

DOI:10.11784/tdxbz202407008

基于机器学习的人体热舒适度建模与预测

邓 斌, 龚 安, 王 江

(天津大学电气自动化与信息工程学院, 天津 300072)

摘要: 热舒适度是衡量室内环境质量和影响人类健康的重要指标之一, 是建筑、空调控制等系统智能化的重要参考依据, 同时能够有效降低建筑热环境控制的能源需求。目前可穿戴设备如智能手环、柔性传感器等已广泛应用, 可构建人体的健康大数据。但由于存在个体差异因素, 不同个体对相同热环境所表现的生理热反应不同, 基于单个人的热舒适模型难以对群体热状态实现有效地预测。考虑到以往研究样本量相对较小、模型复杂难以部署等局限性, 本文建立人工气候室, 利用环境传感器和可穿戴设备收集了 60 名受试者的热舒适数据, 采用机器学习实现人体热舒适度建模与预测。研究考虑身高、体重、性别等个体差异因素, 采用 XGBoost、随机森林和 SVC 共 3 种机器学习算法, 得到了基于人体生理参数的增强型预测热态模型并对热舒适度进行分类。结果表明: 对皮肤温度及其梯度进行归一化处理发现, 归一化过程能够将冷不舒适、舒适、热不舒适 3 种状态拉开, 有利于 SVC 算法在高维空间寻找最优超平面, 对特征进行分类。对比归一化前后随机森林模型的特征重要性发现, 归一化过程降低了体重、身高、性别等个体差异对模型预测效果的影响程度; 在 XGBoost、随机森林和 SVC 这 3 种机器学习算法中, SVC 在测试集上的准确率和 3 种热状态的 AUC 值都高于 XGBoost 和随机森林, 其分类效果和泛化能力最好。

关键词: 可穿戴; 热舒适度; 个体差异; 机器学习; 预测模型

中图分类号: TP274.2

文献标志码: A

文章编号: 0493-2137(2025)03-0237-10

Modeling and Prediction of Human Thermal Comfort Based on Machine Learning

Deng Bin, Gong An, Wang Jiang

(School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China)

Abstract: Thermal comfort is an important indicator of indoor environment quality and affects human health. It is an important reference for intelligent building systems, air conditioning control, and other systems. Moreover, it can effectively reduce the energy demand for controlling the thermal environment in buildings. Currently, wearable devices such as smart watches and flexible sensors are extensively used to compile comprehensive data on human health. However, due to individual differences, physiological thermal responses to identical thermal conditions vary, and it is difficult to effectively predict the group thermal state for a personal thermal comfort model. Considering the limitations of the relatively small sample sizes and complex model deployments in previous studies, this work established an artificial climate chamber with environmental sensors and wearable devices to collect thermal comfort data of 60 subjects and leveraged machine learning to realize human thermal comfort modeling and prediction. Considering individual differences such as height, weight, and gender, three machine-learning algorithms, i.e., extreme gradient boosting (XGBoost), random forest, and support vector classifier (SVC), were used to obtain an enhanced predictive thermal state model based on human physiological parameters and to classify thermal comfort. The results showed that the skin-temperature normalization process and its gradient result in three states (cold discomfort, comfort, and thermal discomfort). This facilitates the SVC algorithm to find the optimal hyperplane in high-dimensional space and

收稿日期: 2024-07-16; 修回日期: 2024-09-02

作者简介: 邓 斌 (1979—), 男, 博士, 教授。

通信作者: 邓 斌, dengbin@tju.edu.cn.

基金项目: 国家自然科学基金资助项目(62171311).

Supported by the National Natural Science Foundation of China (No. 62171311).

classify the features. Comparative analysis of the feature importance of the random forest model before and after skin-temperature normalization revealed that normalization reduces the influence of individual differences such as weight, height, and gender on the predictive effect of the model. Of the three machine learning algorithms, the accuracy of SVC on the test set and the area under the curve (AUC) values of the three thermal states were higher than those of XGBoost and random forest. Hence, SVC has the best classification effect and generalization capability.

Keywords: wearable; thermal comfort; individual difference; machine learning; prediction model

热舒适影响居住者的健康、生产力和幸福感^[1-2]。为了最大限度地提高舒适度,节省环境调节能源,同时了解居住者热舒适性的更多细微方面,尤其是个人需求,提出以居住者为中心的热舒适性预测模型^[3]和干扰解^[4]对改善传统一刀切控制策略^[5-6]尤为重要。

近些年,在人体热舒适研究领域,许多研究者结合生理学来探究热舒适的机理。从生理学的角度出发,探讨热舒适机理是新的思路^[7]。在不同的热环境中,人体会通过一系列生理调节活动来适应热环境。与此同时,人体热反应生理参数,如皮肤温度、脉搏率等,会随之变化。其中,皮肤温度、脉搏率是反映人体与环境热交换状况的重要参数,与人体热舒适性密切相关。Fanger^[5]通过研究发现,在微寒、微冷区及舒适区的环境下,皮肤温度是反映人体热舒适感受的重要指标;Choi 等^[8]的研究表明,心率/脉搏率及其变化在温暖和凉爽条件下都显示出显著的差异。因此,研究人体皮肤温度、脉搏速率与热舒适的关系,对热舒适机理研究具有十分重要的意义。

热舒适中个体差异的研究对于热环境的精准调控至关重要^[9]。在以往的研究中,PMV-PPD 模型^[5]和自适应热舒适模型^[10-11]通常用于预测居住者在各种室内环境中的热感知,但是,模型中没有充分考虑个体差异。因此,这些模型需要进一步改进,以便在室内环境中创造良好的热舒适性,并满足居住者对热舒适性日益增长的需求。目前,热舒适中个体差异的研究有两个方面:人与人之间舒适反应的差异和个体在不同场合同一环境中的感受^[9]。一些研究考虑了在不同热环境下,作用于生理反应的性别^[12]、身高、体重等^[13]个体因素,但并未提出解决或改善方案。

目前,商业化的可穿戴设备能以低成本准确地收集实时心率/脉搏率、皮肤温度等生理数据,因此结合新的传感技术(可穿戴设备等)和数据驱动(机器学习等)的方法,利用环境、生理等方面的多种测量数据,可以实现更加精确的热舒适预测^[14],这为改善传统的一刀切控制策略提供了基础。文献[15]是通过可穿戴传感器和机器学习分别检测和预测个体差异,采取措施(如个人舒适系统)提高空调系统的灵活性,使其满

足不同用户的不同需求。本研究在检测个体差异之后,对其进行分析和处理,以降低其对模型的不良影响。

为了解决以前室内热舒适研究的局限性,如样本量相对较小、模型未充分考虑个体差异、模型复杂难以部署等,本研究从热感觉和热舒适性的角度进行热环境的评价,利用人体热反应生理参数(皮肤温度、脉搏率)和环境参数(环境温度、环境湿度),结合机器学习的方法建立热舒适度预测模型。本研究基于身高、体重、性别等个体差异因素建立的增强型预测热态模型能够利用可穿戴设备的便携性优点,实现更加精确的热舒适预测,并结合室内空调系统控制的特点,对室内热环境进行恰当的调控。

1 人体热舒适实验

1.1 热舒适参数选取

1) 环境温度和湿度

环境温度和湿度是影响人体热感觉和热舒适性最直接的环境因素,同时也是方便测量的环境变量。

2) 皮肤温度

皮肤温度通过其传热特性,在控制热舒适性的热调节原理中起着重要作用。热调节原理通过平衡人体与环境之间的热通量来调节热舒适性^[16]。人体通过血管收缩和血管舒张的过程来调节热量,其中皮肤血液流动被用作在身体核心和周围之间传递热量的媒介。在温暖的环境下,人体通过增加从核心到外围的血液流动来排出热量;这种现象被称为皮肤血管舒张,增加了皮肤的血流量,因此皮肤温度升高。在低温条件下,从核心到外周的血流量减少,以保持核心温度恒定;这种现象被称为皮肤血管收缩,减少皮肤血液流动,导致皮肤温度降低。这个温度调节过程表明皮肤温度具有预测热状态的潜力。

四肢皮肤与环境直接接触,反映人体表层温度,而胸口皮肤温度能够反映人体核心温度^[17]。表层温度与核心温度的结合具有更好的潜力来预测人体热舒适状态。为了模型的可靠性及个体特异性,本实验对手腕、胸口、小臂、小腿等多个部位的皮肤温度进行

采集。

3) 脉搏率

脉搏与热舒适有密切关系,本文研究了与心率类似的脉搏率.与测量心率的侵入式脑电图方法不同,它可以从手指感知,因此可穿戴设备更容易实现测量,并用于实际.脉搏率与心血管系统密切相关.在Choi等^[8]的研究中,心率及其变化在温暖和凉爽条件下都显示出显著的差异,可以使用机器学习算法来评估热舒适。

4) 性别

男性和女性在生理特征和生理上的先天差异会导致他们在热反映上的差异,因此本文研究潜在的生理差异。

性别是一个常见的个人因素,经常被认为与热感知和热舒适需求有关.Liu等^[18]研究结果表明,女性受试者的热感觉和平均皮肤温度普遍低于男性.然而,这些差异在统计学上并不显著.Lan等^[19]研究发现,根据心率变异性分析显示,女性在生理上对凉爽环境更敏感,而男性对温暖环境更敏感。

1.2 实验场地和实验设备

在本实验中,主要用到的仪器、设备有人工气候室、环境传感器和人体体征传感器.人工气候室位于天津大学某教室,大小为 $5.25\text{ m} \times 9.52\text{ m} \times 2.72\text{ m}$,良好契合普通卧室大小.采用隔热棉封窗,室内升温快,保温效果良好.实验场地如图1(a)所示.环境温度和湿度采用建大仁科环境温湿度记录仪记录,范围分别为 $-40 \sim 80\text{ }^{\circ}\text{C}$ 和 $0 \sim 100\%$,精度分别为 $\pm 0.5\text{ }^{\circ}\text{C}$ 和 $\pm 3\%$.环境传感器如图1(b)所示.采用iButton DS1922L温度记录仪记录皮肤温度.采集时间间隔设置为 1 min .体温传感器如图1(c)所示.采用力康Prince-100H型号血氧仪检测被试者的脉搏率和血氧饱和度,血氧仪如图1(d)所示.传感器的参数如表1所示。



(a) 人工气候室



(b) 环境传感器



(c) 体温传感器



(d) 血氧仪

图1 实验场地和设备

Fig.1 Experimental site and equipment

表1 传感器参数

Tab.1 Parameters of sensors

采集参数	传感器名称	测量范围	精度
环境温度	建大仁科 RS-MG101-1	$-40 \sim 80\text{ }^{\circ}\text{C}$	$\pm 0.5\text{ }^{\circ}\text{C}$
环境湿度	建大仁科 RS-MG101-1	$0 \sim 100\%$	$\pm 3\%$
皮肤温度	MAXIM iButton DS1922L	$-40 \sim 85\text{ }^{\circ}\text{C}$	$\pm 0.5\text{ }^{\circ}\text{C}$
脉搏率	Heal Force Prince-100H	$30 \sim 240\text{ 次/min}$	$\pm 2\%$
血氧饱和度	Heal Force Prince-100H	$30\% \sim 100\%$	$\pm 2\%$

1.3 实验方案

实验受试者是60名(男33名,女27名)年龄在19~24岁之间的学生,实验前12h要求受试者保证充足的睡眠和稳定的情绪以及避免饮酒、吸烟和剧烈

体育活动。

受试者的基本信息汇总见表2,表中将各参数的测量结果以“平均值 \pm 标准差”的形式表示.从表2可以看出,在身高、体重和体表面积上,性别差异极

显著(显著性差异 P 值均小于 0.001),在年龄方面,性别差异不显著(显著性差异 P 值等于 0.375)。

表 2 受试者的基本信息

Tab.2 Basic information of subjects

性别	数量	身高/cm	体重/kg	年龄/岁	体表面积/m ²
男	33	176.3 ± 4.2	72.0 ± 12.1	21.9 ± 1.7	1.88 ± 0.15
女	27	163.1 ± 4.5	54.0 ± 6.0	21.5 ± 1.3	1.58 ± 0.09

受试者被要求提前 30 min 到达实验室,并在隔壁温暖舒适的准备室静坐 20 min 以上,以尽量减少之前热状态的影响并佩戴相关测量设备,同时气候室中空调初始温度设置为 18 °C 或 19 °C. 实验工况如表 3 所示。

表 3 实验工况

Tab.3 Experimental conditions

工况	温度范围/°C	温度步长/°C	时间步长/min
1	18 ~ 27	3	30
2	19 ~ 28	3	30

实验过程中受试者保持静止并且情绪稳定,每隔 5 min 填写一次调查问卷,选择相应的指标来评价自己当前的热感觉(thermal sensation vote, TSV)和舒适度. 每隔 30 min 调高空调温度 3 °C. 每次实验时长为 2 h. 受试者的服装热阻是根据 ISO 9920—2007 确定的. 夏季受试者均按照要求统一着装(深色 T 恤,长裤子,休闲鞋),考虑椅子的热阻后,服装热阻为 0.57 Clo.

调查问卷如表 4 所示. 将热状态分类为:冷不舒适(热感觉为-3, -2),舒适(热感觉为-1, 0, 1),热不舒适(热感觉为 2, 3). 使用热状态指数(thermal state index, TSI)^[20]来量化这 3 种热状态: TSI = -1 代表冷不舒适; TSI = 0 代表舒适; TSI = 1 代表热不舒适。

表 4 调查问卷的标尺

Tab.4 Scale of the questionnaire

TSV	热感觉
-3	冷
-2	凉
-1	稍凉
0	中性
1	稍暖
2	暖
3	热

2 实验结果及讨论

2.1 热舒适参数

2.1.1 热舒适参数分布

实验测量的环境温湿度、皮肤温度、血氧饱和

度、脉搏率等热舒适参数统计结果见表 5. 从表 5 可以得出,手腕、胸口、小臂、小腿处皮肤温度的均值依次为 29.5 °C、33.4 °C、30.7 °C、30.4 °C,手腕皮肤温度最低,胸口皮肤温度最高. 另外,血氧饱和度范围和均值分别为 82% ~ 100% 和 97.2%,变化幅度小. 脉搏率范围和均值分别为 43 ~ 111 次/min 和 70.1 次/min,变化幅度大。

为进一步研究环境温湿度、皮肤温度、血氧饱和度、脉搏率等热舒适参数与热感觉的关系,统计了不同热状态下,各热舒适参数的分布情况,结果如图 2 所示。

表 5 热舒适参数统计结果

Tab.5 Statistical results of thermal comfort parameters

热舒适参数	符号	范围	均值	标准差
环境温度/°C	Te	16.7 ~ 29.6	23.5	2.7
环境湿度/%	RH	16.4 ~ 56.4	34.4	8.3
手腕温度/°C	ST1	22.7 ~ 35.0	29.5	2.4
胸口温度/°C	ST2	27.9 ~ 36.2	33.4	1.2
小臂温度/°C	ST3	21.2 ~ 35.8	30.7	2.1
小腿温度/°C	ST4	24.2 ~ 35.2	30.4	1.4
血氧饱和度/%	SP	82 ~ 100	97.2	1.8
脉搏率/(次/min)	PR	43 ~ 111	70.1	9.9

在冷(-3)和热(3)两种热状态下,受试者的脉搏率均值在 75.0 次/min 以上,其他热状态下,受试者的脉搏率均值在 69 ~ 71 次/min,冷、热两种热状态下受试者的脉搏率会有明显提升;在凉(-2)到暖(2)的过程中,脉搏率均值变化幅度不断降低. 可见在极端热环境下,人群的脉搏率变化剧烈,而在较为温和的热环境下,人群的脉搏率变化较为平缓。

2.1.2 热舒适参数与热感觉相关性分析

采用 Spearman 秩相关系数来研究每个受试者的热舒适参数与热感觉之间的相关性,相关性结果见图 3 和表 6,图 3 中展示了每个受试者各热参数与热感觉的相关性的分布情况,表 6 中给出了对应的相关系数的均值和标准差。

从表 6 可以看出,在所有的热舒适参数当中,环境温度与热感觉的相关性均值最高,为 0.72;环境湿度与热感觉的相关系数标准差最高,为 0.52. 可见在不同热感觉下,环境温度具有预测人群中个体热感觉的巨大潜力,环境湿度对热感觉的影响个体差异较大。

手腕、胸口、小臂、小腿处皮肤温度与热感觉的相关系数均值依次是 0.52、0.65、0.56、0.10. 其中胸口、小臂、手腕处皮肤温度与热感觉的相关性较高,大多数个体在 0.50 以上,这表明随着热感觉从冷到热的变化,皮肤温度单调地增加. 脉搏率与热感觉的

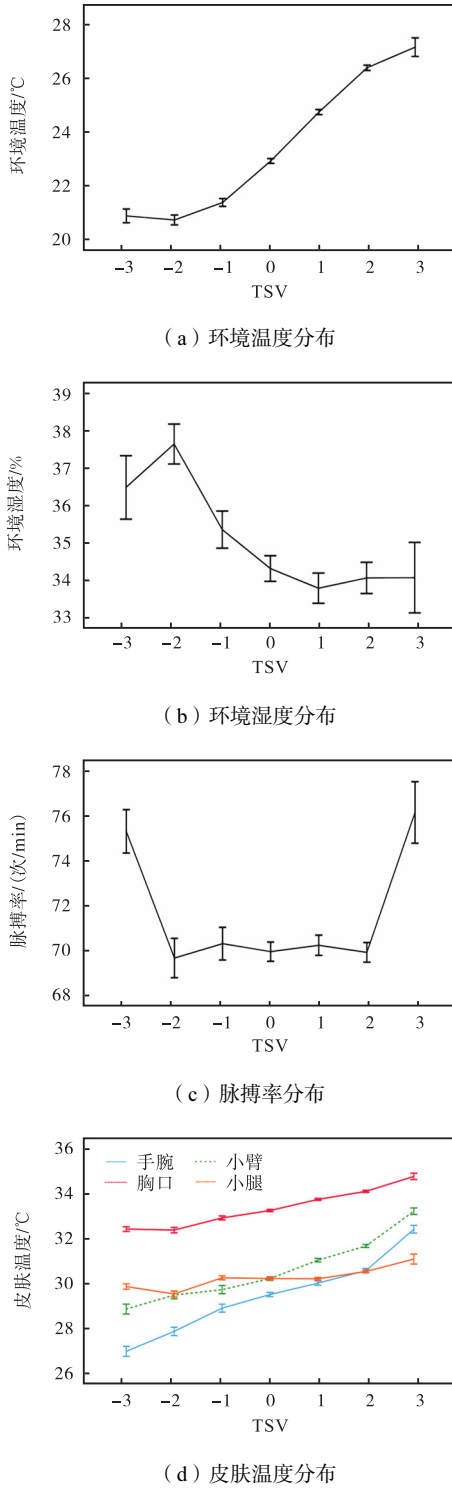


图2 不同热感觉(TSV)下热参数分布情况

Fig.2 Distribution of thermal parameters under different thermal sensation votes

相关系数均值为-0.10,多数受试者为负相关,少数为正相关.可见脉搏率具有预测热状态的潜力,但是其潜力低于环境温度和皮肤温度等热舒适参数.

值得注意的是:血氧饱和度与热感觉相关性均值极低,仅为-0.01,因此本文未将血氧饱和度纳入对热舒适的影响研究.

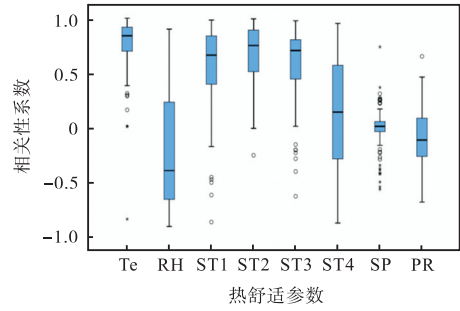


图3 热舒适与热感觉的相关性

Fig.3 Correlation between thermal comfort parameters and thermal sensation votes

表6 Spearman秩相关系数统计结果

Tab.6 Statistical results of Spearman rank correlation coefficient

热舒适参数	符号	相关系数均值	相关系数标准差
环境温度	Te	0.72	0.26
环境湿度	RH	-0.24	0.52
手腕温度	ST1	0.52	0.37
胸口温度	ST2	0.65	0.27
小臂温度	ST3	0.56	0.32
小腿温度	ST4	0.10	0.49
血氧饱和度	SP	-0.01	0.19
脉搏率	PR	-0.10	0.25

2.2 基于个体差异的生理参数归一化热舒适模型

2.2.1 生理参数归一化

不同的服装热阻和体型(身高、体重)组合会导致不同的暴露体表面积,这直接影响到身体的总热量散失,而总热量散失最终决定了整体的热感觉.一个人的体表面积是恒定的,可以解决个体间的差异,而衣服的保温程度因人而异,可以解决个体内部的差异.为了纳入这个简单的现象,引入一个归一化因子 A_{exp} ,具体公式为

$$A_{exp} = (1 - I_{cl}) A_{Du} \tag{1}$$

$$A_{Du} = W^{0.425} \times H^{0.725} \times 0.203 \tag{2}$$

式中: I_{cl} 为服装热阻; A_{Du} 为受试者的体表面积^[21]; W 为体重; H 为身高.

皮肤温度和脉搏率的梯度携带了重要的信息,可以指示居住者如何在给定的热环境中感知自己的热感觉.

因此本文将这一衍生特征纳入研究的分析中.取近 10 min 的皮肤温度和脉搏率变化分别作为皮肤温度梯度和脉搏率梯度,具体计算公式为

$$T_s^g(t) = T_s(t) - T_s(t-10) \tag{3}$$

$$R_p^g(t) = R_p(t) - R_p(t-10) \tag{4}$$

式中: $T_s(t)$ 为当前时刻的皮肤温度; $T_s(t-10)$ 为过去 10 min 的皮肤温度; $T_s^g(t)$ 为当前时刻的皮肤温度梯

度; $R_p(t)$ 为当前时刻的脉搏率; $R_p(t-10)$ 为过去 10 min 的脉搏率; $R_p^g(t)$ 为当前时刻的脉搏率梯度。

对皮肤温度及其梯度进行归一化, 得到归一化皮肤温度和归一化皮肤温度梯度, 具体计算过程为

$$\bar{T}_s = \frac{T_s}{A_{exp}} \quad (5)$$

$$\bar{T}_s^g = \frac{T_s^g}{A_{exp}} \quad (6)$$

式中: T_s 为皮肤温度; \bar{T}_s 为归一化后的皮肤温度; T_s^g 为皮肤温度梯度; \bar{T}_s^g 为归一化后的皮肤温度梯度。

借鉴皮肤温度归一化因子引入脉搏率归一化因子, 得到归一化脉搏率和归一化脉搏率梯度, 具体计算过程为

$$\bar{R}_p = \frac{R_p}{A_{exp} I_{sex}} \quad (7)$$

$$\bar{R}_p^g = \frac{R_p^g}{A_{exp} I_{sex}} \quad (8)$$

式中: R_p 为脉搏率; I_{sex} 为性别因子(男性 $I_{sex} = 1$, 女性 $I_{sex} = 2.4$); \bar{R}_p 为归一化后的脉搏率; R_p^g 为脉搏率梯度; \bar{R}_p^g 为归一化后的脉搏率梯度。

2.2.2 生理参数归一化效果分析

图 4 和图 5 展示了关于皮肤温度(胸口)及其梯度成对组合的热状态的详细分类。在归一化之前, 有

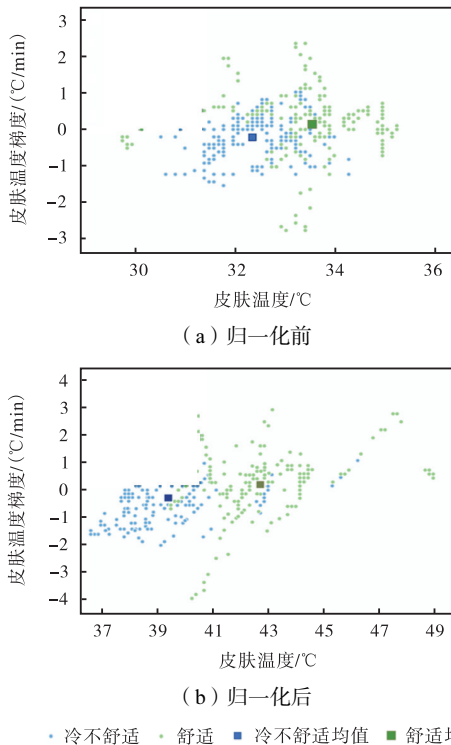


图 4 冷不舒适和舒适两种热状态下归一化前后对比
Fig.4 Comparison before and after normalization in two thermal states of cold discomfort and comfort

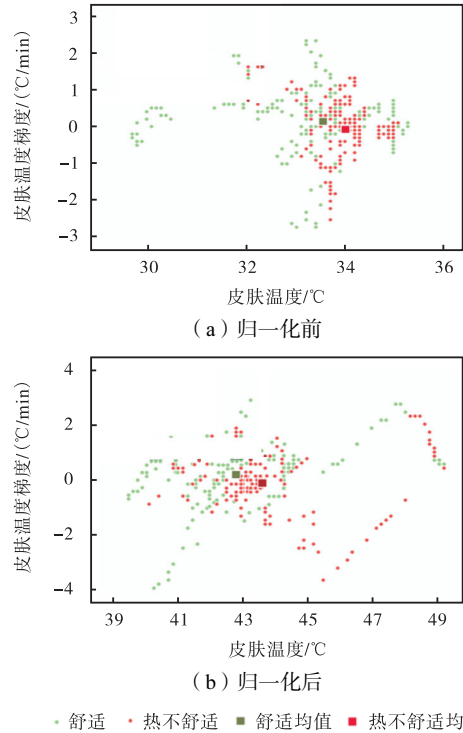


图 5 舒适和热不舒适两种热状态下归一化前后对比
Fig.5 Comparison before and after normalization in two thermal states of comfort and hot discomfort

较多重叠的点, 区分不清楚(图 4(a)和图 5(a)). 然而, 如图 4(b)和图 5(b)所示, 归一化过程成功地将冷不舒适、舒适、热不舒适 3 种状态拉开, 生成更清晰的非线性区域将彼此分开. 虽然在线性尺度上划分可能在视觉上不明显, 但与非归一化情况相比, 归一化过程显著提高了区别。

在不同输入参数下, 利用 SVC 对热状态进行分类预测, 效果如表 7 所示。

由表 7 可知, 皮肤温度/皮肤温度梯度、脉搏率/脉搏率梯度在归一化后作为输入参数, 准确率、召回率、F1 分数都有提高. 引入梯度之后, 准确率、召回率、F1 分数都有提高. 因此, 基于个体差异的归一化过程有助于提高模型的预测能力。

表 7 不同输入参数下模型热状态预测效果

Tab.7 Thermal state prediction effect of the model under different input parameters

输入参数	准确率/%	精度/%	召回率/%	F1 分数/%
$T_s + R_p$	68.19	53.15	35.16	30.49
$\bar{T}_s + \bar{R}_p$	69.11	43.46	39.15	38.00
$\bar{T}_s + \bar{T}_s^g + \bar{R}_p + \bar{R}_p^g$	69.20	42.84	41.00	40.38

2.2.3 模型架构

改进的增强型预测热态 (enhanced predictive thermal state, ePTS) 模型的解析图如图 6 所示。

首先, 使用可穿戴传感器记录环境温度 (T_e) 和环

境湿度(RH),同时记录皮肤温度和脉搏率.其次,分别从皮肤温度和脉搏率得到皮肤温度梯度和脉搏率梯度.

再次,根据服装热阻(I_{cl})、体表面积(A_{Du})、性别因子(I_{sex})等个体因素对这4个生理特征进行归一化处理.

该特征工程包括归一化过程,以从基本的生理数据中提取有意义的信息.生成以下输入特征:皮肤温度(T_s);归一化皮肤温度(\bar{T}_s);归一化皮肤温度梯度(\bar{T}_s^g);脉搏率(R_p);归一化脉搏率(\bar{R}_p);归一化脉搏率梯度(\bar{R}_p^g);环境温度(T_e);环境湿度(RH).

最后,将这些特征输入到经过优化训练的分类器中,以预测热状态指数(TSI),即

$$TSI = g(T_s, \bar{T}_s, \bar{T}_s^g, R_p, \bar{R}_p, \bar{R}_p^g, T_e, RH) \quad (9)$$

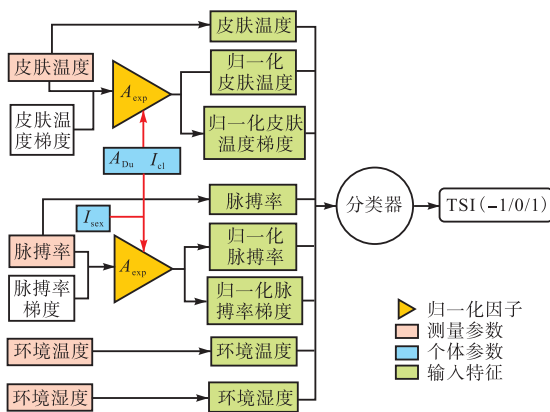


图6 ePTS模型框架

Fig.6 Framework of ePTS model

2.3 基于机器学习的热舒适度预测模型

2.3.1 数据预处理

本文使用一组标准数据进行数据预处理,将热舒适性数据集转换为机器学习模型的兼容格式.下面详细介绍数据预处理的主要阶段.

(1) 资料准备:由于热感觉投票是每5min获得一次,输入数据是每1min获得一次,因此将每次热感觉得分填充前4min,获得每1min一个标签.通过将热感觉(TSV)转化为热状态指数(TSI),热舒适感量表从7分制转换为3分制.之后将冷不舒适、舒适、热不舒适依次编码为0、1、2.

(2) 缺失值填充:用最近邻行中的对应值填充缺失条目.

(3) 离群值检测与替换:计算待检测数据的4分位数Q1、Q2、Q3、Q4和4分位距IQR(IQR=Q3-Q1);在[Q1-1.5IQR, Q3+1.5IQR]以外的点即为离群点,用最近邻的非离群值替换.

(4) 归一化:利用归一化因子对皮肤温度、皮肤

温度梯度、脉搏率、脉搏率梯度进行归一化.

(5) 数据缩放:采用min-max极差标准化方法进行处理.

(6) 数据过采样和欠采样:采用SMOTE Tomek混合采样方法来解决类不平衡问题.该方法通过SMOTE过采样扩大数据集,并通过采用Tomek Link欠采样方法来减少噪声和冗余,以消除最近邻但属于不同类别的样本^[22].

考虑模型的可泛化性,以7:3的比例将数据集划分为训练集和测试集.随之采用SMOTE Tomek混合采样方法分别对训练集和测试集进行过采样.之后在训练集上(训练集和验证集)采用5折交叉验证来预训练模型.最后通过测试集来分析评估模型性能.

2.3.2 模型性能评估

本研究使用网格搜索方法对算法中的超参数进行调整,以优化训练模型的性能.XGBoost、随机森林、SVC算法的超参数的详细设置范围、间隔和最优超参数分别见表8~表10.

表8 XGBoost算法的超参数调整范围和最优超参数
Tab.8 Hyperparameter adjustment range and optimal hyperparameter of XGBoost algorithm

超参数	范围	间隔	最优超参数
学习率	[0.1, 10.0]	0.1	0.2
树的数量	[50, 200]	5	70
基评估器	梯度提升树, 线性模型, 决策树	每个函数	梯度提升树
正则化 γ	[0, 0.2]	0.02	0.1
正则化 λ	[0, 2]	0.2	1
精减枝	[0, 20]	2	9
树的最大深度	[0, 10]	1	6

表9 随机森林算法的超参数调整范围和最优超参数
Tab.9 Hyperparameter adjustment range and optimal hyperparameter of random forest algorithm

超参数	范围	间隔	最优超参数
n_estimators	[30, 100]	10	80
max_features	[10, 20]	1	12
max_samples	[0.1, 1.0]	0.1	0.5
min_samples_leaf	[0, 0.2]	0.02	1
min_samples_split	[1, 5]	0.2	4
max_depth	[10, 140]	1	129
max_leaf_nodes	[200, 400]	20	340

表10 SVC算法的超参数调整范围和最优超参数
Tab.10 Hyperparameter adjustment range and optimal hyperparameter of SVC algorithm

超参数	范围	间隔	最优超参数
核函数	线性核, 多项式核, 径向基函数核, sigmoid核	每个函数	线性核
C	[0.1, 1, 10, 100]	每个值	10
γ	默认值, 0.01, 0.1, 1	每个值	—

表 11 深入地评估了 3 种模型的输出. 根据表 11 可知, SVC 算法在测试集上的表现最好, 准确率为 69.13%, 高于随机森林算法的准确率 62.40% 和 XGBoost 算法的准确率 60.15%. XGBoost、随机森林和 SVC 算法在热不舒适状态下(标签为 2)的表现(召回率)都不如舒适状态(标签为 1), 其原因是热不舒适中的数据量太少, 导致训练集和测试集之间存在一定量的不相关数据, 从而阻碍了算法的学习. 但是在热不舒适情况下, SVC 的表现显著高于 XGBoost 和随机森林的表现. 在冷不舒适方面(标签为 0), SVC 表现也比其他两种算法好.

表 11 3种算法在测试集上的详细评估结果

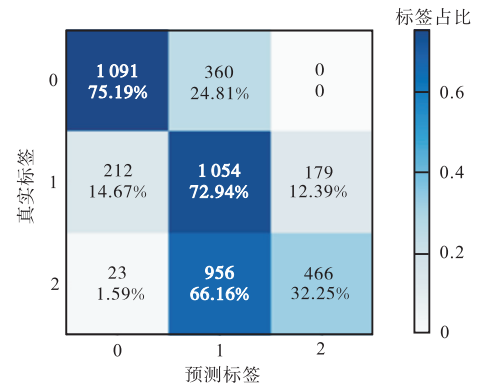
Tab.11 Detailed evaluation results of three algorithms on the test set

算法	标签	精度/%	召回率/%	F1 分数/%	支持度
XGBoost	0	82.28	75.19	78.57	1 451
	1	44.47	72.94	55.26	1 445
	2	72.25	35.25	44.59	1 445
	准确率	—	—	60.15	4 341
	宏平均	66.33	60.13	59.47	4 341
随机森林	0	82.89	78.15	80.45	1 451
	1	47.03	79.38	59.06	1 445
	2	80.15	29.62	43.25	1 445
	准确率	—	—	62.40	4 341
	宏平均	70.02	62.38	60.92	4 341
SVC	0	84.67	86.01	85.33	1 451
	1	53.42	66.44	59.22	1 445
	2	74.11	54.88	63.06	1 445
	准确率	—	—	69.13	4 341
	宏平均	70.73	69.11	69.21	4 341
	加权平均	70.75	69.13	69.23	4 341

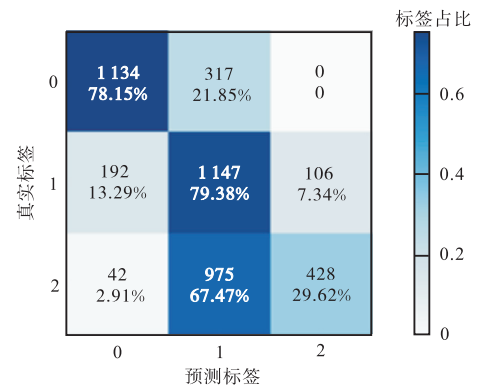
对随机森林、XGBoost、SVC 3 种算法进行优化, 得到 3 种模型在测试集上的混淆矩阵和 ROC 曲线, 分别如图 7 和图 8 所示. 在图 7 混淆矩阵中, 每个单元格中的数值代表样本绝对数量, 百分比数值代表样本比例. 在冷不舒适(标签为 0)和热不舒适(标签为 2)两种热状态下, 分类支持向量机(SVC)的准确率最高, 分别为 86.01% 和 54.88%. 在舒适状态下(标签为 1), 随机森林的分类准确率最高, 为 79.38%. 对比 3 种模型的 ROC 曲线可知, 在 3 种热状态下 SVC 算法的 AUC 值都高于 XGBoost 算法和随机森林算法的 AUC 值. 此外, SVC 算法的宏平均 AUC 值为 0.87, 高于 XGBoost 算法的 0.81 和随机森林算法的 0.78. 从 3 种模型在 3 种热状态下的表现, 可以得出 SVC 算法模型的预测效果最佳.

除了上述 3 种模型外, 还对比了其他的机器学习

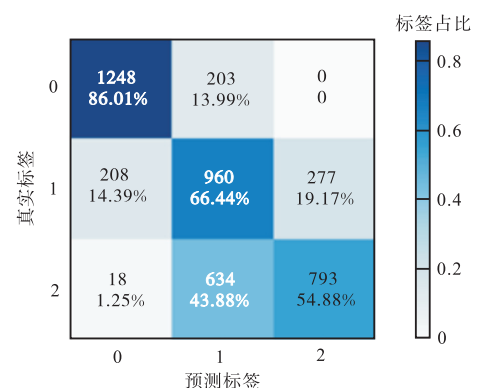
习模型, 结果如表 12 所示. 在这些机器学习算法中, 在精度、召回率、F1 分数和准确率上, SVC 算法表现最好. 由于归一化过程能够将冷不舒适、舒适、热不舒适 3 种状态拉开, 生成更清晰的非线性区域将彼此分开, 有利于 SVC 算法在高维空间寻找最优超平面, 对特征进行分类.



(a) XGBoost 预测模型



(b) 随机森林预测模型



(c) SVC 预测模型

图 7 3种模型的混淆矩阵

Fig.7 Confusion matrix of three models

特征重要性是评估机器学习模型中各个特征对预测结果贡献程度的一种方法. 表 13 显示了归一化前后每个特征对于整个随机森林热舒适预测模型的预测能力的贡献程度的排序.

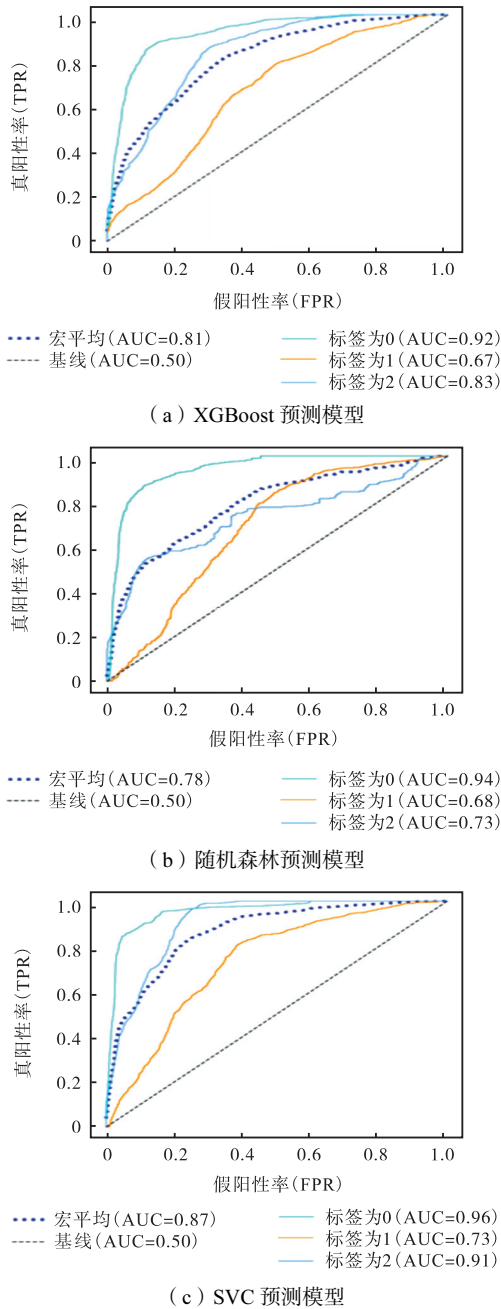


图 8 3种模型中所有类别的 ROC 曲线

Fig.8 ROC curves of all categories in the three models

从归一化后的特征重要性排序可以得出,环境温度对模型预测能力的贡献程度最大,侧面也反映出环境温度与热状态指数的高相关性.之后依次是环境湿度、手腕温度、小臂温度、归一化后的手腕温度梯度、胸口温度等.而性别特征对模型的预测能力的贡献程度最低.对模型输出影响比较大的生理参数为手腕、胸口、小臂温度以及它们衍生的特征.

在基于个体差异的归一化过程前,体重、身高两个个体差异因素的特征重要性在 0.05 以上,经过归一化后,体重、身高两个个体差异因素的特征重要性在 0.05 以下.性别个体差异因素的特征重要性略有

降低.归一化的过程降低了体重、身高、性别等个体差异对模型预测效果的影响程度.

表 12 热舒适预测模型的性能对比

Tab.12 Performance comparison of thermal comfort prediction models

算法名称	精度/%	召回率/%	F1 分数/%	准确率/%
随机森林	70.02	62.38	60.92	62.40
XGBoost	66.33	60.13	59.47	60.15
SVC	70.73	69.11	69.21	69.13
逻辑回归	69.89	67.68	67.76	67.68
决策树	60.75	56.10	56.12	56.10
梯度提升	64.26	60.60	59.94	60.60
K 邻近	65.14	61.06	61.61	61.06
自适应提升	68.70	64.07	63.20	64.07
多层感知机	65.83	62.58	62.03	62.58

表 13 归一化前后特征重要性对比

Tab.13 Comparison of feature importances before and after normalization

输入参数	归一化前的特征重要性数值	归一化后的特征重要性数值
环境温度	0.413	0.336
环境湿度	0.100	0.074
手腕温度	0.129	0.071
小臂温度	0.069	0.062
归一化手腕温度梯度	—	0.059
胸口温度	0.084	0.056
归一化胸口温度	—	0.050
归一化小臂温度梯度	—	0.036
归一化手腕温度	—	0.036
归一化小臂温度	—	0.033
身高	0.051	0.031
小腿温度	0.052	0.027
体重	0.069	0.025
归一化小腿温度	—	0.023
归一化胸口温度梯度	—	0.020
归一化小腿温度梯度	—	0.015
脉搏率	0.031	0.015
脉搏率梯度	—	0.006
性别	0.001	0.0007

3 结 论

本研究基于身高、体重、性别等个体差异因素建立增强型预测热态模型,分析 XGBoost、随机森林、SVC 共 3 种机器学习算法在测试集上的表现,并与其他机器学习算法进行比较,整体上表现最好的算法为 SVC 算法.主要结论如下.

(1) 归一化过程能够将冷不舒适、舒适、热不舒适 3 种状态拉开,生成更清晰的非线性区域将彼此分开,有利于 SVC 算法在高维空间寻找最优超平面,对特征进行分类.

(2) 归一化前后在随机森林模型上对特征的重要性进行分析,结果表明:对模型输出影响比较大的

生理参数为手腕、胸口、小臂温度以及它们衍生的特征。归一化的过程降低了体重、身高、性别等个体差异对模型预测效果的影响程度。

(3) XGBoost、随机森林、SVC 这 3 种机器学习算法在测试集上的准确率分别为 60.15%、62.40%、69.13%，在冷不舒适、舒适、热不舒适上召回率最高的依次为 SVC 算法(86.01%)、随机森林算法(79.38%)、SVC 算法(54.88%)。SVC 算法的宏平均 AUC 值为 0.87，高于 XGBoost 算法的 0.81 和随机森林算法的 0.78。因此，分类效果和泛化能力最好的是 SVC 算法。

参考文献:

- [1] Tham K W, Willem H C. Room air temperature affects occupants' physiology, perceptions and mental alertness[J]. *Building and Environment*, 2010, 45(1): 40-44.
- [2] 余娟, 朱颖心, 欧阳沁, 等. 基于生理指标评价人体热舒适、工作效率和长期健康的研究路线探讨[J]. *暖通空调*, 2010, 40(3): 1-5.
Yu Juan, Zhu Yingxin, Ouyang Qin, et al. Discussion on research routes of using physiological index to evaluate human thermal comfort, work efficiency and long-term health[J]. *Health Ventilating and Air Conditioning*, 2010, 40(3): 1-5 (in Chinese).
- [3] Kim J, Zhou Y X, Schiavon S, et al. Personal comfort models: Predicting individuals' thermal preference using occupant heating and cooling behavior and machine learning[J]. *Building and Environment*, 2017, 129: 96-106.
- [4] Luo M H, Arens E, Zhang H, et al. Thermal comfort evaluated for combinations of energy-efficient personal heating and cooling devices[J]. *Building and Environment*, 2018, 143: 206-216.
- [5] Fanger P O. *Thermal Comfort: Analysis and Applications in Environmental Engineering*[M]. Copenhagen: Danish Technical Press, 1972.
- [6] Gagge A P, Stolwijk J A J, Nishi Y. An effective temperature scale based on a simple model of human physiological regulatory response[J]. *ASHRAE Transactions*, 1971, 77: 21-36.
- [7] 储向阳, 于航, 焦瑜, 等. 与生理参数相关的热舒适实验研究综述[J]. *建筑热能通风空调*, 2018, 37(12): 60-64.
Chu Xiangyang, Yu Hang, Jiao Yu, et al. A review of experimental study on physiological parameters in the field of thermal comfort[J]. *Building Energy and Environment*, 2018, 37(12): 60-64 (in Chinese).
- [8] Choi J H, Loftness V, Lee D W. Investigation of the possibility of the use of heart rate as a human factor for thermal sensation models[J]. *Building and Environment*, 2012, 50: 165-175.
- [9] Wang Z, de Dear R, Luo M H, et al. Individual difference in thermal comfort: A literature review[J]. *Building and Environment*, 2018, 138: 181-193.
- [10] Rana R, Kusy B, Jurdak R, et al. Feasibility analysis of using humidex as an indoor thermal comfort predictor[J]. *Energy and Buildings*, 2013, 64: 17-25.
- [11] Jiang L, Yao R M. Modelling personal thermal sensations using C-support vector classification(C-SVC) algorithm[J]. *Building and Environment*, 2016, 99: 98-106.
- [12] Aqilah N, Rijal H B, Yoshida K. Gender difference regarding thermal comfort in Japanese residential building during free running mode[J]. *Building and Environment*, 2023, 245: 110891.
- [13] Dougherty K A, Chow M, Kenney W L. Critical environmental limits for exercising heat-acclimated lean and obese boys[J]. *European Journal of Applied Physiology*, 2010, 108(4): 779-789.
- [14] Feng Y X, Liu S C, Wang J L, et al. Data-driven personal thermal comfort prediction: A literature review[J]. *Renewable and Sustainable Energy Reviews*, 2022, 161: 112357.
- [15] Lan H F, Hou H Y, Gou Z H. A machine learning led investigation to understand individual difference and the human-environment interactive effect on classroom thermal comfort[J]. *Building and Environment*, 2023, 236: 110259.
- [16] Charkoudian N. Skin blood flow in adult human thermoregulation: How it works, when it does not, and why[J]. *Mayo Clinic Proceedings*, 2003, 78(5): 603-612.
- [17] Tan C L, Knight Z A. Regulation of body temperature by the nervous system[J]. *Neuron*, 2018, 98(1): 31-48.
- [18] Liu W W, Lian Z W, Deng Q H, et al. Evaluation of calculation methods of mean skin temperature for use in thermal comfort study[J]. *Building and Environment*, 2011, 46(2): 478-488.
- [19] Lan L, Lian Z W, Liu W W, et al. Investigation of gender difference in thermal comfort for Chinese people[J]. *European Journal of Applied Physiology*, 2008, 102(4): 471-480.
- [20] Chaudhuri T. *Predictive Modelling of Thermal Comfort Using Physiological Sensing*[D]. Singapore: Nanyang Technological University, 2018.
- [21] Du Bois D, Du Bois E F. A formula to estimate the approximate surface area if height and weight be known[J]. *Nutrition*, 1989, 5(5): 303-311.
- [22] Han X G, Hu Z Q, Li C, et al. Prediction of human thermal comfort preference based on supervised learning[J]. *Journal of Thermal Biology*, 2023, 112: 103484.

(责任编辑:孙立华)