

DOI:10.19651/j.cnki.emt.2108646

基于贝叶斯优化 XGBoost 的无创血压预测方法*

孙斌 储芳芳 陈小惠

(南京邮电大学自动化学院、人工智能学院 南京 210023)

摘要: 为降低个体特征差异对无创血压预测模型精度的影响,提高预测准确率,提出建立贝叶斯优化(BO)XGBoost 的无创血压预测方法。该方法首先通过脉搏波传导时间(PTT)和身体质量指数(BMI)建立多元线性模型获得初步血压预测值;进一步结合人体特征参数作为 XGBoost 血压预测模型的输入;再运用贝叶斯优化对 XGBoost 超参数自动寻优,最终建立 BO-XGBoost 模型进行血压预测,并与其他方法对比。实验结果表明,BO-XGBoost 血压预测模型舒张压和收缩压测量值的平均误差满足美国医疗仪器促进协会(AAMI)制定的小于 5 mmHg 的标准,与水银血压计具有更好的一致性。

关键词: 血压预测;人体特征;贝叶斯优化;XGBoost

中图分类号: TP274 **文献标识码:** A **国家标准学科分类代码:** 510.8040

Non-invasive blood pressure detection method based on Bayesian optimization XGBoost

Sun Bin Chu Fangfang Chen Xiaohui

(School of Automation & School of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract: In order to reduce the impact of individual characteristic differences on the accuracy of the non-invasive blood pressure prediction model and improve the prediction accuracy, a Bayesian optimization (BO) XGBoost non-invasive blood pressure prediction method is proposed. Firstly, the multivariate linear model is established to obtain preliminary blood pressure prediction values based on pulse transit time (PTT) and body mass index (BMI). Then combine human characteristic parameters as the input of XGBoost blood pressure prediction model. Then use Bayesian optimization to automatically optimize XGBoost hyperparameters. Finally, the BO-XGBoost model is used to predict the blood pressure and compare with other methods. The experimental results show that the average absolute error of diastolic and systolic blood pressure based on the BO-XGBoost blood pressure prediction model meets the standard of less than 5 mmHg formulated by AAMI (American Medical Instrument Promotion Association), which is better consistent with the method of mercury sphygmomanometer.

Keywords: blood pressure detection; human characteristics; Bayesian optimization; extreme gradient boosting

0 引言

血压是日常生活和术前术后健康监测的重要指标之一,掌握血压的波动状态能够及时的进行疾病的预防和治疗^[1]。高血压是引起脑梗塞和心力衰竭等心血管疾病的关键危险因素,这些疾病具有较长的潜伏期,人体无法明显感知不适,然而此时血压已存在异常。高精度的血压检测是监测血压变异性^[2]的基础,因此实现无创血压的精确测量具有重要意义。

血压的测量方法可分为有创测量方法和无创测量方法。

有创测量方法采用侵入式测量,其中,动脉导管法能够及时准确的获取血压值^[3],但是具有感染、出血等多种风险,主要用于外科手术和重症监护室。无创血压测量方法能克服这些问题,并适用于日常血压监测^[4]。近年来,为提高连续无创血压预测精度,国内外学者进行了深入研究。文献[5]通过对光电容积脉搏波(PPG)和心电信号(ECG)预处理计算脉搏波传导时间(PTT),建立一元线性 BP-PTT 回归方程估算血压。张大可等^[6]基于脉搏波和心电信号提取时域特征参数,PPG 上升时间、周期比和峰值,并结合单一脉搏波传导时间(PWTT)建立血压模型。但是模型单一,在保证血压预

收稿日期:2021-12-21

* 基金项目:国家自然科学基金(61801239)项目资助

测精度的同时应考虑模型泛化能力。作为进一步改进,一些学者将机器学习算法应用于血压预测^[7-8]。Khalid 等^[9]从 PPG 信号提取时域特征,使用支持向量机(SVM)^[10]、决策树回归等算法进行血压预测,该方法需要准确的进行 PPG 时域特征提取并标准化处理。朱海龙等^[11]利用粒子群优化(PSO)算法优化 SVM 模型参数,建立基于差分自回归移动平均和支持向量机混合血压预测模型,未考虑人体特参数和血压预测模型的联系。师荣堃等^[12]实现基于循环神经网络(RNN)的血压预测模型,该模型预测精度优于传统的机器学习方法,但复杂度高,计算代价大。

针对上述存在的问题,本文提出一种基于贝叶斯优化的 XGBoost 无创血压预测方法。通过建立多元线性方程得到初步血压预测值,结合人体特征参数作为 XGBoost 血压预测模型的输入,进一步降低因个体特征差异造成的预测偏差。同时利用贝叶斯优化算法实现 XGBoost 模型的超参数自动寻优,建立 BO-XGBoost 血压预测模型。采用平均绝对误差,决定系数以及准确率对模型预测性能进行评估,实验结果表明 BO-XGBoost 模型预测血压具备更高精度。

1 基本原理

1.1 XGBoost 算法

XGBoost 属于 Boosting 集成学习算法,既能处理分类问题,也能处理回归问题^[13],多个弱学习器迭代学习实现强学习器。XGBoost 基于若干分类与回归树(classification and regression tree,CART)实现加法模型,每增加一棵新的决策树作为基学习器,去拟合上一次的预测残差,累加所有决策树的预测结果得到预测模型的最终结果。XGBoost 目标函数包括损失函数和正则项两部分,在梯度提升决策树(gradient boosting decision tree,GBDT)基础上加以改进,对损失函数使用二阶泰勒展开,引入一阶导数和二阶导数,可以有效控制模型过拟合,提高预测精度^[14-15]。适用于无创血压检测方法。

XGBoost 中 CART 树的加法模型表达式如下:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (1)$$

式中: \hat{y}_i 表示第*i*个样本预测值;*K*表示树的数目; x_i 表示输入的第*i*个样本数据; $f_k(x_i)$ 表示第*K*棵决策树; f_k 为树集合空间*F*中的一个函数。

XGBoost 回归模型在训练过程中向目标函数优化,原模型在每次迭代中保持不变,第*t*次迭代模型中加入新的函数,对新生成的决策树与之前所有树预测值总和的残差进行拟合,迭代过程表达式如下:

$$\begin{aligned} \hat{y}_i^{(0)} &= 0 \\ \hat{y}_i^{(1)} &= f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i) \\ \hat{y}_i^{(2)} &= f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i) \\ &\dots \\ \hat{y}_i^{(t)} &= \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \end{aligned} \quad (2)$$

XGBoost 的目标函数表达式如下:

$$Obj = \sum_{i=1}^N l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) = \sum_{i=1}^N l\{y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\} + \sum_{k=1}^K \Omega(f_k) \quad (3)$$

其中,第 1 部分为损失函数,可描述预测值和真实值的误差,第 2 部分为正则项,有效控制模型构建树结构的复杂度,防止过拟合,正则化项表达式如下:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad (4)$$

其中, γ 和 λ 表示加权因子,*T*为叶子节点个数, ω 表示叶子节点权重。

XGBoost 对目标函数使用泰勒展开法进行优化,对式(3)目标函数使用二阶泰勒展开式展开:

$$Obj = \sum_{i=1}^N \{l(y_i, \hat{y}_i^{(t-1)}) + g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i)\} + \Omega(f_i) \quad (5)$$

其中, g_i 表示损失函数一阶导数项; h_i 表示损失函数二阶导数项; $l(y_i, \hat{y}_i^{(t-1)})$ 为常数项,在目标函数简化过程中可去除并不影响结果。将正则项式(4)代入目标函数式(5),进一步简化并对 ω_j 求导,得到最优解如下:

$$\omega_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (6)$$

其中, I_j 表示叶子节点样本集; $\sum_{i \in I_j} g_i$ 可表示为 G_j ,

$\sum_{i \in I_j} h_i$ 可表示为 H_j ,则目标函数的最优解可表示为:

$$\omega_j^* = - \frac{G_j}{H_j + \lambda} \quad (7)$$

XGBoost 的最优目标函数为:

$$Obj = - \frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (8)$$

1.2 贝叶斯优化算法

贝叶斯优化(bayesian optimization,BO)是一种基于概率分布的全局优化算法。针对 XGBoost 模型的超参数寻优问题,在一组超参数的决策空间中,贝叶斯优化为待优化函数 $f: \mathcal{X} \rightarrow R^d$ 构造概率模型,进一步利用该模型选择一个评估点,依次迭代循环得到超参数最优解^[16]:

$$x^* = \operatorname{argmin}_{x \in \mathcal{X} \subseteq R^d} f(x) \quad (9)$$

其中, x^* 为最优超参数组合, \mathcal{X} 为决策空间, $f(x)$ 为目标函数。

贝叶斯优化算法主要核心步骤是先验函数和采集函数两部分。先验函数采用高斯回归过程,基于贝叶斯定理,将先验概率模型转换为后验概率分布。采集函数采用改进概率(probability of improvement,PI)选择下一个评估点^[17]。

1) 高斯回归过程

高斯过程属于非参数模型,也是随机变量的一个集合,由均值函数和核函数(协方差函数)确定,即:

$$f(x) \sim gp(m(x), k(x, x')) \quad (10)$$

式中: $m(x)$ 为均值函数, $m(x) = E[f(x)]$, $k(x, x')$ 为协方差函数, $k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))]$ 。

XGBoost 超参数优化问题中,建立超参数的样本数据集 $D = (X, y)$, 其中 $X = (x_1, x_2, \dots, x_t)$ 为训练集, $y = \{f(x_1), f(x_2), \dots, f(x_t)\}$ 为连续函数 $f(x)$ 的集合。则存在高斯分布可表示如下:

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_t) \end{bmatrix} \sim N \left(0, \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \dots & k(x_1, x_t) \\ k(x_1, x_1) & k(x_1, x_2) & \dots & k(x_1, x_t) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_t, x_1) & k(x_t, x_2) & \dots & k(x_t, x_t) \end{bmatrix} \right) \quad (11)$$

若加入新的样本 x_{t+1} 并更新协方差矩阵,记为 \mathbf{K} , 则联合高斯分布可表示为:

$$\begin{bmatrix} f_{1:t} \\ f_{t+1} \end{bmatrix} \sim N \left(0, \begin{bmatrix} \mathbf{K} & k \\ k^\top & k(x_{t+1}, x_{t+1}) \end{bmatrix} \right) \quad (12)$$

其中, $k = [k(x_{t+1}, x_1) \quad k(x_{t+1}, x_2) \quad \dots \quad k(x_{t+1}, x_t)]$, 进一步可得到 f_{t+1} 的后验概率分布:

$$P(f_{t+1} | D_{1:t}, x_{t+1}) \sim N(u, \sigma^2) \quad (13)$$

$$u = k^\top \mathbf{K}^{-1} f_{1:t} \quad (14)$$

$$\sigma^2 = k(x_{t+1}, x_{t+1}) - k^\top \mathbf{K}^{-1} k \quad (15)$$

2) 采集函数

采集函数在超参数的决策空间中指导选择下一个采样点。常见的主要有 3 种:PI(probability of improvement)、EI (expected improvement)、UCB (upper confidence bound)。文中使用 PI 作为采样函数,表达式为:

$$f_{PI}(x) = \Phi \left(\frac{u(x) - f(x^+) - \xi}{\sigma(x)} \right) \quad (16)$$

式中: $\Phi(\cdot)$ 为正态分布累积密度函数; $u(x)$ 、 $\sigma(x)$ 基于高斯过程所得,分别是目标数值的均值和方差; $f(x^+)$ 为当前最佳目标函数值; ξ 为参数。

2 构建血压预测模型

2.1 初步血压预测模型

来自数据集中同一时间段的 PPG 信号和 ECG 信号经过滤波、降噪、消除基线漂移等预处理^[18-19], 计算得到 PTT。在基于单一 PTT 的线性血压预测模型基础上,综合身高和体重计算身体质量指数(BMI)值,水银血压计收缩压(systolic blood pressure, SBP)和舒张压(diastolic blood pressure, DBP)标准测量值进行多元线性回归拟合,计算得到血压预测初始值。收缩压和舒张压的多元线性拟合公式下:

$$SBP = a_1 \times PTT + b_1 \times BMI + c_1 \quad (17)$$

$$DBP = a_2 \times SBP + b_2 \times BMI + c_2 \quad (18)$$

其中, PTT 为脉搏波传导时间, BMI 为身体质量指数, a_1, b_1, c_1 和 a_2, b_2, c_2 为常量。

2.2 BO-XGBoost 血压预测模型

基于 BO-XGBoost 模型进一步对血压预测值进行精确预测。将初步血压预测值和人体特征参数加入数据集,并对数据进行剔除异常值、缺失值、归一化等预处理,预处理后数据集 80% 作为训练集,训练血压预测模型,优化超参数,并对超参数优化方法进行性能比较。20% 作为测试集,对 XGBoost 血压预测模型效果进行全面评估。XGBoost 超参数多且复杂,人工调参具有较高难度,各参数的取值大小直接影响模型的预测精度。贝叶斯优化可实现超参数快速自动寻优,为便于对超参数优化效果进行实验对比,引入网格搜索(grid search, GS)和 10 折交叉验证^[20]相结合调节 XGBoost 超参数。基于贝叶斯优化的 XGBoost 模型血压预测流程如图 1 所示。

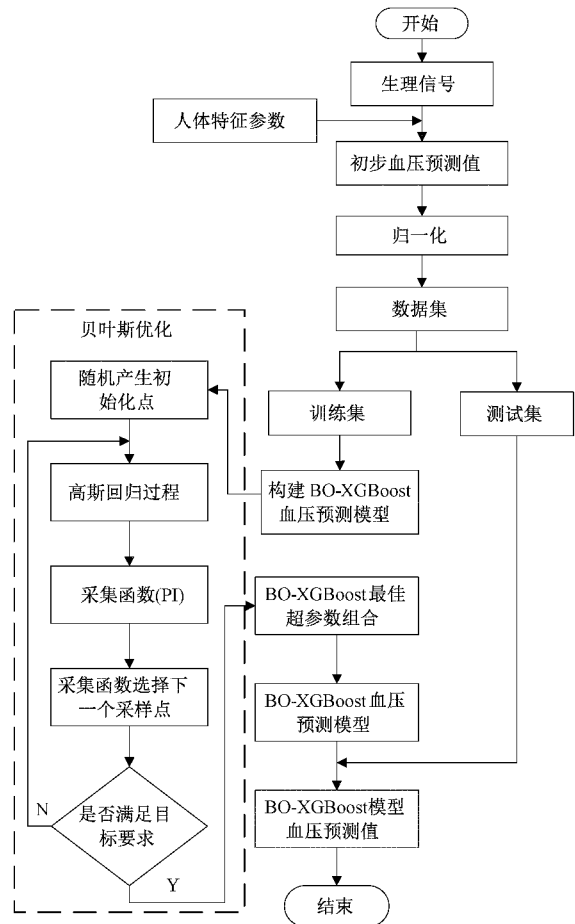


图 1 基于贝叶斯优化的 XGBoost 模型预测血压流程

1) 数据预处理

人体特征参数(年龄、身高、体重、性别、心率、BMI),以及初步血压预测值构成 BO-XGBoost 血压预测模型数据集,对数据集进行预处理,剔除异常值和缺失值。转换类别数据,将女性和男性分别记为 0 和 1。由于实验数据包含

多种人体特征参数,这些特征参数处于不同的量纲和数量级,极差较大,作为血压预测模型的输入会增大血压预测结果误差,因此对数据进行归一化处理。数据归一化计算公式如下:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (19)$$

其中, x 表示数据集中的一个特征参数, x_{\max} 为此特征参数最大值, x_{\min} 为此特征参数最小值。

2) 基于贝叶斯优化算法的超参数调优

XGBoost 算法的超参数主要归为 3 类:任务参数、常规参数、提升器参数。对于已确定的超参数组合,在相同超参数取值范围内,分别使用贝叶斯优化算法,网格搜索法结合 10 折交叉验证寻找最佳超参数组合。主要调节参数的含义及范围如表 1 所示。

表 1 XGBoost 超参数含义及其范围

XGBoost 参数	参数含义	参数范围
n_estimators	决策树数量	(200, 1 000)
max_depth	树的最大深度	(2, 16)
min_child_weight	叶子结点权重之和	(0, 10)
learning_rate	学习率	(0.01, 0.3)
subsample	子样本占训练样本集比例	(0.5, 1)
Colsample_bytree	特征随机采样比例	(0.5, 1)

XGBoost 算法超参数寻优范围确定后,贝叶斯优化超参数调优主要步骤如下:

(1) 构建 BO-XGBoost 血压预测模型,根据确定的待优化超参数,在参数范围内得到超参数的随机初始点 $X = [x_1, x_2, \dots, x_n]$, 输入实验数据,训练 BO-XGBoost 模型,目标函数响应值为 $Y = [f(x_1), f(x_2), \dots, f(x_n)]$ 建立高斯回归过程。

(2) 基于采集函数 PI 从高斯回归模型中选取下一个超参数组合采样点 x_{n+1} 。

(3) XGBoost 模型的超参数取值为新的超参数组合,训练模型,得到模型血压预测值。若预测值与标准值误差达到目标要求,输出最佳超参数组合。

(4) 若血压预测值与标准值的误差未达到目标要求,将 $(x_{n+1}, f(x_{n+1}))$ 加入高斯回归模型,重复执行步骤(2)和(3),找到符合目标要求的超参数组合,停止迭代。

网格搜索和 10 折交叉验证结合对 XGBoost 超参数调优主要步骤如下:

(1) 确定超参数取值范围,已优化参数取优化后取值,未优化参数取初始值。

(2) 使用网格搜索和 10 折交叉验证训练模型,获得最佳分数超参数取值。

(3) 根据步骤(2)所得参数值缩小参数取值范围和步长,重复步骤(2)得到最佳超参数值。

(4) 对参数 n_estimators、max_depth、min_child_weight、learning_rate、subsample、Colsample_bytree 重复步骤(1)~(3),得到参数组合值作为 XGBoost 血压预测模型超参数。

3) 超参数调优方法性能比较

通过贝叶斯优化和网格搜索分别对 XGBoost 算法超参数寻优,对应得到两组超参数组合。在数据集中随机选取 56 组测试数据,采用均方根误差(RMSE)和调优时间(t/s)作为评价指标,两种调优方法的对应血压预测模型 BO-XGBoost 和 GS-XGBoost 的收缩压(P_s)和舒张压(P_d)性能比较结果如表 2 所示。

RMSE 的计算公式为:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (20)$$

其中, m 表示样本数量, y_i 表示第 i 个样本的血压真实值, \hat{y}_i 表示第 i 个样本的血压预测值。

表 2 XGBoost 超参数调优性能对比

血压	预测模型	RMSE	t/s
P_s	BO-XGBoost	2.58	1.22
	GS-XGBoost	3.88	3.43
P_d	BO-XGBoost	2.83	1.54
	GS-XGBoost	4.15	3.92

由表 2 可知,贝叶斯超参数优化方法展现出更好的性能。网格搜索优化方法运行时间约是贝叶斯优化方法的 3 倍,并且其收缩压和舒张压的均方根误差都大于贝叶斯优化方法。因此,通过贝叶斯优化超参数的 XGBoost 血压预测模型具备更高精度和泛化能力。进一步分析两种超参数组合下收缩压和舒张压与真实值的误差,得到如图 2 所示的 GS-XGBoost 和 BO-XGBoost 血压预测结果误差对比图。GS-XGBoost 血压预测模型收缩压最大绝对误差为 8.45 mmHg,舒张压最大绝对误差为 7.28 mmHg,而 BO-XGBoost 血压预测模型收缩压最大绝对误差为 5.12 mmHg,舒张压最大绝对误差为 5.73 mmHg,说明贝叶斯优化 XGBoost 超参数能使血压预测模型性能误差更小,性能更优。

3 实验结果与分析

3.1 数据源

本文使用的数据来源于数据库 (<http://dataset.kangdollar.com/>),由本团队自主采集创建。每组数据信息由两部分组成,包括实时采集的连续 PPG 和 ECG 信号波形以及被测人员的人体特征参数:年龄、性别、身高、体重、心率以及标准收缩压与舒张压,其中标准收缩压与舒张压是在采集生理信号时用水银血压计测量获得。本文中的实验数据是从数据库中随机挑选的 1 000 组数据,为保证

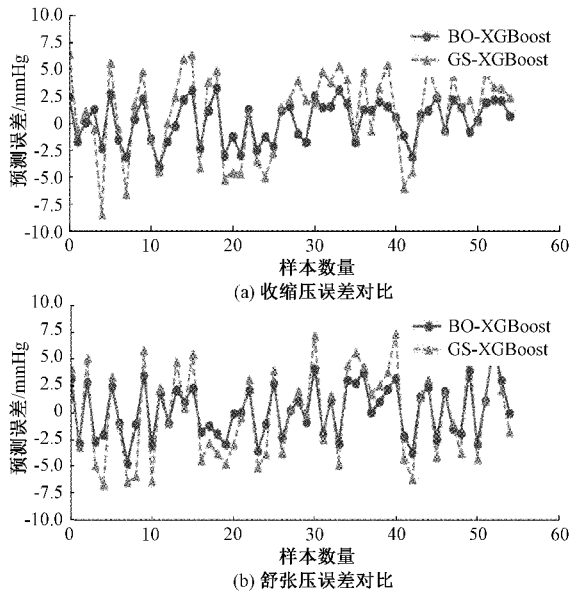


图 2 超参数调优预测结果误差对比图

数据的健壮性与完备性,收缩压的范围为 80~184 mmHg,舒张压分布范围为 45~100 mmHg,包含 16~80 岁年龄段人群,其中男性 44%,女性 56%。能为连续无创血压预测模型提供可靠的数据基础。具体的实验数据分布如图 3 所示。

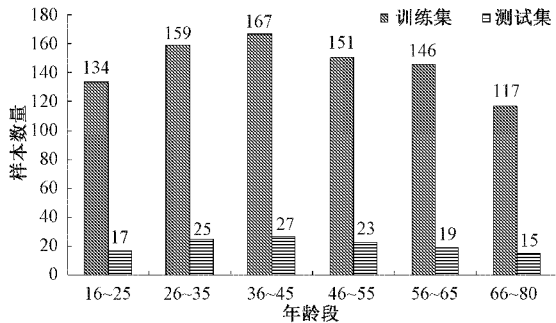


图 3 数据集年龄分布图

3.2 实验分析

为检验模型的预测效果,将 126 组数据导入训练好的 3 种血压预测模型,即贝叶斯优化的 XGBoost 模型(BO-XGBoost),网格搜索法优化的 XGBoost 模型(GS-XGBoost),支持向量机回归模型(SVR)^[21],将同一时间段水银血压计的观测值记为 True。保存 3 种血压预测模型结果并分析。BO-XGBoost 模型与其余两种方法的收缩压与舒张压预测结果对比如图 4 所示。

从图 4 的血压预测结果对比图可以清晰地看出,BO-XGBoost 回归模型血压预测方法经过优化超参数,对比 GS-XGBoost 血压预测效果上有更大提高。与支持向量机回归预测模型相比,BO-XGBoost 模型收缩压和舒张压的预测结果更接近于标准测量值 True,具有更好的拟合效果,收缩压的预测优势显著。

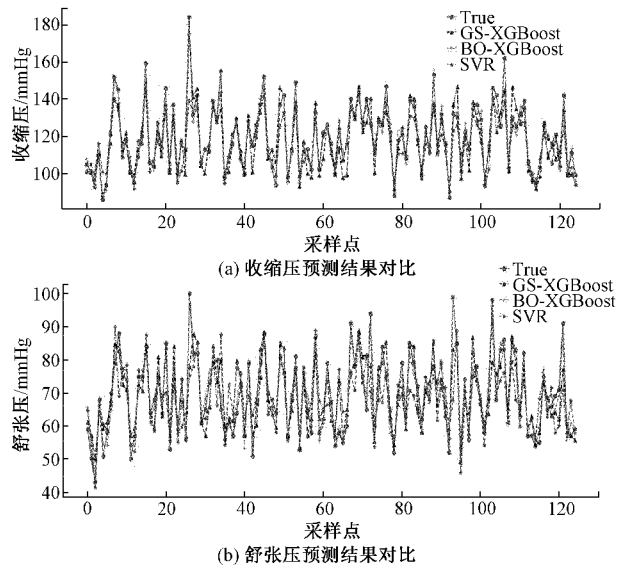


图 4 BO-XGBoost 与其他算法血压预测结果对比图

为了进一步验证血压预测结果,对 3 种模型预测结果进行误差分析,根据 AAMI 标准制定误差带大小为 ± 5 mmHg,得到误差分析图如图 5、6 所示。

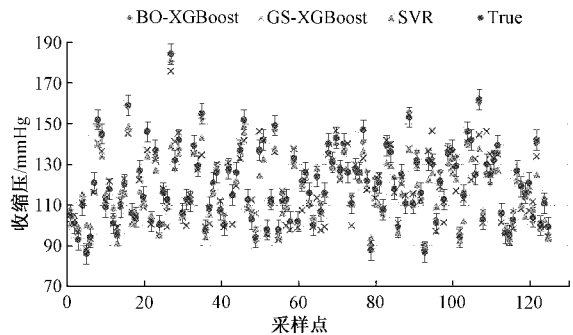


图 5 收缩压误差分析图

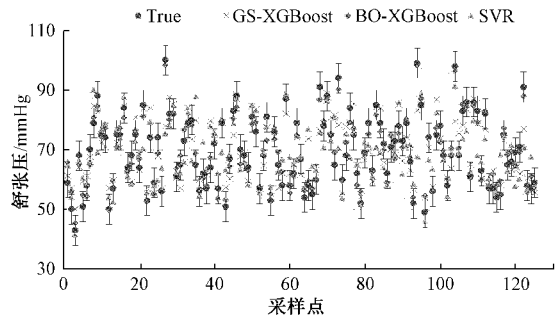


图 6 舒张压误差分析图

为了评估 BO-XGBoost 血压预测模型和其他血压预测算法的性能,本文针对收缩压(P_s)和舒张压(P_d)分别从平均绝对误差(MAE)、决定系数(R^2)、准确率(AR)进行模型评估。计算公式如下:

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (21)$$

$$R^2 = \frac{\sum_{i=1}^m (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (22)$$

$$AR = \frac{\text{符合误差允许的样本数}}{\text{总测试样本数}} \times 100\% \quad (23)$$

其中, y_i 为第 i 个样本的血压真实值, \hat{y}_i 为第 i 个样本的血压预测值, \bar{y} 为 m 个样本的血压标准值的平均值。AR 值越大表明模型预测结果越准确, 在 AR 计算过程中符合误差允许的样本数由图 5 和 6 所得。MAE 值越大则模型的准确度越低。 R^2 用来衡量模型的拟合程度, R^2 值越接近于 1, 则模型的拟合程度越好。通过以上 3 个评价指标对实验结果进行分析, 得到如表 3 所示的结果对比信息。

表 3 模型各项评价指标结果对比

血压	预测模型	MAE	R^2	AR/%
P_s	BO-XGBoost	2.45	0.974 8	96
	GS-XGBoost	5.89	0.823 9	65
	SVR	5.11	0.822 5	61
P_d	BO-XGBoost	2.58	0.937 8	92
	GS-XGBoost	5.02	0.704 2	64
	SVR	6.58	0.471 7	51

由表 3 可知, BO-XGBoost 模型收缩压和舒张压的 MAE 均满足 AAMI 制定的小于 ± 5 mmHg 的要求, 且明显小于支持向量机回归模型和 GS-XGBoost 模型的 MAE, 其 R^2 值分别为 0.974 8 和 0.937 8, 更接近于 1。相比 GS-XGBoost 回归血压预测模型, BO-XGBoost 模型对收缩压的预测准确度提高了 31%, 舒张压提高了 28%。并且 SVR 血压预测模型预测收缩压运行时间为 4.96 s, 舒张压预测时间为 5.28 s, 大于表 2 中 GS-XGBoost 和 BO-XGBoost 血压预测模型运行时间。表明基于贝叶斯优化的极端梯度提升血压预测模型泛化性能好, 预测准确度高, 更具可靠性。

进一步检验 BO-XGBoost 模型预测血压的可行性, 运用 Bland-Altman 分析法对 BO-XGBoost 血压预测值和水银计测量的标准血压值进行一致性分析检验, 得到如图 7、8 所示的一致性分析图。图中 SD 是两种检测方法得到的血压差值均值的标准差, Mean 为血压差值的均值。

由图 7、8 可以看出, 通过对 BO-XGBoost 模型血压预测值和水银计血压测量值进行一致性检验, 这两种方法的差值 96% 位于 $\text{Mean} \pm 1.96 \times \text{SD}$ 一致性界限之间, 收缩压的 Mean 值为 -0.3 , 舒张压的 Mean 值为 -0.1 , 接近于 0, 则基于贝叶斯优化的 XGBoost 的连续无创血压测量方法与水银血压计测量法具有很好的一致性, 同时说明本文提出的 BO-XGBoost 模型用于连续无创血压测量是可行的。

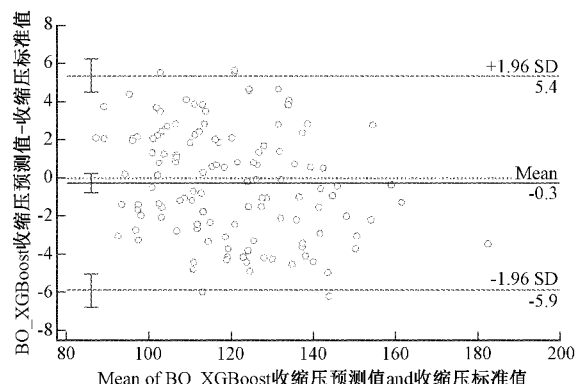


图 7 BO-XGBoost 预测收缩压 Bland-Altman 分析图

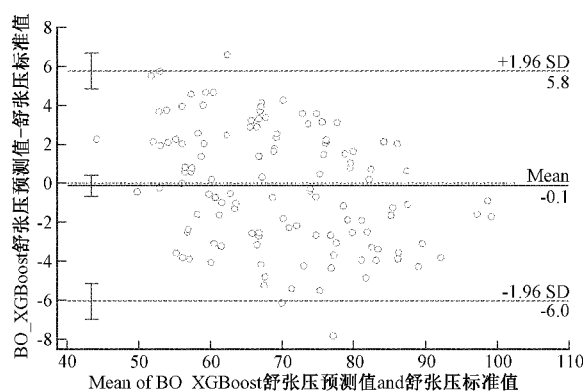


图 8 BO-XGBoost 预测舒张压 Bland-Altman 分析图

4 结 论

本文提出一种基于贝叶斯优化 XGBoost 的无创血压检测方法。针对实时采集的连续 PPG 和 ECG 信号进行预处理, 提取优质波形特征点计算 PTT, 结合人体特征参数年龄、身高、体重等建立血压模型输入参数。并利用贝叶斯优化算法确定 XGBoost 血压预测模型超参数最佳组合, 对比 GS-XGBoost 和 SVR 模型。通过 MAE、 R^2 、AR 和运行时间综合评估 3 种模型的血压预测效果, Bland-Altman 一致性分析结果表明, BO-XGBoost 具备良好的泛化能力和准确性, 克服了连续无创血压检测误差大的缺点, 有助于连续无创血压测量方法的改进和广泛应用。

参考文献

- [1] 赵冬. 中国成人高血压流行病学现状[J]. 中国心血管杂志, 2020, 25(6): 513-515.
- [2] KOSHIMIZU H, KOJIMA R, KARIO K, et al. Prediction of blood pressure variability using deep neural networks[J]. International Journal of Medical Informatics, 2020, 136(C): 104067.
- [3] REBESCO M R, PINKSTON M C, SMYRNIOS N A, et al. A Comparison of non-invasive blood pressure measurement strategies with intra-arterial

- measurement[J]. *Prehospital and Disaster Medicine*, 2020, 35(5): 516-523.
- [4] 吴金桐, 桑海峰, 徐州, 等. 实时血压监测系统设计[J]. *电子测量技术*, 2019, 42(24): 63-70.
- [5] GHOSH S, BANERJEE A, RAY N, et al. Continuous blood pressure prediction from pulse transit time using ECG and PPG signals [C]. *Healthcare Innovation Point-of-care Technologies Conference, IEEE*, 2016, DOI: 10.1109/HIC.2016.7797728.
- [6] 张大可, 刘立红, 王钢, 等. 基于心电信号-光电容积脉搏波信号多特征参数无创血压监测系统设计[J]. *生物医学工程研究*, 2018, 37(4): 447-453.
- [7] EL-HAJJ C, KYRIACOU P A. A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure [J]. *Biomedical Signal Processing and Control*, 2020, 58(C): 101870-101870.
- [8] 淳新益, 郑秀娟, 张畅, 等. 遗传算法优化 BP 神经网络的非接触式血压估计方法[J]. *电子测量与仪器学报*, 2021, 35(7): 53-59.
- [9] KHALID S G, ZHANG J, CHEN F, et al. Blood pressure estimation using photoplethysmography only: Comparison between different machine learning approaches [J]. *Journal of Healthcare Engineering*, 2018, 2018: 1548647.
- [10] 李嘉莹, 赵丽, 边琰, 等. 基于 LDA 和 KNN 的下肢运动想象脑电信号分类研究[J]. *国外电子测量技术*, 2021, 40(1): 9-14.
- [11] 朱海龙, 袁贞明, 俞凯. 基于 ARIMA 与 SVM 混合模型的连续血压预测[J]. *杭州师范大学学报(自然科学版)*, 2018, 17(5): 555-560.
- [12] 师荣堃, MOHAMMED H, 林炳辉, 等. 特征自提取的深度学习血压预测模型研究[J]. *传感器与微系统*, 2020, 39(4): 16-19.
- [13] CHEN T, GUESTRIN C. XGBoost: A scalable tree boosting system [J]. *ACM*, 2016, DOI: 10.1145/2939672.2939785.
- [14] 曲文龙, 李一漪, 周磊. XGBoost 算法在糖尿病血糖预测中的应用[J]. *吉林师范大学学报(自然科学版)*, 2019, 40(4): 118-125.
- [15] 周盛山, 汤占军, 王金轩, 等. EEMD 和 CNN-XGBoost 在风电功率短期预测的应用研究[J]. *电子测量技术*, 2020, 43(22): 55-61.
- [16] 崔佳旭, 杨博. 贝叶斯优化方法和应用综述[J]. *软件学报*, 2018, 29(10): 3068-3090.
- [17] 龚雪娇, 朱瑞金, 唐波. 基于贝叶斯优化 XGBoost 的短期峰值负荷预测[J]. *电力工程技术*, 2020, 39(6): 76-81.
- [18] SUN B, WANG C, CHEN X, et al. PPG signal motion artifacts correction algorithm based on feature estimation[J]. *Optik International Journal for Light & Electron Optics*, 2019, 176: 337-349.
- [19] 孙斌, 王成超, 陈建飞, 等. 受运动伪影干扰 PPG 序列的优质信号提取算法[J]. *仪器仪表学报*, 2018, 39(9): 171-178.
- [20] 王燕, 郭元凯. 改进的 XGBoost 模型在股票预测中的应用[J]. *计算机工程与应用*, 2019, 55(20): 202-207.
- [21] 赵谓博, 赫英迪, 李信政, 等. 基于支持向量回归的人体血压预测方法[J]. *燕山大学学报*, 2017, 41(5): 438-443.

作者简介

孙斌, 博士, 副教授, 主要研究方向为生物医学信号处理、多光谱图像融合、智能测控与仪器。

E-mail: ffgz366@163.com

储芳芳(通信作者), 硕士研究生, 主要研究方向为网络化测控与网络测试技术。

E-mail: 876829309@qq.com

陈小惠, 博士, 教授, 主要研究方向为网络化测控系统、嵌入式系统与智能仪器、传感器网络与信息融合。

E-mail: chenxiaohui@njupt.edu.cn