

基于迭代决策树的ICU临床干预预测

张娅峰, 龚 振

(华南理工大学, 广东 广州 510641)

摘要:在重症监护室中,临床干预的实时预测仍然是一个挑战。由于近来数字化趋势的发展,医院记录的信息越来越多。医生可以访问有关患者的大量数据,但能用来处理数据的时间和工具很少。智能的临床决策支持可以为医生提供患者何时需要特定干预的预测信息。面对ICU病房数据密度大、质量高的特点,善于处理海量数据的机器学习算法吸引了医疗界的关注。通过提取患者在ICU期间产生的动态时间序列数据及患者静态人口学数据进行整合,使用GBDT、SVM等机器学习算法开发模型,学习这些数据的表现以预测患者何时需要进行连续肾脏治疗干预。预测是以一种前瞻性的方式进行的,以实现“实时”的性能。实验表明GBDT模型的准确率和召回率均达到80%以上。基于GBDT的ICU临床干预预测模型能够辅助临床医生进行风险预警,及时采取干预措施从而改善患者预后。

关键词:机器学习;重症监护室;干预;预测;迭代决策树

中图分类号:TP181

文献标识码:A

文章编号:1673-629X(2020)12-0118-05

doi:10.3969/j.issn.1673-629X.2020.12.021

ICU Clinical Intervention Prediction Based on GBDT

ZHANG Ya-feng, GONG Zhen

(South China University of Technology, Guangzhou 510641, China)

Abstract: In the intensive care unit, real-time prediction of clinical interventions remains a challenge. As a result of the recent trend towards digitization, a growing amount of information is recorded in hospitals. Doctors have access to a wealth of data about patients, but they have little time and tools to process the data. Intelligent clinical decision support can provide doctors with predictive information about when patients need specific interventions. Facing the characteristics of high density and high quality of data in ICU wards, machine learning algorithms that are good at processing massive data have attracted the attention of the medical profession. By extracting the dynamic time series data and static demographic data generated by the patient during the ICU, the machine learning algorithms such as GBDT and SVM are used to develop models to learn the representation of these data to predict when the patient needs continuous renal therapy intervention. Forecasting is done in a forward-looking way to achieve “real-time” performance. Experiment shows that the precision and recall of the GBDT model are both above 80%. The GBDT-based ICU clinical intervention prediction model can assist clinicians in early warning of risks and timely intervention measures to improve patient prognosis.

Key words: machine learning; intensive care unit; intervention; prediction; gradient boosting decision tree

0 引言

“AI+医疗”目前成为热门领域,将AI技术应用到医疗领域是发展的大趋势,近年来,许多成功的研究成果发表在《Nature》、《Cell》、《JAMA》、《NEJM》等顶级期刊上,受到了学术界、工业界和监管机构的极大关注。人工智能,特别是深度学习,是开始用于解释医学影像和电子健康记录的主要技术工具之一。而在ICU科室,因为收治患者的特殊性,每位患者都包含大量的实时生命体征数据以及其他干预、实验室检查等数据,是人工智能应用的完美场景^[1-2]。目前已有大量研究

以患者死亡风险作为预测任务进行建模^[3-5],因为预测患者的死亡风险,可以合理分配抢救的资源,及早地进行干预以提高患者的生存率,该任务主要根据固定的时间间隔(通常是入ICU后的前24小时)的数据进行预测,不包含患者在整个ICU住院期间不断产生的临床数据。一些学者则以特定的疾病发生为预测任务建立预测模型,通常选定的疾病是会对患者生命造成严重影响的。Savin等人通过树学习模型的机器学习方法对NICU(神经重症监护)医源性脑室炎、脑膜炎的发生进行预测,并试图探究主要影响因素^[6]。Jawad

收稿日期:2020-01-01

修回日期:2020-05-06

基金项目:广东省医学科研基金项目(A2019409)

作者简介:张娅峰(1994-),女,硕士,CCF会员(A3909G),研究方向为大数据分析与应用。

I 等人则对 ICU 患者感染性休克发生的概率进行预测,建立的感染性休克预警模型,使得患者的 ICU 存活率得到明显提升^[7]。ICU 是临床科室中数字医疗设备最多的地方。ICU 的设备必须配有床边监护仪、中心监护仪、多功能呼吸治疗机、麻醉机、心电图机、除颤仪、起搏器、输液泵等。这些数字医疗设备产生的数据、医护人员操作的数据以及患者在其他科室及实验室产生的数据使得 ICU 成为多种信息交汇的地方。通常可以大致分为以下几大类:(1)人口统计学数据,即患者的姓名、出生日期、出生地等等;(2)生命体征数据包括血压、体温、心率、心电等,由仪器自动实时采集,频率为每五秒钟记录一次;(3)实验室检验检查数据,通常医生根据患者病情做出是否需要做该项检验的判断,从而产生相应的数据,因此数据频率不确定,可能是一天也可能是几天;(4)患者的用药记录也包括患者的出入量的详细记录。这部分数据大部分由护士实际操作后记录下来;(5)干预措施,例如气管插管或者发生抢救行为等;(6)医生写的文本格式的病程记录。而基于这些高维的监测数据,数据源嘈杂,稀疏,异构且结果不平衡,预判患者在何时需要进行干预治疗对 ICU 临床工作人员来说是一项非常具有挑战性的工作,在不恰当的时机进行干预可能会对患者造成二次伤害。相比于人类专家,AI 算法在执行任务的过程中具有更快的速度、更高的一致性和可重复性,使得临床医生和人工智能在共同工作时能产生 1+1>2 的协同效应。

急性肾损伤(AKI)是重症监护病房中感染性休克的患者的常见并发症,通常伴随较高的死亡率^[8-10]。大约 70% 的 AKI 患者需要进行肾脏替代治疗(RRT),住院死亡率超过 60%^[11]。连续肾脏替代治疗(CRRT)是重症监护室中一种常见的临床干预措施。通常使用各种用于代谢,溶质和体液控制的参数来指导治疗的开始和终止^[12]。CRRT 可以提高 ICU 的存活率,因为它通过调节乳酸盐的清除,不可测的阴离子以及磷酸盐和氯化物的含量来纠正代谢性酸中毒。但 ICU 患者的连续肾脏替代疗法(CRRT)的提供情况千差万别,主要还是基于经验主义和当地机构的实践和资源。有经验的医生通过长时间在床旁观察患者实时的生命体征数据可以及时发现患者病情好转或者恶化的信号,但这个过程非常耗时,并且一位医生不能同时兼顾多位患者。机器学习算法可以利用所有可用信息来预测每个患者未来可能发生的事件,并且能够捕获大量随时间变化的变量之间的复杂关系。因此该文使用机器学习算法对 CRRT 干预启动进行建模预测,从历史干预的数据中学习,从而形成相应的预警模型,使得医生能够在忙于抢救某一患者的同时,也不耽误对

其他患者病情的监测和预警;同时,在适当的时候采取适当的干预措施对于改善败血症性休克肾衰竭患者的医院预后非常重要。

1 梯度提升迭代决策树相关理论

GBDT 是一种迭代的决策树算法,全称为 (gradient boosting decision tree)。Freidman 在 1997 年提出 GBDT 算法,GBDT 属于一种集成学习算法,通过构建并结合多个学习器来完成学习任务,常常可以获得比单一学习器显著优越的泛化性能^[13]。针对输入的训练数据集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, $x_i \in X \subseteq R^N, y_i \in Y \subseteq R$; 损失函数 $L(y, f(x))$; 输出为回归树 $f(x)$ 。GBDT 无论是用于分类和回归,采用的都是回归树,分类问题即最终将拟合值转换为概率来进行分类。

算法的迭代步骤如下:

(1) 初始化弱学习器:

$$f_0(x) = \arg \min_c \sum_{i=1}^N L(y_i, c) \quad (1)$$

(2) 对于 $m = 1, 2, \dots, M$:

(a) 对 $i = 1, 2, \dots, M$; 计算负梯度,即残差。

$$r_{mi} = - \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x) = f_{m-1}(x)} \quad (2)$$

(b) 对 r_{mi} 拟合一个回归树,得到第 m 棵树的叶节点区域 $R_{mj}, j = 1, 2, \dots, J$ 。

(c) 对 $j = 1, 2, \dots, J$, 计算最佳拟合值。

$$c_{mj} = \arg \min_c \sum_{x_i \in R_{mj}} L(y_i, f_{m-1}(x_i) + c) \quad (3)$$

(d) 更新强学习器。

$$f_m(x) = f_{(m-1)}(x) + \sum_{j=1}^J c_{mj} I(x \in R_{mj}) \quad (4)$$

(3) 得到最终学习器。

$$\hat{f}(x) = f_M(x) = \sum_{m=1}^M \sum_{j=1}^J c_{mj} I(x \in R_{mj}) \quad (5)$$

算法第 1 步初始化,估计使损失函数极小化的常数值,它是只有一个根节点的树。第 2(a)步计算损失函数的负梯度在当前模型的值,将它作为残差的估计。对于平方损失函数,它就是通常所说的残差;对于一般损失函数,它就是残差的近似值。第 2(b)步估计回归树叶节点区域,以拟合残差的近似值。第 2(c)步利用线性搜索估计节点区域的值,使损失函数极小化。第 2(d)步更新回归树。第 3 步得到输出的最终模型 $\hat{f}(x)$ ^[14]。

2 数据来源及预测任务

2.1 数据来源

该文数据来自于某三甲医院,该 ICU 是国家重点

专科所在科室,在功能配置、医疗技术和管理方面达到世界先进水平。自重症医疗信息系统 2017 年底上线以来,对于患者的数据进行了非常详实的记录,主要包括:病人基本信息,生命体征观察项,导管管理,液体管理,护理记录,病程记录等,能够为该文的研究提供精细化的数据和优质的经验。截至 2019 年 7 月,系统采集了大约 752 位患者 ICU 记录。由于 80% 患者的住 ICU 时长少于 15 天,因此将总时长小于 6 小时(避免没有充足的信息用来做决策)以及大于 360 小时(避免一些病情过重的患者)的患者排除,这样就可以专注于那些通过干预有机会康复的患者。正如大多数文献所提到的,该文只考虑每位患者的首次入科的完整记录。根据这些条件过滤后,最终获得了 522 名患者的队列。

2.2 数据提取与预处理

数据预处理是数据挖掘中的关键步骤,是指在分析建模之前对数据进行人为处理,数据预处理的质量对后续数据的正确分析具有巨大的影响,甚至决定了最终的性能表现。原始数据往往是脏数据,含有大量的噪声,数据预处理主要是对原数据进行清洗,筛选出与研究目的相关的数据,并去除噪声。因此对于符合队列选择标准的 N 个患者队列中的每个患者 n ,从重症信息数据库中提取以下三个数组:临床观察的时间序列 x_n 以及相对应的临床干预状态 y_n ,与此同时,还提取了患者静态人口学相关信息。

(1) 临床观察数组 $x_n = [x_{n1}, x_{n2}, \dots, x_{nt}, \dots, x_{nT}]$ 在每一个时刻有 34 个变量,其中包含生命体征变量:中心静脉压,收缩压,舒张压,心率,呼吸频率,体温,血氧饱和度,吸氧浓度;实验室检查指标:阴离子间隙,肌酐,尿素氮,酸碱度,血红蛋白浓度,钾离子浓度,乳酸等以及出入量指标(尿量);生命体征数据由监护仪每

5 秒钟产生一次,但是系统通常在 5 ~ 60 分钟保存一条记录,实验室检查指标通常 ICU 患者每日至少会抽取动脉血或者静脉血一次进行检测从而报告相应的数值。出入量中的尿量值一定程度上能够反映患者肾功能的情况,因此通常每小时记录一次,数值为此刻与上一记录时刻的时间间隔内患者的尿量累计值。在数据预处理阶段,首先将患者的所有临床观测记录在时间轴上进行对齐,时间索引处理为入科小时数;若同一小时内有多个信号测量值,其中生命体征以及实验室检查变量取这些测量结果的平均值;对于尿量值,则取多条记录的总和。对于记录中的缺失值,采用前向填充法,即用上一时刻的取值替代,直到有新的测量值产生。

(2) 患者 n 的干预状态 $y_n = [y_{n1}, y_{n2}, \dots, y_{nt}, \dots, y_{nT}]$ 是一个二进制时间序列,每一个 y_{nt} 表示患者 n 在 t 时刻是否进行了干预。一位患者进入 ICU 后可能会发生多次 CRRT 干预,该文考虑对患者进入 ICU 后的首次干预进行预测。

(3) 静态人口学信息 s_n 包含性别、年龄、身高、体重;以入科时记录为准,并且在患者的该次 ICU 住院期间不会发生改变。基于患者的身高和体重,计算了一个新的 BMI 变量。对于其中部分患者身高或者是体重的缺失,采取的方法是分男女分别拟合一个身高与体重的回归方程用于填补缺失值。对于静态变量,则将其在每位患者自身的所有时间序列上不断往后复制。

因此,对于选定的患者队列 N ,每位患者都具有与之关联的数据 $\{x_n, y_n, s_n\}_{n=1}^N$ 。最终通过数据预处理之后,形成每位患者每小时一条记录。数据处理流程图见图 1。

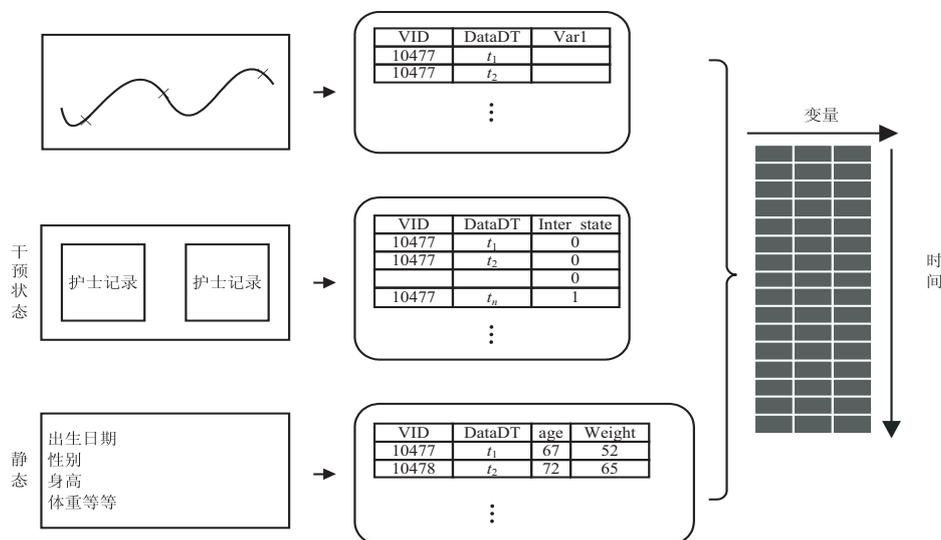


图 1 数据预处理及特征提取

2.3 预测任务及模型构建

文中的 ICU 临床干预的预测任务是一个有监督的二分类问题^[15-16]。经过数据预处理后,此时每个时间步长的间隔为一小时。因此,在患者每次 ICU 住院期间的一小时间隔中,必须预测是否实施干预。为实现“实时”预测的效果,设计了如图 2 所示的预测框架,并依此在患者队列 $\{x_n, y_n, s_n\}_{n=1}^N$ 上进行采样。目标是对需要进行 CRRT 干预的患者进行早期预警,使得医护人员能够及时采取干预措施,从而改善患者预后。即对任一患者 n ,在其住 ICU 期间的任一时刻 t ,使用 $\{x_{nt-2}, s_n\}, \{x_{nt-1}, s_n\}$ 的 2 小时观察窗口的患者数据,预测间隔一小时后的时长为 3 小时 ($t+1, t+4$) 的预测窗口内是否需要采用干预措施。间隔的小时数

可以给予临床医护人员充足的时间做好干预的相应准备工作。整个采样的过程以滑动窗口的形式进行,以患者入 ICU 科室时间作为起点,以 2-1-3 的时间窗口向后滑动,每次滑动 1 小时。对于纳入研究范围的无干预患者,划分样本时使用其完整的时间序列数据;对于有干预的患者,仅对其首次干预进行预测,即该患者首次干预后的时间序列数据予以丢弃。窗口沿着整个患者记录滑动,从而形成多个样本示例^[17-18]。在 3 小时的预测窗口内若出现过干预视为正样本,否则视为负样本。采样的具体方式如图 2 所示。根据上述采样方法,共形成 2 289 条样本,其中正样本 1 982 条,负样本 307 条。

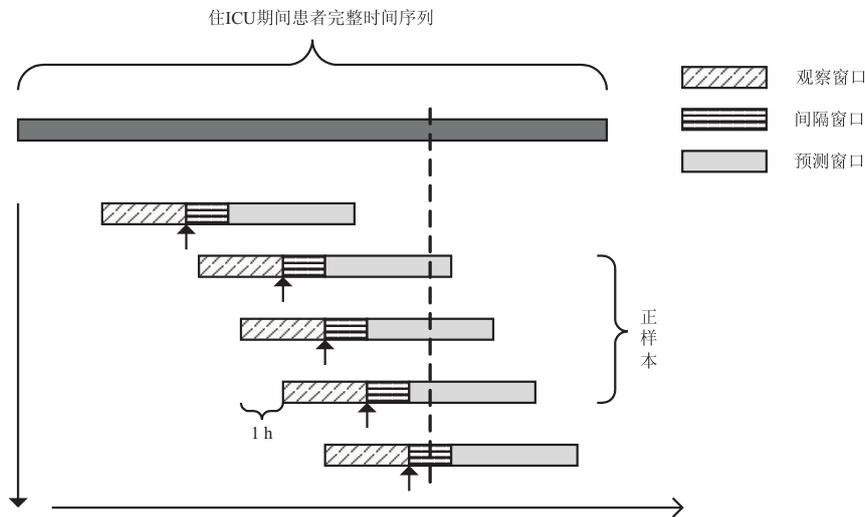


图 2 采样方式

3 实验及结果分析

3.1 评估指标

该文将临床干预预测问题抽象为机器学习中的二分类问题,因此采用机器学习中常用的准确率 (precision)、召回率 (recall) 以及 F1 分数作为实验的评估指标。准确率是指:对于给定的测试集数据,被分类器判断为需要干预的样本中正确的比例。

$$\text{precision} = \frac{TP}{TP + FP} \tag{6}$$

召回率即所有实际发生干预的患者中被模型预测对的比重。

$$\text{recall} = \frac{TP}{TP + FN} \tag{7}$$

与此同时, F1-Score 指标可以看作是准确率和召回率的加权平均,综合了二者的产出的结果。F1-Score 的取值范围从 0 到 1, 1 代表模型的输出最好, 0 代表模型的输出最差。该文分别对 6 种参考模型计算 precision、recall 以及 F1-Score 进行分类性能对比^[19]。

3.2 实验及结果分析

按照 7 : 3 的比例将所有样本划分为建模样本和测试样本,为保证模型性能的稳定性,选用 5 折交叉检验方法进行评估。除迭代决策树模型外,还对比了逻辑回归、随机森林、多层感知机、支持向量机四种参考模型,并使用网格搜索方法,对上述模型分别尝试了 72, 72, 64, 72, 72 种参数组合。实验报告了每个模型在最优参数组合下的预测结果,如表 1 所示。

表 1 五种机器学习模型结果比较

模型	测试集准确率	测试集召回率	测试集 F1 值
逻辑回归	0	0	0
随机森林	0.871 795	0.781 609	0.824 242
多层感知机	0.538 462	0.080 46	0.14
支持向量机	0.583 333	0.885 057	0.703 196
迭代决策树	0.829 545	0.839 08	0.839 08

在使用机器学习的分类问题中,准确率和召回率通常是相互影响的。理想的情况是两者都取得比较高的值。因此,从综合结果来看,迭代决策树的准确率和

召回率均在0.8以上,F1分数最高,表现最优。其中,逻辑回归的结果最差,其性能远低于迭代决策树模型,这可能是数据线性不可分造成的。

4 结束语

就临床干预预测问题而言,目前利用机器学习方法对ICU临床数据进行干预预测的现有相关文献大多集中于MIMIC III医疗信息市场的重症监护(MIMIC)数据集^[20-21]。有研究表明,基于本地电子病历数据开发的预测模型的预测性能往往高于基于大队列研究的预测模型。因此,本研究中的结果基于本地数据,尝试多种机器学习模型,已经达到一定的准确率和召回率,能够满足一定的应用需求,高效的分类算法能够大大节省临床医生们的时间和精力,并提供重要的辅助信息。在未来的工作中,可尝试使用深度学习的方法处理这类时间序列数据,以便能够更好地提取和利用时间序列数据中包含的患者信息,提高预测性能。

参考文献:

- [1] SURESH H, GONG J J, GUTTAG J V. Learning tasks for multitask learning: heterogenous patient populations in the ICU[C]//Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. London: ACM, 2018: 802-810.
- [2] 朱志勇, 陈一昕, 李建功. 大数据技术在医疗急重症领域的应用[J]. 邮电设计技术, 2016(8): 28-32.
- [3] AWAD A, BADER-EL-DEN M, MCNICHOLAS J, et al. Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach[J]. International Journal of Medical Informatics, 2017, 108: 185-195.
- [4] PIRRACCHIO R, PETERSEN M L, CARONE M, et al. Mortality prediction in intensive care units with the super ICU learner algorithm (SICULA): a population-based study[J]. Lancet Respiratory Medicine, 2015, 3(1): 42-52.
- [5] 蔺轲, 林瑜, 孔桂兰. 基于XGBoost算法的ICU脓毒症患者住院死亡风险预测研究[J]. 中国卫生信息管理杂志, 2018, 15(5): 536-540.
- [6] SAVIN I, ERSHOVA K, KURDYUMOVA N, et al. Health-care-associated ventriculitis and meningitis in a neuro-ICU: incidence and risk factors selected by machine learning approach[J]. Journal of Critical Care, 2018, 45: 95-104.
- [7] JAWAD I, LUKŠIĆ I, RAFNSSON S B. Assessing available information on the burden of sepsis: global estimates of incidence, prevalence and mortality[J]. Journal of Global Health, 2012, 2(1): 010404.
- [8] HOSTE E A J, BAGSHAW S M, BELLOMO R, et al. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study[J]. Intensive Care Medicine, 2015, 41(8): 1411-1423.
- [9] MEHTA R L, CERDÁ J, BURDMANN E A, et al. International society of nephrology's Oby25 initiative for acute kidney injury (zero preventable deaths by 2025): a human rights case for nephrology[J]. The Lancet, 2015, 385(9987): 2616-2643.
- [10] QUENOT J P, BINGUET C, KARA F, et al. The epidemiology of septic shock in French intensive care units: the prospective multicenter cohort EPISS study[J]. Critical Care, 2013, 17(2): R65.
- [11] YOON B R, LEEM A Y, PARK M S, et al. Optimal timing of initiating continuous renal replacement therapy in septic shock patients with acute kidney injury[J]. Scientific Reports, 2019, 9(1): 11981.
- [12] SEABRA V F, BALK E M, LIANGOS O, et al. Timing of renal replacement therapy initiation in acute renal failure: a meta-analysis[J]. American Journal of Kidney Diseases, 2008, 52(2): 272-284.
- [13] FRIEDMAN J H. Stochastic gradient boosting[J]. Computational Statistics & Data Analysis, 2002, 38(4): 367-378.
- [14] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012: 146-152.
- [15] WU M, GHASSEMI M, FENG M, et al. Understanding vasopressor intervention and weaning: risk prediction in a public heterogeneous clinical time series database[J]. Journal of the American Medical Informatics Association, 2016, 24(3): 488-495.
- [16] LILLO-CASTELLANO J M, MORA-JIMENEZ I, SANTIAGO-MOZOS R, et al. Weaning outcome prediction from heterogeneous time series using normalized compression distance and multidimensional scaling[J]. Expert Systems with Application, 2013, 40(5): 1737-1747.
- [17] 胡满满, 陈旭, 孙毓忠, 等. 基于动态采样和迁移学习的疾病预测模型[J]. 计算机学报, 2019, 42(10): 2339-2354.
- [18] 陈旭, 刘鹏鹤, 孙毓忠, 等. 面向不平衡医学数据集的疾病预测模型研究[J]. 计算机学报, 2019, 42(3): 596-609.
- [19] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 28-37.
- [20] JOHNSON A E W, POLLARD T J, SHEN L, et al. MIMIC-III, a freely accessible critical care database[J]. Scientific Data, 2016, 3: 160035.
- [21] 张家艳, 郑建立, 郑西川, 等. MIMIC数据库智能挖掘研究概述[J]. 计算机技术与发展, 2020, 30(1): 144-148.