An Ensemble Machine Learning Model For the Early Detection of sepsis from Clinical Data

Mengsha Fu[1],Jiabin Yuan[1],Menglin Lu[2],Pengfei Hong[3],Mei Zeng[4]
[1] Nanjing University of Aeronautics and Astronautics, Nanjing, China
[2] Dalian University of Technology,Dalian,China
[3] Beijing Wodong Tianjun Information Technology Co.Ltd, Beijing, China
[4] QiLu Hospital of ShanDong University,Jinan,China

Abstract

Sepsis is a life-threatening disease with high mortality and expensive treatment. In order to improve the outcomes of patients, it is important to detect at-risk patients with sepsis at an early stage. This study focused on improving predicting sepsis 6-12 hours before the clinical diagnosis by using the most recent definition of Sepsis-3. Unlike the previous work, the scoring metric that rewards early predictions and penalizes false alarms. A stacking model, which combined boosting and bagging tree models (lightgbm, xgboost and random forest as base-learner and Logistic regression as meta-learner) was designed to predict whether a patient will develop sepsis or not in the next 6 hours based on the current hour and past record clinical data. The stacking model utilized the preprocessing data and achieved a higher evaluate utility score than a single ensemble tree model.

1. Introduction

Sepsis is a life-threatening disease when the body's response to infections cause tissue damage, organ failure or death occurred [1]. Sepsis has become a global public health problem due to high morbidity, mortality and complex pathogenesis, especially in the intensive care unit (ICU). Sepsis is also regarded as a costly disease, the United States cost from $20 billion in 2011 increased to more than $23 billion in 2013, approximately is 6.2 % of all US hospital fee[2]. Early detection and targeted treatment such as antibiotics have been shown are critical to improve sepsis outcomes. Delayed treatment per hour is associated with an approximately 4-8 % increase in mortality[3,4]. The definition of sepsis is also constantly updated. The new recent definition of sepsis-3 is different from the previous criterion, diagnose patients septic if their Sequential Organ Failure Assessment(SOFA) score identified by a two-point deterioration within a 24-hour period[1]. The aim of the Physionet/CinC 2019 challenge[5] was for early 6-12 hours detection of sepsis (sepsis-3 definition) from clinical data.

Most of the researches about sepsis focused on the specific patient cohorts and used a different sepsis definition. Calvert et al[6] proposed a model called InSight for early detection of sepsis by systematic inflammatory response syndrome(SIRs) criteria. Jin, He, et al[7] research focused on trauma sepsis patients and Calvert, Jacob, et al [8] studied high-risk group aged 45 years or older patients for diagnosis of sepsis. The majority of previous work concentrated on a single-center hospital, data mainly from the public MIMIC database[9]. A robust model should be performed similarly when generalized to other hospital systems. So multi-center clinical data provides the possibility of testing for the versatility of the model. Recently many studies have used new definitions of sepsis-3. Nemati [10] demonstrated an interpretable machine learning for predicting sepsis onset 4-12 hour prior to clinical diagnosis. Multiscale blood pressure and heart rate dynamic feature extraction and Elastic Net logistic model were used to predict sepsis 4h prior to its onset by Shashikumar[11], Roman Z Wang [12]compared three models (LR/SVM/LMT) by extracting a random time window 48 to 6 hours prior to the onset of sepsis . Their model mostly applied a single machine learning method, stacking model is not utilized for early sepsis issue to enhance the effect. In addition, evaluation matrices used in these studies were traditional scoring function, such as AUROC and AUPRC), those evaluations are not a clinically significant way to reward or punish early detection of false positives or overtreatment. Therefore, the challenge describes a novel measure of a solution to the problem.

2. Methods

Data was prepared by the challenge organizers, which were from three different electronic medical record systems and hospitals. Hospital-A included 20336 patients (sepsis patients: 1790(8.8%), non - sepsis:18546(91.2%)). Hospital-B included 20000 pa-

tients (sepsis:1142(5.7%),non-sepsis:18860(94.3%)).The two labeled sets were posted for public download and Hospital-C contained 24819 patients from the third hospital system were sequestered as hidden test sets. Every patient has a file hourly record the clinical data with 40 variables (e.g. heart rate, systolic blood pressure) and 1 sepsis label. (0 means not sepsis in the next 6 hour,1 means sepsis happen in the next 6 hour). A large number of values were missing because measurements were not so frequent and were condensed into hourly bins. Positive and negative samples are extremely imbalance. To build a robust model, preprocessing data is necessary.

2.1. Preprocessing

Considering close to the real world of true clinical data, missing and erroneous data were intentionally not removed as part of the challenge. Firstly we analyze the distribution of data: The shortest ICU stay record was 8 hours and the longest was 336 hours, most hospital length of ICU stay are 20-35 hours in the two hospitals. About the sepsis patients, we excluded sepsis patients who labeled 1 from the first hour. 203 and 223 sepsis patients were respectively removed from hospital A and B. Additional sepsis patients record from the first time labeled1 to the end record less than 6 hours were also excluded to prevent a condition that only in the last few hours patients has a label of 1. At last, A/B hospital kept 1587/909 sepsis patients, non-sepsis patients did not change.

Then dealing with the missing data values. We summarized 40 features missing rate and founded variables were missing very badly especially laboratory values. EtCO2 was eliminated for the reason that hospital A has no value records. We computed the mean values of each feature from the two hospitals separately . Our impute strategy was using the" pad" method, called carry-forward , where filling the missing value with the previous non-missing value. If a patient lack of any values of records variable, groups mean values were utilized to impute. In the medical field, generally speaking, missing value sometimes represents the normal value or keep the same as last measurements. Table 1 showed part of hospital A/B overall mean values of each feature.

As for sample imbalances problem, the negative and positive ratio was close to 10:1. In our method, every hour record was taken a sample, so we just chose different sample sizes. Not all the negative sample were needed. For example, 1587 sepsis patients record files from hospitals A were used to train, which included 103196 samples (only 15368 hours labeled as 1). In this case, We just used septic patients records that already included enough negative samples, adding more negative samples was not necessary. According to the score function,it rewards 6-12 hours early prediction, We shifted the label, and the first time labeled 1 been moved forward for 12 hours based on the original label in order to get more rewards.

Table 1. A part of A/B sets overall feature mean value. The overall feature mean values were calculated to fill NaN values. if a patients missing all feature values, then filled with mean value, otherwise used the previous record values to pad.

Features	A_mean	B_mean	AB_mean
HR	84.99	84.14	84.56
Temp	37.02	36.92	36.97
SBP	120.96	126.5	123.7
MAP	78.76	86.36	82.56
DBP	59.98	66.23	63.109
Resp	18.77	18.67	18.72
pH	7.38	7.37	7.37
Sa02	91.25	96.56	93.89
Creatinine	1.40	1.64	1.52
Bilirubin_direct	3.11	1.06	2.05
Lactate	2.46	2.98	2.72
Bilirubin_total	2.69	1.69	2.1
WBC	11.93	10.72	11.32
Platelets	199.61	191.45	195.53
Age	63.01	60.96	62.00

2.2. Feature extraction

Based on the medical knowledge, we excluded 6 variables :EtCO2,Unit1,Unit2,Gender,BaseExcess, HospAdmTime, which were not related to sepsis or lack of values in hospital system.34 variables were kept and we want to find the remained features which are important and related to sepsis. Three tree-based models (lightgbm,xgboost and random forest) were used to rank the importance of the features by 5-fold cross-validation. The Figure 1 showed the ranked average feature importance by the three models. We conducted two experiments about different numbers of features. First, we used the original 34 features and to add rich representations from the hours records data, maximum, minimum, mean values, and trend information also been calculated. Features extended to 170 dimensions. The second experiment was used 15 features:HR,Temp,SBP,MAP,DBP,Resp,pH,SaO2, Creatinine, Bilirubin-direct,Lactate,Bilirubin-total,WBC,Platelets,-Age. Those were selected according to feature importance and reference to related literature[13, 14].The experiments in these articles were based on the selection of few important and easily available features for prediction. Besides, trend information about the difference between the predicted current hour and the previous hour also added.At last 30 features were used to train the model.

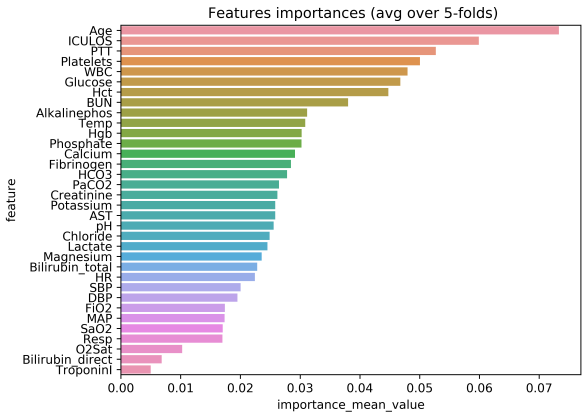

Figure 1. Feature importances score

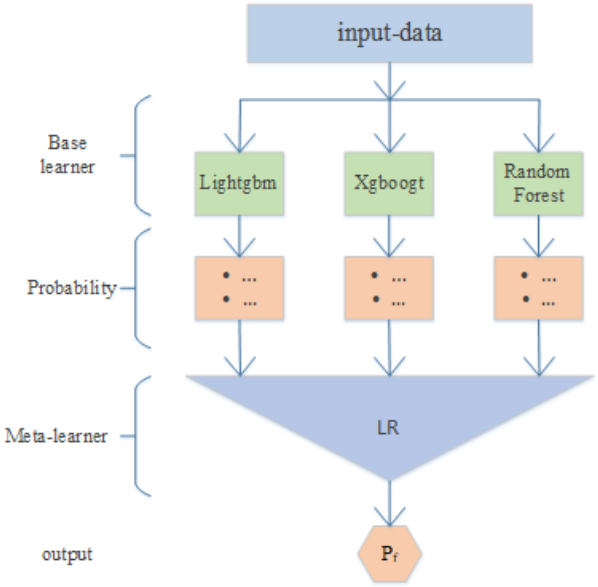

Figure 2. structure of stacking model. Base-learner included (lgb,xgb,rf) three models,each model trained input data and generated predicted probability then concatenate them to the meta-learner(LR) to train and got the last output.

2.3. Model development

With the development of data-driven, machine learning approaches about clinical data were applied widely. Stacking is an ensemble learning technique that combines multiple base models with a meta-learner. Different from the homogenous ensembles like bagging and boosting models use the same type of learner, stacking is a typical representative of heterogeneous ensembles can provide better prediction results than a single model. We proposed a stacking model which combined three different predictive models (lightgbm,xgboost and random forest) as a base-learner and Logistic Regression (LR) as a meta-learner. The structure of the stacking model was shown in Figure 2.

3. Results

Limited on the number of submissions, We cant test and compare every single model and the stacking model with different features and parameters setting. We mainly compared lightgbm,as a typical boosting model and the stacking model in the online hidden datasets.The method was evaluated by using a novel metric that created for this challenge. The utility-based scoring metric[5] would reward early detection or penalize false alarms. It is quite unlike the tradition scoring metrics, the ideal value for the utility score is 1, and higher values indicate better discrimination.

In the first experiment, 170 expanded features from 34 original variables, about 56000 samples were used to train the lightgbm and stacking model. After testing full hidden datasets online, The stacking model received a slight better utility result than the single lightgbm (0.04 versus 0.02).

In the second experiment, we changed the feature dimension based on the first submission results.15 selected features add their trend features to construct 30 features. 147172 hourly records from two hospitals as training samples were employed. The stacking model score improved to 0.14. Although the prediction effect of the stacking model was not outstanding on the online test datasets. It proved that the stacking model was better than a single model to a certain extent and provided a new idea to predict sepsis at least.

4. Discussion and Conclusions

We have developed a stacking model for the early 6-12 hours detection of sepsis from clinical data. Compared to a single ensemble tree-based model, our method can provide a slight improvement utility-based score.The parameters and structure of the model also need to be optimized to get a better prediction. Early sepsis prediction for patients in ICU is still a challenging but significant problem. A stacking model was developed to predict 6-12 hour before sepsis happening according to Sepsis-3 clinical criteria. The proposed method provides a new forecasting idea and has a slightly better result than a single ensemble model. Moreover, We can also score the importance of each feature to find the impact factors that are closely related to sepsis. The limitation of the model about generalization ability needs to be improved. When dealing with real-world clinical data, data preprocessing is more important. Not only some domain knowledge to build meaningful feature engineering is needed, but how to deal with miss-

ing and imbalances medical data problem is worth studying. Further studies will be conducted to more exploration and data analysis work.Trying different fill strategies(e.g. K-means clustering, Expectation maximization and Multiple Imputation) for missing values and using oversampling or undersampling methods for unbalanced data processing, those are very worthy of comparison and exploration. Moreover, model fusion methods about integrating deep learning techniques, such as LSTM, which could learn the intrinsic link between time series data, or other machine learning models maybe produce a potential performance improvement.

Acknowledgments

No

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] Torio C, Andrews R. National inpatient hospital costs: the most expensive conditions by payer, 2011: statistical brief# 160 2006;.

[3] 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.

[4] 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.

[5] Reyna MA, Josef C, Jeter R, Shashikumar SP, M. Brandon Westover MB, Nemati S, Clifford GD, Sharma A. Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019. Critical Care Medicine 2019;In press.

[6] Calvert JS, Price DA, Chettipally UK, Barton CW, Feldman MD, Hoffman JL, Jay M, Das R. A computational approach to early sepsis detection. Computers in biology and medicine 2016;74:69–73.

[7] Jin H, Liu Z, Xiao Y, Fan X, Yan J, Liang H. Prediction of sepsis in trauma patients. Burns trauma 2014;2(3):106.

[8] Calvert J, Saber N, Hoffman J, Das R. Machine-learning-based laboratory developed test for the diagnosis of sepsis in high-risk patients. Diagnostics 2019;9(1):20.

[9] Johnson AE, Pollard TJ, Shen L, Li-wei HL, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG. Mimic-iii, a freely accessible critical care database. Scientific data 2016;3:160035.

[10] Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the icu. Critical care medicine 2018;46(4):547–553.

[11] Shashikumar SP, Stanley MD, Sadiq I, Li Q, Holder A, Clifford GD, Nemati S. Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. Journal of electrocardiology 2017;50(6):739–743.

[12] Wang RZ, Sun CH, Schroeder PH, Ameko MK, Moore CC, Barnes LE. Predictive models of sepsis in adult icu patients. In 2018 IEEE International Conference on Healthcare Informatics (ICHI). IEEE, 2018; 390–391.

[13] van Wyk F, Khojandi A, Mohammed A, Begoli E, Davis RL, Kamaleswaran R. A minimal set of physiomarkers in continuous high frequency data streams predict adult sepsis onset earlier. International journal of medical informatics 2019;122:55–62.

[14] Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, Shimabukuro D, Chettipally U, Feldman MD, Barton C, et al. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. JMIR medical informatics 2016;4(3):e28.

Address for correspondence:

Mengsha Fu
School of Computer Science and Technology

No.29 JiangJun Road,Jiangning district,Nanjing,China

mengshafu@nuaa.edu.cn