Ring-Topology Echo State Networks for ICU Sepsis Classification

Miquel Alfaras[1,2], Rui Varandas[1,3], Hugo Gamboa[3]

[1] PLUX Wireless Biosignals, Lisboa, Portugal
[2] Universitat Jaume I, Castelló de la Plana, Spain
[3] Laboratório de Instrumentação, Engenharia Biomédica e Física da Radiação (LIBPhys-UNL), Departamento de Física, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Caparica, Portugal

Abstract

Sepsis is a life threatening condition that can be treated if detected early. This paper presents a study of the application of a Ring Topology Echo State Network (ESN) algorithm to a sepsis prediction task based on ICU records. The implemented algorithm is compared with commonly used classifiers and a combination of both approaches. Finally, we address how different causal strategies on filling missing record values affected the final classification performances. Having a dataset with a limited number of time entries per patient, the utility score U = 0.188 obtained (team 51: PLUX) suggests that further research is needed in order for the ESN to capture the temporal dynamics of the problem at hand.

1. Introduction

Sepsis is defined as "life-threatening organ dysfunction caused by a dysregulated host response to infection" [1]. This condition can be detected using physiological measures, such as, heart rate, temperature and laboratory measures of body fluids analysis. Furthermore, it was identified that any 2 of 3 clinical variables Glasgow Coma Scale score of 13 or less, systolic blood pressure of 100 mm Hg or less, and respiratory rate 22/min or greater offered predictive validity. For patients in the ICU, sepsis prediction is compromised due to the effects of the prescribed treatment. For instance, drugs that help to attenuate the patient symptoms have an impact on physiological measurements (temperature, heart rate, etc.). In turn, mechanical ventilators do not allow to take the respiration rate into account in order to detect sepsis. In [2], a data-driven probabilistic model was found to be able to improve the prediction of sepsis relative to some of commonly used theoretical thresholds.

On the other hand, Machine Learning approaches have been applied in the medical field in contexts such as arrhythmia detection [3], death prediction [4], and even death prediction in the presence of sepsis [5]. Hence, we decided to study such techniques for sepsis detection in ICU patients.

Our study was aimed at assessing the application of Echo State Network (ESN) algorithms to predict the occurrence of sepsis in ICU data, with an anticipation of 6 hours. For this goal, we explored the influence of the input parameters and optimised them and compared the results obtained using the ESN with standard classifiers. Finally, we implemented a simple combination of both classic classifiers and ESNs. Additionally, our work addressed the effect that different strategies in the replacement of the missing ICU data had on the results obtained using the ESN.

The trained algorithms can be tested with a dataset containing data that has never been shown to the classifiers.

2. Methods

The dataset utilised was provided by the organisation of the *Physionet Challenge 2019* [6] and consists of physiological, laboratory and demographic measurements of 40,336 patients from the ICU of two different hospitals. The readings were registered hourly, with missing values appearing as NaN (not a number). At any point in time, a Sepsis Label indicating the presence of sepsis is provided. Specifically, we count on 40 measurements over 1,552,210 hours, thus making 62,088,400 values, in which 70% are NaN. Furthermore, sepsis samples amount to 1.80% of the total. Algorithm performance is rated according to a utility function $U(s,t)$ that assesses output septic labels (s) overtime (t), penalising excessively early, late, or false detection and favouring timely sepsis detection.

2.1. Preprocessing

The provided data counts on a shift applied to the Sepsis Labels, to train for early (6 hour anticipation) prediction.

One of the first issues addressed, was the replacement

 ISSN: 2325-887X DOI: 10.22489/CinC.2019.327

of the NaN missing ICU record values (see section 3.4) by means of:

1. Substitution by zero (constant value);
2. Substitution by the mean value of the patient records;
3. Substitution by a random value assuming a normal distribution with the mean and standard deviation of the original records;
4. Substitution by the value corresponding to the interpolation of the measurements around the missing entries.

The substitutions were applied taking into account the causality of the problem, in which we could only have access to the past measurements until the current time and ignored the future values to simulate the real-time use case.

2.2. Echo State Network

The main algorithm chosen for the sepsis classification task consisted in a Ring Topology Echo State Network (ESN). ESNs are a particular subset of recurrent neural networks that leverage Reservoir Computing properties of mapping the input data to a high-dimensional space and simplifying the training given the fact that only output layer weights are trained. Random inner connections are set and kept constant. In a successful implementation, 3 conditions are met: a)The recurrent network exhibits different dynamics for inputs that differ significantly, b)The network provides similar outputs for similar inputs, c)The reservoir exhibits a fading memory where temporal dynamics play a relevant role that decays over time.

An ESN consists of 3 layers: Input, Reservoir and Output. First, random weights (i.e., mask) map the input to a high dimensional space, distributing the input along neurons that behave differently so that unique responses are shown for different inputs. Once distributed, inputs need to be converted to a nonlinear space by means of an activation function (such as a sigmoid or hyperbolic tangents $tanh$). The reservoir's internal weights, must recreate a sparsely connected network that preserves the reservoir properties. Finally, the output weights are the only weights subject to training. The training procedure is often guided by the problem of solving the reservoir system presented below.

$$\hat{S} = Esn \times W \tag{1}$$

In eq.1, the classification of samples consists in obtaining the sepsis labels $\hat{S}$, i.e. the product of the the reservoir state holding the sample data **Esn** and the output weights **W** that take into account the contribution of the different neurons.

2.2.1. Topology and ESN formulation

Recent ESN research drawing upon ring topologies has shown how these networks have successfully been applied to cross-database clinical contexts such as the detection of ventricular heartbeats [7]. We chose to work with the same topology and a zero centred sigmoid nonlinear mapping function $f(x) = (1 + e^{-x})^{-1} - 0.5$. Our mask was continuous, randomly generated, uniformly distributed around 0 and counted on an additive offset per neuron.

Since we counted on a full dataset where the number of samples (s) exceeded 1 million, the ESN reservoir matrix that needs to be numerically inverted has dimensions $Esn = Esn_{s,N}$, which turn to be the computationally limiting factor. Instead of trying to numerically solve a system based on computing the pseudo-inverse matrix of the ESN matrix, we adopted the normal equation formulation presented in [8]. Following eq.1, the reservoir equations are:

$$Esn^T \times S = \left[Esn^T Esn\right] W \tag{2}$$

$$\left[Esn^T Esn\right]^* \times \left[Esn^T \times S\right] = W \tag{3}$$

where S represents the known sepsis labels for the training records, Esn is the reservoir matrix holding the information, W are the output weights, $(^T)$ stands for matrix transposition and $(^*)$ corresponds to the numerical pseudo-inverse. The advantage of adopting the linear equation formulation lies in the fact that the matrix that needs to be inverted has a constrained dimension of $R \equiv \left[Esn^T Esn\right] = R_{N,N}$, significantly less computationally demanding using the LU and Moore-Penrose methods.

The ESN states are represented by:

$$Esn_{n,t} = f(\gamma Input_t + \eta(W_{Esn} \times Esn_{n-1,t-1})) \tag{4}$$

where, $Esn_{n,t}$ is the state of neuron n at time t, f is the activation function, γ is the input scaling factor applied to the masked data $Input_t$ and η is the memory or leakage parameter controlling the strength between the fixed internal neuron connections W_{Esn}. The equation captures the temporal and neighbouring neuron dependencies. In the case of the ring topology, we have a W_{Esn} connection matrix that consists of a square matrix with ones in the subdiagonal.

The optimisation approach was based in the exploration of different regimes for the input scaling and memory parameters (γ, η). We computed bidimensional grid searches and looked for the best Area Under the Receiver Operating Characteristic Curve (AUC) score by means of a 10-Fold stratified cross-validation strategy, ensuring that no patient is used simultaneously in the training and test subsets. A threshold is then set to obtain the best F1-score.

3. Results

In this section we present a comparison of sepsis classification performances of the ESN and other classifiers.

Regarding the problem that the absence of values posed, we first adopted the strategy of substituting the missing NaN values by zeros and proceeded with the exploration of other approaches reported below.

We used the AUC to guide our parameter optimisation, and proceeded to apply the optimal threshold and compute the corresponding F1-score, Precision (Pr), Recall (Re) and Accuracy (ACC) for the highest AUC point (η, γ). We checked for robustness of the ESN and input mask dependence by means of ensembles (5-10 networks) that yield 2-3% increases in AUC. When reporting performances we use F1-score, given that it combines Pr and Re, which are the most important metrics in extremely imbalanced data sets like ours.

3.1. Echo State Network Parameters

In the case of the ESN, we started assessing the influence of the amount of neurons employed (N). The reservoir size is a key element for a reservoir computer to be able to either find similarities or tell samples apart. Since its use is motivated by a high dimensional nonlinear mapping, the number of neurons must be greater than the number of features provided at any given point. In order to find out an optimal regime for our network to work, we used the first data subset of 5000 patients and conducted a grid search on several reservoir sizes (of N neurons) scanning over the same ranges of memory and input scaling parameters ($\eta, \gamma \in [0; 10]$).

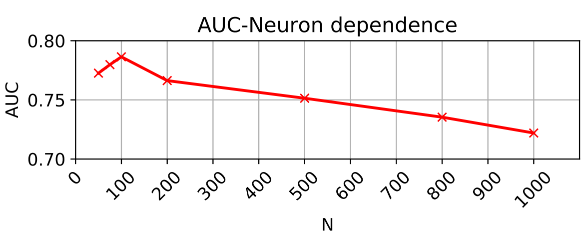

Figure 1. Study of the AUC-Neuron dependence

Fig.1 shows that N=100 provided the highest AUC values. Once the N size was set, we conducted grid searches in the (γ, η) parameter space using the whole dataset. Fig.2 shows a representation of the resulting parameter map.

Table 1 shows the results for the optimal parameters in terms of AUC (N=100, $\gamma \leq 0.001$, $\eta \in [0.1; 2.5]$).

The trained ESN received a $U(s,t) = 0.188$ utility score [6], (team 51: *PLUX*). The results indicate that there was overfitting relative to the training sets A and B, because no relevant classification is achieved in test C (unseen data). However, the overfitting that prevented ESN generalisation is likely to happen in the threshold optimisation, as AUC remained high through test sets.

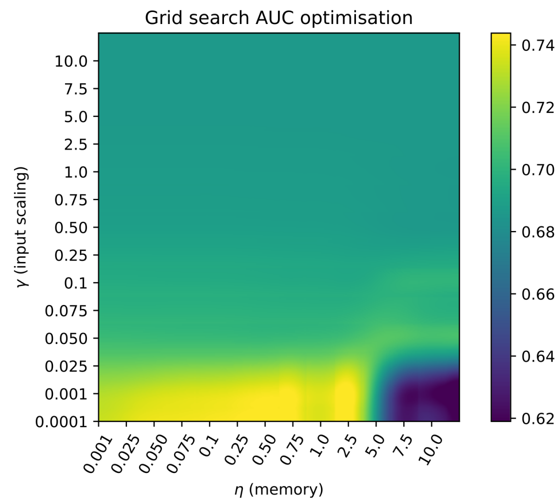

Figure 2. AUC - Grid search ESN parameter optimisation

Table 1. ESN training and official test results.

	Dataset	U	ACC	F1-score	AUC
Train.	10-fold (A+B)	n/a	0.960	0.148	0.744
Test	A	0.206	0.943	0.160	0.667
	B	0.214	0.942	0.127	0.688
	C	0.055	0.877	0.058	0.702
	Official	0.188	n/a	n/a	n/a

3.2. Classifier comparison

In this work we used four different classifiers available in the scikit-learn Python package [9]: Decision Trees (DT), Gaussian Naive-Bayes (GNB), Random Forest (RF) and Gradient Boosting (GB) classifier. Here, we wanted to compare the performance of our ESN in relation to the performance obtained using different classifiers (Table 2).

Table 2. Results of different classifiers.

Classifier	Accuracy	F1-score	Recall	AUC
DT	0.956	0.088	0.113	0.543
GNB	0.810	0.073	0.415	0.665
RF	0.953	0.168	0.265	0.772
GB	0.955	0.183	0.283	0.801

These results indicate that some classifiers may be better suited for this specific problem than the application of ESN, namely, the RF and GB classifiers.

3.3. Combination of ESN and classifiers

The combination of the ESN with other classifiers could possibly increase the performance of both methods. The combination was made by implementing the network described in section 2 and providing the resulting neuron states as the input for the classifiers (see results in Table 3).

Table 3. ESN-classifier combination results.

Combination	Accuracy	F1-score	Recall	AUC
ESN + DT	0.960	0.068	0.083	0.523
ESN + GNB	0.889	0.061	0.199	0.620
ESN + RF	0.950	0.144	0.235	0.749
ESN + GB	0.951	0.153	0.245	0.725

The combined classifications suggest that despite achieving values in some cases comparable to those of the ESN (e.g. ESN+RF, ESN+GB), classifier performances using the network as input are worsened when compared to their raw input counterparts (see Tables 3 and 2).

3.4. Causal NaN Substitution Influence

In an early stage, we addressed noncausal scenarios where a relevant cross-hospital classification (AUC $>$ 0.80) was achieved, enhanced by a minmax scaling and NaN substitution. Given the causal nature of the problem at hand, we investigated further different strategies to substitute missing values, described in section 2.1 and present the corresponding results in Table 4. The low F1-score and AUC influence indicates that no strategy can be considered best.

Table 4. Influence of the NaN substitution on the ESN.

Substitution	Accuracy	F1-score	Recall	AUC
(1) Const.	0.956	0.148	0.194	0.744
(2) Mean	0.962	0.139	0.171	0.759
(3) Norm.	0.923	0.152	0.352	0.756
(4) Interp.	0.953	0.155	0.240	0.759

4. Conclusions

An ESN algorithm was implemented and ranked positively for sepsis prediction. However, the low performance achieved relative to other classifiers suggests that further research is needed in order for the ESN to capture the temporal dynamics of the problem at hand. Having a highly imbalanced dataset with a limited number of time entries per patient, we hypothesise that research into data augmentation approaches maintaining the ESN data resampling in a causal scenario could yield a better sepsis prediction. Although enhanced by the use of other classifiers drawing upon the mapped reservoir states, we can conclude that no implementation based on our ESN outperforms that of commonly used classifiers so far. Finally, while the NaN replacement strategy plays a key role in non-causal use case scenarios, no significant F1-score improvements were observed in the causal classification task.

Acknowledgements

This work was supported by Marie Skłodowska Curie Actions ITN AffecTech (ERC H2020 Project ID: 722022) and Fundação para a Ciência e a Tecnologia (FCT, Portugal), Phd grant PD/BDE/150304/2019.

References

[1] 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.

[2] Tsoukalas A, Albertson T, Tagkopoulos I. From data to optimal decision making: a data-driven, probabilistic machine learning approach to decision support for patients with sepsis. JMIR medical informatics 2015;3(1):e11.

[3] Eerikäinen LM, Vanschoren J, Rooijakkers MJ, Vullings R, Aarts RM. Decreasing the false alarm rate of arrhythmias in intensive care using a machine learning approach. In Computing in Cardiology Conference. IEEE, 2015; 293–296.

[4] Johnson AE, Kramer AA, Clifford GD. Data preprocessing and mortality prediction: The physionet/cinc 2012 challenge revisited. In Computing in Cardiology. IEEE, 2014; 157–160.

[5] Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W, Hall MK. Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data–driven, machine learning approach. Academic emergency medicine 2016;23(3):269–278.

[6] Reyna M, Josef C, Jeter R, Shashikumar S, M. Brandon Westover M, Nemati S, Clifford G, Sharma A. Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019. Critical Care Medicine In Press;.

[7] Alfaras M, Soriano MC, Ortín S. A fast machine learning model for ECG-based heartbeat classification and arrhythmia detection. Frontiers in Physics 2019;7:103.

[8] Lukoševičius M. A practical guide to applying echo state networks. In Neural networks: Tricks of the trade. Springer, 2012; 659–686.

[9] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 2011;12:2825–2830.

Address for correspondence:

Miquel Alfaras - PLUX S.A.
Avenida 5 de Outubro 70, Lisboa (Portugal)
malfaras@plux.info