value). More advanced possibilities for prediction of missing values are regression techniques e.g. usage of LSTM network [12]. A different approach is the use of the prediction technique, which can deal with missing values. The example of such technique is naive Bayes classifier [13], which can take into account known values only; however, this is a simple technique, which did not achieve satisfying results. Since LSTM networks are the main topic of this paper, mentioned LSTM network method was applied, but did not reach any reasonable results. The best result, which was achieved with this method, was the replacement of missing values with nothing more accurate than the average value in most of the cases.

In this paper, feature normalization into the fixed range of values is applied including the replacement of missing values with numerical representation from outside the normalized range. Specifically, values of every feature were fit into the range $1, 5$ and missing values were replaced by 0. Resulting values y are then

$$y = \begin{cases} 0 & \text{if } x \text{ is missing} \\ \frac{4(x - x_{min})}{x_{max} - x_{min}} + 1 & \text{otherwise} \end{cases} \quad (1)$$

where x is the original value, x_{min} and x_{max} are minimum and maximum of the specific feature over all the patients and all the time points, respectively. This approach is motivated by the expectation that the network can recognize zeros as non-informative and learn to ignore them.

When preprocessing data, it was noticed that for some features there are values not corresponding to physiological possibilities of the human body. In some cases, the recorded values were significantly (multiple-times) lower (or higher) than the lowest (or the highest) known value from the physiological point of view. According to the consultation with medical expert, all values outside the specified range were replaced by the range limit values. The normalization was performed in the same way as in the previous case (eq. 1). Clinical features, which were mostly represented by missing values or according to the medical expert did not provide any important information, were removed. However, this data preprocessing did not provide any improvement. Even though it is believed that there should not be present data which do not make sense from the physiological view.

2.2. Class imbalance problem

The second important problem toward clinical time-series data is the class imbalance. The sepsis occurs at 1.80 % of all time points across all the patients. The percentage of patients diagnosed as septic is 7.27 %. A low number of representatives of sepsis class causes trouble in most of the standard prediction techniques, leading to the prediction of non-sepsis class in all cases. This obstacle can be overcome by oversampling of the rarer (septic) class or the introduction of higher weights for samples from rarer class. In the case of neural networks, the standard approach is the introduction of weighted loss function such as weighed cross-entropy or Generalized Dice Loss (GDL)[14].

In this paper, the class imbalance problem was dealt with by specialized lost function, which can automatically set the weights appropriately for each of the classes [15], which is for two classes defined as

$$GDL = 1 - 2 \frac{\sum_{i=1}^{2} w_l \sum_n r_{ln} p_{ln}}{\sum_{i=1}^{2} w_l \sum_n r_{ln} + p_{ln}} \quad (2)$$

where $w_l = 1/(\sum_{n=1}^{N} r_{ln})^2$, l is an index of the class, n is an index over time and batch. r_{ln} is value of one-hot encoded ground truth and p_{ln} is score output of network (output of softmax layer).

2.3. Network architecture and implementation

The proposed network consists of 7 blocks, where each consist of LSTM layer followed by 3 fully connected layers. Inspired by ResNet [16] and DenseNet [17], we add also residual skip connections, where the input of every block is a concatenation of output of the previous block, skip over the previous block and network input, as is shown on Fig. 1. Every fully connected block is followed by ReLU and dropout layer (with 0.5 drop probability). Inputs to our network are all previous time points before the evaluated time point. Output is a prediction score of sepsis/non-sepsis at evaluated time point. Output softmax layer ensures the mapping of the output values into the range 0-1.

2.4. Threshold adjustment

Proposed loss function does not guarantee that best threshold of output score (with respect to the utility score) will be 0.5. The best threshold was selected with a grid search, where utility measure was maximized for the validation set.

2.5. Training details and implementation

Network was trained with Adam optimizer [18] with $\beta_1 = 0.9$, $\beta_2 = 0.999$) and learning rate 0.0001 decreasing to $1/10$ every 15 epochs. Additionally L2 weight regularization with factor 10^{-8} and gradient clipping [19]
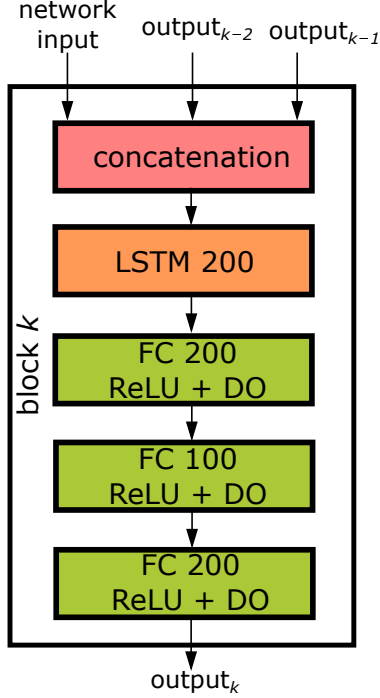
Figure 1. One block of proposed network. Numbers refer to number of neurons. Whole network consist of 7 blocks.

were used. During training session all time points were used. Batch 64 was used, in case of different signal length, signal was padded with zeros. Matlab 2019a with Deep Learning Toolbox was used for network implementation and training of proposed network. Code is available at https://github.com/ECGuru10/PhysioNetCHallenge2019.

3. Results and Discussion

Clinical time-series data for sepsis prediction was provided by the PhysioNet/Computing in Cardiology (CinC) Challenge 2019 [20]. The dataset contains data from 40336 patients, where 40 clinical features were recorded for every patient by 1-hour interval. The training dataset was randomly divided into internal training (90 %) and validation (10 %) sets. The training set was used for network training. The validation set was used for threshold adjustment and evaluation. Sepsis labels are shifted by 6 hours, thus algorithm should predict sepsis 6 hours before it starts. Results were evaluated in terms of dice coefficient and normalized utility (official challenge metric).

Final best model uses 7 blocks, with normalization/missing value replacement from Eq. 1. This leads to 0.350 utility and 0.131 dice on our validation set, and 0.372 utility on challenge official partial test set A. Additionally, our model was evaluated for different preprocessing methods, model settings and model sizes.

Our method for replacement of missing values (Eq. 1),

Table 1. Results for different preprocessing and network setting.

		Normalized Utility	Dice
Preproc.	Last known value/average	0.315	0.119
	Normalization (Eq. 1)	0.350	0.131
	Physiological correction + Normalization (Eq. 1)	0.349	0.120
Network	w/o Dropout w/o skip con.	0.336	0.123
	w/ Dropout w/o skip con.	0.297	0.111
	w/o Dropout w/ skip con.	0.327	0.121
	w/ Dropout w/ skip con.	0.350	0.131

achieved significantly better results than replacement by last known value or by average from the whole training set if there is no previous known value for current signal (see Tab. 1). This comparison was performed for the network with dropout and skip connections and with 7 blocks. Prediction process was applied on both sets of data - with and without physiological correction (see section 2.1). Since the correction did not provide improvement in the outcome, the original data set was used to gain the final result in the challenge.

The improvement achieved by residual connections and dropout was evaluated. The very similar utility was achieved by network with and without skip connections, when dropout was not used. However, use of the network with both dropout and skip connections leads to significant improvement. This comparison was performed for normalization by Eg. 1 and for the network with 7 blocks.

Additionally, the performance was tested for a different number of blocks (see Tab. 2). The utility is increasing with the number of blocks until the number of 5, where it saturates. Using more blocks did not provide any additional improvement. This comparison was performed for normalization by Eq. 1 and for the network with both dropout and skip connections.

Table 2. Effect of blocks number on the prediction performance of the network.

		Normalized Utility	Dice
Num. of blocks	1 block	0.328	0.112
	2 blocks	0.335	0.121
	3 blocks	0.342	0.117
	4 blocks	0.341	0.130
	5 blocks	0.350	0.133
	6 blocks	0.349	0.129
	7 blocks	0.350	0.131

4.   Conclusions

The use of the RNN (LSTM network) is perfectly suitable for this type of task. Proposed method addressed the missing values problem effectively by simple normalization and replacing missing values with numerical representation from outside the normalized range. The network can recognize replaced values as non-informative and therefore ignore them. The rarity of sepsis occurrence in the provided dataset is addressed by the application of dice loss, which provides automatically weighted classes by the occurrence of the feature. The final model consists of a newly designed block, which is repeated multiple-times. There are three inputs to each block - output of the previous block, skip connection over the previous block and the initial network input.

The proposed method leads to 0.281 normalized utility score on full test set (0.372 on test set A) as the best official PhysioNet/Computing in Cardiology(CinC) Challenge 2019 entry of the ECGuru10 team.

References

[1] Hotchkiss RS, Moldawer LL, Opal SM, Reinhart K, Turnbull IR, Vincent JL. Sepsis and septic shock. Nature reviews Disease primers 2016;2:16045.

[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] Vellido A, Ribas V, Morales C, Sanmartín AR, Rodríguez JCR. Machine learning in critical care: state-of-the-art and a sepsis case study. Biomedical engineering online 2018; 17(1):135.

[4] Ghosh S, Li J, Cao L, Ramamohanarao K. Septic shock prediction for icu patients via coupled hmm walking on sequential contrast patterns. Journal of biomedical informatics 2017;66:19–31.

[5] Barton C, Chettipally U, Zhou Y, Jiang Z, Lynn-Palevsky A, Le S, Calvert J, Das R. Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. Computers in biology and medicine 2019;109:79–84.

[6] Darwiche A, Mukherjee S. Machine learning methods for septic shock prediction. In Proceedings of the 2018 International Conference on Artificial Intelligence and Virtual Reality. ACM, 2018; 104–110.

[7] Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J. Lstm: A search space odyssey. IEEE transactions on neural networks and learning systems 2016; 28(10):2222–2232.

[8] Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. Journal of the American Medical Informatics Association 2016;24(2):361–370.

[9] Liu P, Qiu X, Huang X. Recurrent neural network for text classification with multi-task learning. arXiv preprint arXiv160505101 2016;.

[10] Futoma J, Hariharan S, Sendak M, Brajer N, Clement M, Bedoya A, O'Brien C, Heller K. An improved multi-output gaussian process rnn with real-time validation for early sepsis detection. arXiv preprint arXiv170805894 2017;.

[11] Lin C, Zhangy Y, Ivy J, Capan M, Arnold R, Huddleston JM, Chi M. Early diagnosis and prediction of sepsis shock by combining static and dynamic information using convolutional-lstm. In 2018 IEEE International Conference on Healthcare Informatics (ICHI). IEEE, 2018; 219–228.

[12] Verma H, Kumar S. An accurate missing data prediction method using lstm based deep learning for health care. In Proceedings of the 20th International Conference on Distributed Computing and Networking. ACM, 2019; 371–376.

[13] Rish I, et al. An empirical study of the naive bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence, volume 3. 2001; 41–46.

[14] Sudre CH, Li W, Vercauteren T, Ourselin S, Cardoso MJ. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, 2017; 240–248.

[15] Crum WR, Camara O, Hill DL. Generalized overlap measures for evaluation and validation in medical image analysis. IEEE transactions on medical imaging 2006; 25(11):1451–1461.

[16] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2016; 770–778.

[17] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2017; 4700–4708.

[18] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv14126980 2014;.

[19] Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks. In International conference on machine learning. 2013; 1310–1318.

[20] 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. Critical Care Medicine 2019 2019;.

Address for correspondence:

Petra Novotna
Department of Biomedical Engineering,
Faculty of Electrical Engineering and Communication,
Brno University of Technology,
Technicka 12, 616 00 Brno, Czech Republic.
novotnap@vutbr.cz