TASP: A Time-phased Model for Sepsis Prediction

Xiang Li, Yanni Kang, Xiaoyu Jia, Junmei Wang, Guotong Xie

Ping An Technology, Beijing, China

Abstract

***Background**: As a lift-threatening condition, sepsis is one of the major public health issues around the world. Early prediction can improve the sepsis outcomes by prompt treatment.*

***Method**: As part of the Physionet/Computing in Cardiology Challenge 2019, this study proposed a TimephAsed model for Sepsis Prediction (TASP). Realizing the fact that the incidence of sepsis is time-dependent, our model is a fusion of different frameworks along the time dimension. In the beginning stage of ICU stay, a gradient boosting tree model is used to figure out the patients with relatively high risk of sepsis. Following that, another tree model with more features is adopted to identify the risk in middle stage. If a patient stayed in ICU more than 50 hours, a deep learning framework will be used to capture the long-term relations for sepsis risk prediction in late stage. We construct proper features for each sub-models with different missing value imputation strategies.*

***Result**: The proposed model obtains a score of 0.415 by means of 10-fold cross-validation on the training set. Two simplified versions of the model respectively get scores 0.420 and 0.419 on official leaderboard.*

1. Introduction

Sepsis is a serious condition that caused by a life-threatening and overwhelming immune response to infection [1]. It is of great importance to predict sepsis as early as possible. In Physionet/Computing in Cardiology Challenge 2019, a sepsis dataset has been released for modelling and evaluation. It contains 40 clinical variables for each hour of a ICU stay that collected from over 60,000 ICU patients. A novel clinical utility measurement is used to evaluate the predict performance by considering the reward and penalization for different prediction time-stamps.

In this study, we aim to propose a machine learning model to accurately estimate the risk on each hour of a ICU stay. Through a data exploration on the training dataset, we got an interesting finding that the incidence of sepsis is highly related to ICU length-of-stay (LOS). Therefore, we consider taking the advantages of different frameworks to make predictions for different ICU stages. In particular, the proposed time-phased model for sepsis prediction is called **TASP**. It is composed by three parts:

1) **Early stage** (1-9 hours). We use a gradient boosting tree that trained on early hours of all the ICU patients in training set to predict the sepsis risk of each hour in early stage.
2) **Middle stage** (10-49 hours). More time-related features have been constructed as the input for another tree model to make prediction in middle stage.
3) **Late stage** (50+ hours). To better capture the long-term dependencies in data, a recurrent neural network (RNN) [2] is presented as the predictor.

Due to the fact that part of the variables is of high missing rates, we have adopted different missing value estimation strategies for the feature construction of different sub-models. After the model training, given a new ICU patient data hour by hour, we will predict the sepsis risks by the corresponded sub-models according to the time-stamps. We conducted a 10-fold cross-validation on the offline training data, which resulted in score 0.415. We also submitted two simplified versions of TASP (without time-phased) for online test which respectively got scores 0.420 (tree model) and 0.419 (RNN model)[1].

The main contributions of this study are summarized as follows:

- The proposed time-phrase model TASP mimics the incidence regularity of sepsis for ICU patients, so that it can better align with the clinical practice.
- In TASP, we can construct different features and set different cut-offs for different sub-models, which is conducive to more stable prediction.
- Each sub-model in TASP is a single model without ensemble learning techniques, which is of good

---

[1] The complete TASP failed to get a score on online test data with a run time exception.

potential for the model interpretability and extensibility.

2.   Methodology

In this section, we firstly introduce our data exploration. After that, we give the architecture of TASP, followed by the details of each sub-model.

2.1.   Data exploration

Data exploration is an essential procedure for model design. We did a thorough data analysis, including variable distribution and relation, missing pattern, incidence ratio and time, etc. There are three interesting findings which are important for model selection.

1) **Missing pattern**. Some variables, especially laboratory test, are in high missing rate. It is essential to impute the missing parts for various machine learning methods.
2) **Imbalance data**. In the offline training dataset, the number of patients who developed sepsis in ICU stay are 2,932 (7.27%). The ratio of hourly sepsis labels is 98.2 (label=0):1.8 (label=1).
3) **Incidence of sepsis**. We found that the incidence of sepsis is non-linearly related to ICU LOS (as shown in Figure 1). The incidence in the beginning 9 hours (**early stage**) is relative higher than the following 40 hours (10-49, **middle stage**). We infer that it may stem from the condition difference of different patients on the ICU admission time. The patients with more serious condition are easier to get sepsis. The incidence between 10-49 hours is relative stable. While after 50 hours (**late stage**), the incidence raises rapidly. The possible reason is that longer ICU LOS may represent that the patient condition is getting worse and worse.
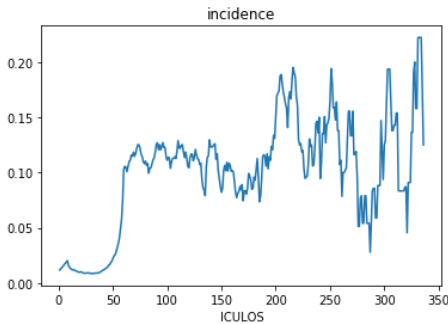

Figure 1. The incidence of sepsis along ICU LOS

2.2.   TASP architecture

According to the findings of data exploration, we propose **TASP**, a time-phased mode for sepsis prediction.

The core principle is that applying different approaches for different ICU stages. Figure 2 shows the architecture of TASP, which is a combination of three sub-models along time dimension. These sub-models, including two tree-based methods and a RNN-based method, are trained on different parts of data with different feature sets. Each hour data is an instance with the officially defined label. Given an hour data of a new ICU stay, our model will choose the corresponded sub-model to make a prediction with the pre-defined cut-off.
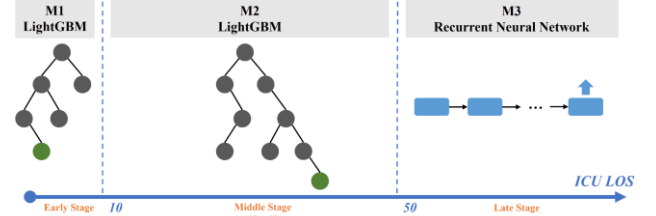

Figure 2. The architecture of TASP

M1: Sub-model for early stage

Among all the ICU patients, we found that the incidence of sepsis in the first 9 hours was significantly different from the following middle stage. Further study identified that there are 370 (12.6% of 2,932) patients got sepsis from the first hour of the ICU stay. It means that these patients have already been in serious conditions when they entered in ICU. For these patients, there are little historical data for prediction. Therefore, we proposed an independent tree-based model that trained on the first 9 hours of all patients to predict the sepsis risk in early stage. We constructed a feature set for each instance (an hour) in early stage as follows (the number in bracket is the feature number for each subset):

- **F1** (39): The original variables in this hour.
- **F2** (4): According to the prior knowledge of sepsis, we constructed HR/SBP, SaO2/FiO2, SOFA [3] and QSOFA [4] as the features.
- **F3** (40): Statistics of the 8 vital signs in the previous 6-hour slide window, including min, max, mean, std and difference between max and min.
- **F4** (48): The information from the 8 vital signs in the previous 1 and 2 hours, including the values in the time-stamps, the division and difference between the values in current hour and pre-hour.
- **F5** (8). The worst values (highest or lowest) of the 8 vital signs in the previous 6-hour slide window.

Note that all the erroneous values have been rectified or removed before feature set construction. With the 139 features, we utilized LightGBM [5] to train the model for early stage. LightGBM is an effective and efficient gradient boosting tree which has been widely used for classification and regression problems. Considering the ability of LightGBM that less prone to overfitting and more sensitive to outliers, we adopted a forward missing

imputation for all variables in early stage. The binary cross entropy between ground truth and predicted result is used as the loss function, which is same to the following two sub-models.

M2: Sub-model for middle stage

There are nearly 70% instances located in middle stage (10-49 hours). Compared to the early stage, the incidence of sepsis in this stage is relative low, while the observation time becomes longer (more historical hours). Therefore, we modified the tree-based method for early stage from two aspects.

- **Feature set**. We introduced two new slide windows, 12-hour and 24-hour, for F3 and F8. In addition, we added the number of measurements of all variables (without imputation) into the feature set, which represents the measure frequency. Similarly, forward missing imputation is used in this stage.
- **Label weight**. Due to the low incidence, we increased the weights for positive instances (label=1).

It is worth mentioning that this model is trained on all data, rather than the instances in middle stage. The reason is that the model can benefit from the information in other stages.

M3: Sub-model for late stage

Compared to the previous two stages, the instances in late stage take the lowest percentage, while the incidence of sepsis of them is the highest. It basically aligns with the clinical practice that if a ICU patient can't be transferred out for a long time, the condition may become more serious. However, the time-stamps (ICU LOS) of instances are varied, from 50 hours to 336 hours in offline training dataset. There exist long-term dependencies between the instances and the previous hourly data.

Recurrent Neural Network (RNN) is an excellent framework to capture the temporal relations in sequential data. A series of works adopted RNNs for different clinical event prediction tasks, including heart failure [6], readmission [7], future diagnosis [8], etc. In this study, we presented a RNN for sepsis prediction in late stage. It is an extension of GRU-D [9], which is designed for multivariate time series with missing values. With a learnable decay rates, GRU-D can measure influence of the missing parts for final prediction. We firstly use the original 39 variables as the input for GRU-D to generate the hidden layer for each hour. Then, we concatenate the hidden layer with the constructed feature set (like M2, without original variables). After that, a multi-layer perceptron is used to make a prediction for sepsis in each hour.

3. Experiments

In this section, we will show the implementation details and experimental results, including the offline scores of the complete TASP and the online scores of two simplified TASP.

3.1. Implementation details

The two tree-based sub-models for early and middle stages are implemented by official library. The core common parameters include learning rate as 0.01, number of leaves as 70, minimum number of instances in each leaf as 1,000. We set the positive instance weight as 1 for early stage and 5 for middle stage.

The RNN sub-model for late stage is implemented by Pytorch 1.1. The core parameters include hidden layer size as 39, learning rate as 0.0005. Adam [10] is used as the optimizer.

Actually, most of the parameters in the three sub-models are in default values. For each sub-model, we used a 10-fold cross validation on offline training dataset to determine its best cut-off.

3.2. Results

Offline

Table 1 illustrates the score of TASP on offline training dataset by a 10-fold cross validation. As we can see, the performance of TASP on each fold ranges from 0.3985 to 0.4462. The score in late stage is the highest due to the high incidence of sepsis. The total score by 10-fold cross validation is 0.4149 (it is not the average value of all the folds).

Table 1. Score of TASP on offline training dataset.

Fold	Early (1 ~ 9)	Middle (10 ~ 49)	Late (50+)	All Stage
1	0.2385	0.1779	0.7824	0.4248
2	0.0686	0.2031	0.8046	0.4218
3	0.2139	0.1721	0.7873	0.4112
4	0.2492	0.1453	0.7787	0.4220
5	0.3039	0.1996	0.7087	0.3931
6	0.2324	0.1003	0.6641	0.3442
7	0.2817	0.1821	0.6726	0.4066
8	0.3064	0.1926	0.7307	0.4462
9	0.1220	0.2686	0.8011	0.4423
10	0.2276	0.2117	0.6851	0.3985
Total	**0.2225**	**0.1891**	**0.7532**	**0.4149**

As a combination of three single models, we can use SHAP [11] to interpret the prediction result. SHAP values can be seen as a unified measure of feature importance for

outcome. Figure 3 shows an example of the non-linear relation in M2 sub-model between temperature (X-axis) and the corresponded SHAP value (Y-axis). As we can see, when temperature is below normal value (around 37), the risk of sepsis is relative low, regardless of the ICU LOS. When temperature becomes larger than normal value, the risk of sepsis will increase rapidly, and longer ICU LOS usually higher risk. It basically aligns with the medical knowledge about sepsis.
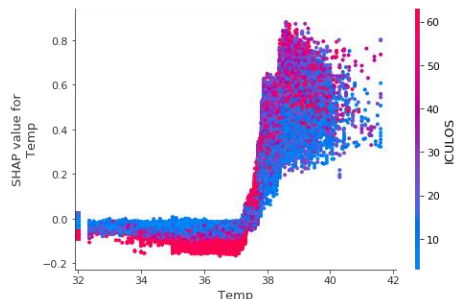

Figure 3. SHAP value of temperature

Online

We successfully submitted two simplified versions of TASP for online test. The first version is to use the sub-model of TASP for middle stage (M2) as the submitted model. It means that all the testing online instances will be inferenced by the single model with one cut-off. Similarly, the second version is the deep learning model for late stage (M3). The scores of the two simplified versions of TASP on leaderboard are 0.420 and 0.419, respectively.

4.   Discussion

According to the data exploration, TASP fuses three sub-models for different time segmentations. Tree-based models are used for early and middle stages with limited historical hourly data, while RNN-based model is utilized to capture the long-term time dependencies in late stage.

Compared to a unified prediction model with fixed feature set and cut-off, we believe that TASP is of better generalization and stability. It can be illustrated from the performance comparison in the offline training dataset that TASP outperforms the two submitted simplified versions of TASP, which do not distinguish the time phases. Moreover, as a combination of single models along time dimension, we can use some methods to interpret the prediction result.

Future work will investigate more concise feature sets for each sub-model, as well as a proper objective function which can better align with the official scoring mechanism.

5.   Conclusion

In this study, we proposed a time-phased model TASP for sepsis prediction. The core insight is that the incidence of sepsis is highly related to the ICU LOS. Two tree-based and one RNN-based sub-models are presented for the early, middle and late stage with different feature sets and different cut-offs. Experiments have shown that TASP achieves good performance in offline training dataset by a 10-fold cross validation. The two simplified versions of TASP obtains 0.420 and 0.419 scores in online test data.

References

[1] Singer, Mervyn, et al. "The third international consensus definitions for sepsis and septic shock (Sepsis-3)," *JAMA*, vol. 315, no. 8, pp. 801-810, Feb. 2016.

[2] Bengio, Yoshua, et al. "Representation learning: A review and new perspectives," *PAMI*, vol. 35, no. 8, pp. 1798-1828 Aug. 2013.

[3] Vincent, J-L., et al. "The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure," *Intensive Care Medicine*, vol. 22, no. 7, pp. 707-710, Jul. 1996.

[4] Angus, Derek C., et al. "A framework for the development and interpretation of different sepsis definitions and clinical criteria," *Critical Care Medicine*, vol. 44, no. 3, pp. 113, Mar. 2017.

[5] Ke, Guolin, et al. "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in Neural Information Processing Systems*, pp. 3146-3154, 2017.

[6] Choi, Edward, et al. "Using recurrent neural network models for early detection of heart failure onset," JAMIA, vol. 24, no. 2, pp: 361-370, Aug. 2016.

[7] Pham, Trang, et al. "Predicting healthcare trajectories from medical records: A deep learning approach," Journal of biomedical informatics, vol. 69, pp: 218-229, May. 2017.

[8] Lipton, Zachary C., et al. "Learning to diagnose with LSTM recurrent neural networks," arXiv preprint arXiv:1511.03677, Nov. 2015.

[9] Che, Zhengping, et al. "Recurrent neural networks for multivariate time series with missing values," Scientific reports, vol. 8, no. 1, Apr. 2018.

[10] Kingma, Diederik P., et al. "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, Dec. 2014.

[11] Lundberg, Scott M., et al. "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Sy*stems, pp. 4765-4774, 2017.

Address for correspondence:

Xiang Li.
B-9 PingAn International Financial Center, Beijing, China.
LIXIANG453@pingan.com.cn