

DOI:10.19322/j.cnki.issn.1006-4710.2026.02.014

<https://xuebao.xaut.edu.cn>

引文格式:李媛,张潇文,舒平.住宅室内风舒适的敏感参数分析与预测方法比较研究[J].西安理工大学学报,2026,42(2):292-300.

Li Yuan,Zhang Xiaowen,Shu Ping. Analysis of sensitive parameters and comparative study of predictive methods for indoor wind comfort in residential buildings[J]. Journal of Xi'an University of Technology,2026,42(2):292-300.

住宅室内风舒适的敏感参数分析与 预测方法比较研究

李媛^{1,2,3}, 张潇文¹, 舒平^{1,2,3}

(1. 河北工业大学 建筑与艺术设计学院, 天津 300130; 2. 河北工业大学 城乡更新与建筑遗产保护中心, 天津 300130; 3. 河北工业大学 河北省健康人居环境重点实验室, 天津 300130)

摘要:为解决住宅室内风舒适的敏感影响因素和快速预测问题,本文首先从流体力学理论出发假设了六个套型参数,并通过CFD模拟抽样套型的室内风舒适面积比。而后应用随机森林重要特征排序方法,得出套型距地高度、进出风口面积比、进风夹角和隔墙密度值为影响室内风舒适的高敏感参数。最后应用三种机器学习算法构建高敏感套型参数与风舒适面积比的预测模型,通过回归拟合误差对比得出梯度提升回归模型的预测效果最优,并通过误差统计检验法验证了该模型的稳定性和准确性。

关键词:居住建筑;室内风舒适;套型参数;敏感性分析;预测模型

中图分类号: TU241.8

文献标志码: A

文章编号: 1006-4710(2026)02-0292-09

Analysis of sensitive parameters and comparative study of predictive methods for indoor wind comfort in residential buildings

Li Yuan^{1,2,3}, Zhang Xiaowen¹, Shu Ping^{1,2,3}

(1. School of Architecture & Art Design, Hebei University of Technology, Tianjin 300130, China;

2. Urban and Rural Renewal and Architectural Heritage Protection Center, Hebei University of Technology, Tianjin 300130, China; 3. Key Laboratory of Healthy Human Settlements in Hebei Province,

Hebei University of Technology, Tianjin 300130, China)

Abstract: To address the sensitive influencing factors of indoor wind comfort in residential buildings and the issue of rapid prediction, this article firstly hypothesizes six parameters for a unit according to fluid mechanics theory, and applies CFD to simulate the indoor wind comfort area ratio of sampled apartment types. Then, it analyzes four highly sensitive parameters, including the distance from the ground to the apartment type, the ratio of the area of the air inlet to outlet, the angle of the air inlet and the density of the partition wall, by the random forest feature importance ranking method. Finally, it uses three machine learning algorithms to construct the prediction models of the highly sensitive parameters and the wind comfort area ratio. Through a comparison of regression fitting errors, it concludes that the prediction effect by the gradient boosting regression model is the best, whose stability and accuracy is verified through error statistical tests.

收稿日期: 2025-09-26; 网络首发日期: 2026-02-13

网络首发地址: <https://link.cnki.net/urlid/61.1294.n.20260212.1445.002>

基金项目: 国家自然科学基金资助项目(52078178)

第一作者: 李媛,女,博士,高级工程师,研究方向为建筑性能化设计、建筑参数化设计。E-mail:498541446@qq.com

通信作者: 舒平,男,博士,教授。研究方向为微型居住空间健康适居性、建筑与设计教育。E-mail:shuping@hebut.edu.cn

Key words: residential buildings; indoor ventilation comfort; unit parameter; sensitivity analysis; predictive mode

随着城市化进程加快和能源消耗加剧,营造良好的自然通风环境愈发重要。自然通风的合理利用不仅有助于降低建筑能耗,实现节能减排^[1],还能改善城市居民的健康状况^[2],对城市可持续发展具有积极意义。因此,通过住宅套型设计来优化室内自然通风舒适度一直是研究者们关注的问题。

研究内容方面,国内外学者对影响居住建筑室内通风性能的套型特征进行了广泛研究。Hu等^[3]通过现场测量和数值模拟研究了风速一定时,不同朝向向下住宅室内的换气效率差异可达2.6倍,其影响显著。另外,Zhang等^[4]通过CFD模拟发现建筑不同进风口数量与尺寸对内部通风量具有明显影响,其中双进风口与较大出风口组合时通风量最大。Gao和Lee^[5]探讨了香港住宅中开窗位置、朝向和各房间户门位置对平均空气龄的影响依次减小。除以上参数之外,有学者指出室内布局与隔断对室内通风的影响不容忽视^[6]。以上文献的研究对象均为一间卧室或一个典型套型的常见形态参数(朝向、开窗位置、进出风口面积比等),但现实中套型及布局多种多样。

研究方法方面,软件模拟通风性能的方法操作复杂、耗时长,随着机器学习技术的发展,数据驱动预测模型应运而生,在保持准确度的前提下大幅提升预测效率。预测方法一般分为线性与非线性两种。姚佳伟等^[7]利用多元线性回归分析了上海工人新村9种室外空间特征变量与行人风舒适的关系,并讨论了线性模型的局限性。Ding等^[8]开发出基于CFD模拟和机器学习耦合的高密度城市通风数据驱动模型,并对比得出非线性回归模型的准确性优于线性模型。Qin等^[9]基于实时监测与CFD模拟获取室内空气环境数据,选择支持向量回归非线性方法创建较准确的室内空气环境预测模型。由此可见通风性能(因变量)受多个影响因素(自变量)耦合作用的特点适合采用非线性方

法构建模型,但不同非线性回归方法的适用性与准确性还需对比研究。

由于上述文献对住宅室内风舒适影响因素研究不全面,以及缺乏人工智能预测该方法的方法择优,本文将套型布局纳入到影响因素中,并将研究对象拓展至多种套型,在采用随机森林特征重要性方法筛选出对室内风舒适高敏感的影响参数后,对比支持向量回归法、多层感知回归法和梯度提升回归法构建风舒适预测模型的准确性。

1 研究对象与数据获取

1.1 拟定影响室内风舒适的套型参数

合理假定影响室内通风的套型参数是确保研究有效性的前提,本文类比流体在管状物中流速的影响因素拟定这些参数。在流体力学中,流体在管道中的流速计算公式,以及从影响流速的参数类比推理或假设的套型参数见表1。其中,套型进风风压与两个因素相关:进风风速和进风夹角。进风风速与进风口距离地面的高度呈正相关关系^[10],在进风风速相同的前提下,进风风压与进风夹角也呈正相关关系,垂直进风时风压最大。管口截面积和管道宽长比可直接类比为进出风口面积比与面宽进深比。管道的粗糙程度可以理解为阻碍流速的因素,由于没有直接对应的套型参数,所以本研究假设了隔墙密度值和通风阻碍度两个特征参数来表征阻碍通风的程度。其中,通风阻碍度描述了风在套型中通过路径的曲折程度:比如单一空间的通风路径由于进出风口的相对位置不同导致通风路径的转折,每次转折记为一个节点,见图1;住宅套型整体空间通风阻碍度系累加每个进出风口之间路径的转向节点数除以进出风口数量乘积得到的加权值(图2)。套型参数的定义及取值范围见表2。

表1 流体力学中的流速计算公式^[11]

Tab. 1 Flow velocity calculation formulas in fluid mechanics

公式名称	影响参数	与流速的关系	套型类比参数
伯努利公式 ^①	管道两端压差	压差与流速正相关	距地高度与进风夹角
连续性公式 ^②	管口截面积	流量一定时,截面积与流速负相关	进出风口面积比
Darcy-Weisbach 公式 ^③	管壁粗糙度/宽长比	压差一定时,粗糙度、宽长比与流速负相关	面宽进深比、隔墙密度、通风阻碍度

注: ① $\Delta P = \frac{1}{2} \rho (V_2^2 - V_1^2)$, ΔP 为压力差(Pa), ρ 为流体密度($\text{kg} \cdot \text{m}^{-3}$), V_1 和 V_2 为进出口流速($\text{m} \cdot \text{s}^{-1}$)。② $V_1 A_1 = V_2 A_2$, A_1 和 A_2 为进出口截面积(m^2), V_1 和 V_2 为进出口流速($\text{m} \cdot \text{s}^{-1}$)。③ $\Delta P = f \frac{L}{D} \frac{\bar{v}^2}{2g}$, ΔP 为压力差(Pa), f 为摩擦因数,无量纲, L 为管道长度(m), D 为管道直径(m), $\bar{v}$ 为平均流速($\text{m} \cdot \text{s}^{-1}$), g 为重力加速度,一般取 9.8 m/s^2 。

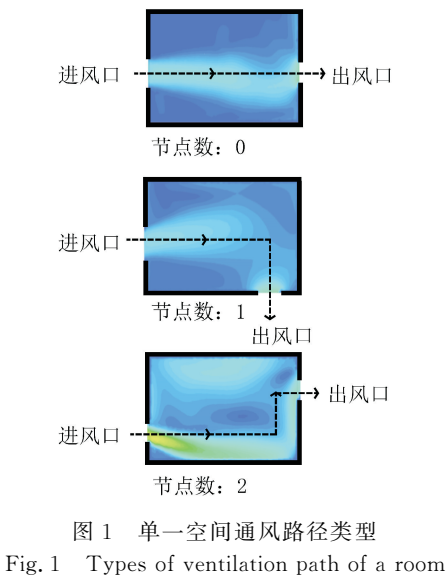
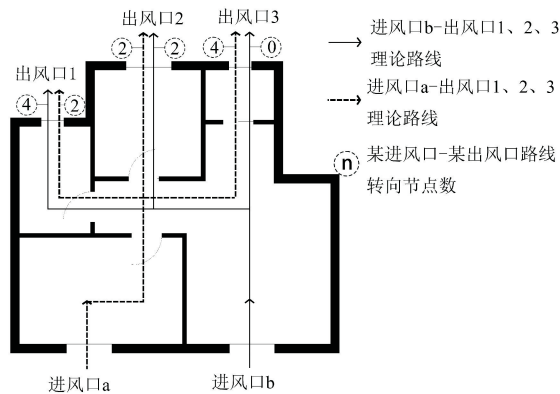


图 1 单一空间通风路径类型
Fig. 1 Types of ventilation path of a room



注:该套型通风阻碍度计算:
$$[(4+2+4)+(2+2+0)]\div(2\times3)=2.32$$

图 2 某套型的通风阻碍度
Fig. 2 Ventilation obstruction degree of a unit

表 2 套型参数定义及取值范围
Tab. 2 Definition and value range of unit parameters

编号	参数名称	含义	取值范围及步长
1	距地高度/m	套型地面距室外地坪高度	9~45 m(步长 9 m) ^①
2	进风夹角/(°)	风向与建筑迎风面所呈角度	0°~90°(步长 15°) ^②
3	进风口面积比	进风口有效面积/出风口有效面积	
4	面宽进深比	面宽/进深	
5	隔墙密度值	套型内隔断面积/套型面积	
6	通风阻碍度	计算公式: $\frac{\sum (N_1 + N_2 + \cdots + N_n)}{n_{\text{进}} \times n_{\text{出}}}$ (式中: N_n 为进风通过某条进风口及室内户门形成的通风路径的转向节点数; $n_{\text{进}}$ 、 $n_{\text{出}}$ 分别为进、出风口数量)	根据实际套型计算 ^③

注:①距地高度考虑到近地面影响与梯度风分布,选取住宅中间层数进行试验,避免极低和极高速对室内通风环境的不利影响以及节省计算资源;②进风角度根据天津高层住宅的主要朝向与天津夏季主导风向(东南风)的夹角决定;③进风口面积比、面宽进深比、隔墙密度值和通风阻碍度的取值是由典型套型的若干增扩套型计算得来,增扩套型的房间布局、尺寸和开启窗扇面积须符合住宅设计规范,因此取值范围因具体套型而异,无法统一,步长也不连续。

1.2 确定通风评价指标

舒适风速面积比是指套型内处于舒适风速范围内的面积与其使用面积之比,这一比值是衡量住宅通风舒适度的重要指标之一,本文将该指标设定为自然通风性能评价指标。现阶段国内外的规范中并未对室内舒适风速的范围进行明确规定,不少学者认为可根据具体研究问题在合理范围内自行设定。本文参考《建筑节能设计与技术》中风速对人体舒适度影响的描述(表 3),将舒适风速范围设置为 0.15~1 m/s。

表 3 风速对人体舒适度影响^[12]
Tab. 3 Wind speed's impact on human comfort

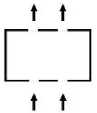
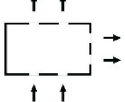
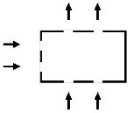
风速/(m·s ⁻¹)	对舒适度的影响
[0,0.15]	不易察觉,稍有不适
(0.15,0.5]	愉快,不影响工作
(0.5,1.0]	稍微察觉到吹风感,一般愉快,但必须提防纸被吹散
(1.0,1.5]	有吹风感,工作台面纸张易吹散
(1.5,7.0]	明显的风击感

1.3 构建套型数据集

为遴选出对室内风舒适敏感性高的套型参数并建立二者间的关系,本文构建了天津市内六区既有小区的常见套型数据集。首先利用爬虫工具爬取天津链家网的套型,统计结果为二居室套型占比最大(44.4%),三居室其次(32.5%),两类套型总占比76.9%。因此数据库只包含二居室和三居室,共计920套,并将全部套型归纳为Ⅰ、Ⅱ、Ⅲ、Ⅳ四种不同进出风类型(表4)。后续研究将在数据库中随机抽取四种进出风类型的套型,以确保抽样代表性。

表4 套型的不同通风类型

Tab.4 Different ventilation types of residential units

通风类型	进风口位置	图示
Ⅰ	出风口在进风口对面	
Ⅱ	出风口在进风口侧面	
Ⅲ	出风口在进风口的对面和侧面	
Ⅳ	相邻两侧进风只有一个出风口	

1.4 获取通风模拟数据

本文训练模型所使用的数据均来自于CFD数值模拟,该方法是近十年来建筑自然通风领域学者们最常用的数值模拟方法^[13]。本文选用ANSYS Fluent作为数值模拟软件,利用Space Claim软件进行建模。模拟风向选取天津夏季高频风向——东南风,入射风速采用梯度风^[10]来表示立面的风速分布,以提高模拟结果的准确性和可靠性,计算公式为:

$$U_z = U_0 (Z/Z_0)^\alpha \quad (1)$$

式中, U_z 为高度 Z 处的风速,单位为m/s; U_0 设置为国家气象官网2023年天津夏季(6~8月)风速均值^①2.5 m/s; Z 为住宅高度,单位为m; Z_0 为气象站风观测点基准高度,10 m; α 为地面粗糙系数,按照实际情况取值0.3,无量纲^[14]。

在模型参数设置上,假设建筑高度为 H ,模拟范围设定入流方向、建筑两侧和上方均为 $5H$,流出距离为 $10H$;设定最小网格尺寸为100 mm,最大网

格尺寸6 000 mm,网格过渡比为1.2,处理后网格数量约为250万个,各工况网格数目略有不同^[14]。模拟中,模型假设每个房间的门与窗户完全开启,模型的构建只考虑套型空间本身,暂不考虑其它因素如室内陈设、结构构件等;计算时选用标准 $k-\epsilon$ 湍流模型进行通风模拟,利用SIMPLE算法对模型进行稳态模拟的通风迭代计算,设定收敛精度为 10^{-4} ,迭代步数为1 000次,计算达到收敛即停止。

1.5 数据预处理

为保证数据分析的可靠性,模拟数据在训练分析前均进行了数据预处理,包括:重复值剔除、缺失值处理与数据归一化处理。其中,归一化处理使用(0,1)标准化方法,将自变量数据统一缩放在0~1的范围内,可解决输入变量数值偏差大、单位不统一等问题。

2 风舒适的参数敏感性分析

为筛选高敏感的套型参数,需分析各套型参数对室内风舒适的全局影响。考虑到本实验研究自变量与因变量间的非线性关系及自变量之间交互影响的特点,更适合采用非参数回归方法中的随机森林(random forest, RF)重要性分析法。该方法作为一种集成学习技术,拥有泛化能力良好、有效防止过拟合、适用于离散型数据、考虑特征间交互作用,并提供全局变量重要性度量^[15]等优点,广泛应用于城市与建筑设计的参数敏感性分析中。

2.1 特征重要性计算

随机森林法在训练模型的过程中,采用有放回的随机抽样方式抽取训练数据,有的样本会被重复抽取,而有的样本不会被抽取,后者被称为袋外数据(out-of-bag, OOB),可用来评估模型泛化性能或者计算特征重要性。具体算法为:随机置换每个特征的值,观察模型预测性能的变化,回归误差变化越大,表示特征重要性越高。特征重要性计算公式^[16]为:

$$P_i = \frac{1}{N} \sum_k (L_k^{\text{OOB}} - L_{k,i}^{\text{OOB}}) \quad (2)$$

式中, P_i 为第 i 个自变量特征的重要性度量; N 为随机森林中树的总数量; k 为第 k 棵树; L_k^{OOB} 为第 k

① 本文的室外入射风向和风速为天津夏季主导风向和平均风速,实时风的多样性某种程度上体现在套型参数中建筑高度导致的梯度风和入射风与迎风面的不同夹角。待本文比较预测方法后,再考虑实时风向、风速和住宅群体布局对入射风的影响,使入射风的取值范围更全面。

棵树对未施加扰动的 OOB 数据计算出的回归误差 (默认为均方误差); $L_{k,i}^{OOB}$ 为对样本数据中第 i 个自变量特征实施扰动后第 k 棵树重新计算的 OOB 数据的回归误差。

2.2 构建随机森林回归模型

机器学习回归模型均通过调整自身特有的超参数来调整模型性能,RF 模型的主要超参数为树的数量与叶子数,其余参数为默认设置。首先通过遍历超参数取值训练 RF 模型,利用其特有的 OOB 数据均方误差判定模型是否稳定、准确来确定最优超参数取值。

在套型数据集中抽取二、三居室共 30 个不同套型,变化六个套型参数的取值将 30 个不同套型增扩为 234 个套型,以增加各参数的取值范围,并模拟得到各套型的舒适风速面积比。现以通风类型 I 中的套型 A(图 3)为例说明增扩过程:距地高度和进风夹角按表 2 取值;根据居住建筑设计规范,调整套型 A 的墙体位置和外窗尺寸,进出风口面积比可取值 0.55、0.60、0.65、0.70 和 0.80;面宽进深比可取值 0.8、0.9、1.1、1.2 和 1.3;隔墙密度和通风阻碍度取值见表 5,后四个参数将套型 A 增扩至 20 个套型,再考虑与距地高度和进风夹角排列组合,可增扩至 $20\times5\times5=500$ 个套型供随机抽取。

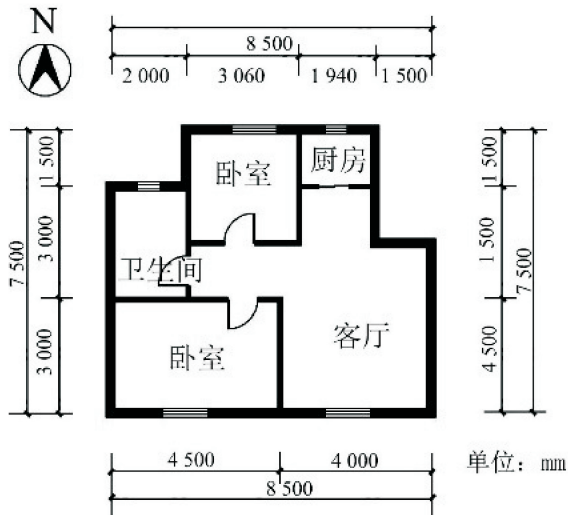


图 3 套型 A 平面图
Fig. 3 Plan of unit A

最后,利用 MATLAB R2024b 软件构建 RF 模型,并使用 OOB 误差监控模型性能确定超参数组合:将树的数量的取值范围设置为 0~1 000,步长为 100,叶子数取值范围设置为 1~20,步长为 5,训练结果见图 4。随着叶子数变少,模型最小误差的收敛值也在变小。为保证模型收敛后均方误差最小且节省计算资源,选择树的数量为 500,叶子数为 1 的

最佳 RF 模型。

表 5 套型 A 的隔墙密度与通风阻碍度取值
Tab. 5 Partition wall density and ventilation obstruction degree of unit A

隔墙密度值	通风阻碍度
 0.034	 1.300
 0.036	 2.000
 0.037	 2.300
 0.038	 3.000
 0.040	 3.200

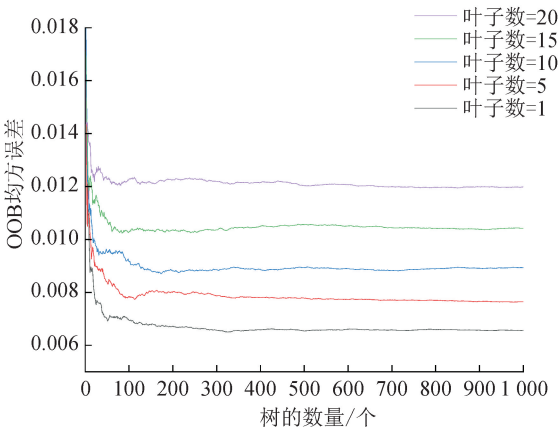


图 4 RF 模型 OOB 误差变化图
Fig. 4 OOB error in the RF model

2.3 筛选高敏感的套型参数

为避免敏感参数筛选过程中存在的偶然性,需多次计算重要性评分后取均值。本文采用上述最佳模型计算各参数重要性评分 500 次,统计结果见图 5。参数的敏感程度均值排序依次为距地高度、进出风口面积比、风向夹角、隔墙密度值、面宽进深比和通风阻碍度。其中,前四项的重要性评分均值全部大于阈值 0.15,总贡献度达 80%,被认定为高敏感参数,而后两项由于影响较小,可被忽略。接下来将采用四个高敏感参数构建预测模型。

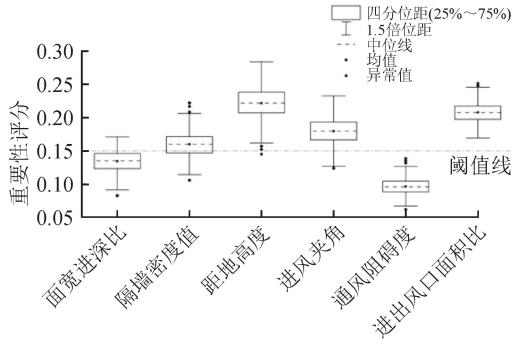


图5 训练500次特征重要性评分箱线图

Fig. 5 Box plot of feature importance scores from 500 trainings

3 预测模型构建方法与验证

由于本研究的因变量和自变量之间具有复杂的非线性关系,因此选择机器学习中的支持向量回归,多层感知回归和梯度提升回归三种适用于非线性回归的算法构建预测模型,并对比择优。

3.1 机器学习方法

3.1.1 支持向量回归

支持向量回归 (support vector regression, SVR) 适用于解决线性及非线性数据回归问题,其核心思想是通过定义不敏感损失函数来构造“容忍带”,允许预测值在真实值的固定容差范围(默认为 ± 0.1)内波动来拟合数据^[17],流程如下^[18]:

1) 线性映射:

$$f(x_i) = \omega^T(x_i) + b \quad (3)$$

2) 拟合优化:

$$\min \left\{ \frac{1}{2} \omega \cdot \omega + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right\}$$

$$\text{约束条件} \begin{cases} f(x_i) - y_i \leq \varepsilon + \xi_i \\ y_i - f(x_i) \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (4)$$

3) 非线性映射:

$$f(x_i) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(x_i, x) + b \quad (5)$$

式中: x_i 为第 i 个输入的样本; $f(x_i)$ 为第 i 个样本的预测值; y_i 为第 i 个样本的真实值; ω 为权重向量; ω^T 为转置后的权重向量; b 为偏置项; ξ_i 与 ξ_i^* 为松弛因子; ε 为拟合误差; C 为误差惩罚系数; α_i 与 α_i^* 均为拉格朗日乘子; $k(x_i, x)$ 为非线性映射核函数。该方法的主要超参数为惩罚因子与核函数参数。

3.1.2 多层感知回归

多层感知回归 (multi-layer perception regression, MLPR) 是一种监督学习算法,一般包含输入

层、隐藏层和输出层,该算法具有多个隐藏层和误差反向传播 (back propagation, BP) 的特点。其训练过程为首先正向传播初步预测,而后反向传播计算误差进行反馈调整,反复迭代此过程^[19],逐步优化训练参数直到损失函数达到最小化,即预测结果逼近真实值时训练完毕。多层感知回归^[20]训练过程如下:

1) 正向传播:

$$y_n^a(k) = f\left(\sum_b \omega_{n,b}^a(k) y_{n-1}^b(k)\right) \quad (6)$$

2) 反向调整:

$$\omega_{n,b}^a(k+1) = \omega_{n,b}^a(k) + \nu \Delta \omega_{n,b}^a(k) \quad (7)$$

3) 误差迭代:

$$E = 1/2N \sum_{k=1}^N [d(k) - y(k)]^2 \quad (8)$$

式中: k 为输入的第 k 个样本; $y_n^a(k)$ 为 n 层第 a 个神经元的输出; $\omega_{n,b}^a(k)$ 为连接第 $n-1$ 层第 b 个神经元与第 n 层第 a 个神经元的权重值; $y_{n-1}^b(k)$ 为第 $n-1$ 层第 b 个神经元的输出; f 为激活函数(默认设定为 tansig); $\omega_{n,b}^a(k+1)$ 为第 $k+1$ 个样本时前一层的第 b 个神经元指向 n 层第 a 个神经元的权重值; ν 为学习率; $\omega_{n,b}^a(k)$ 为权重调整值; E 为模型均方误差,默认为 10^{-2} ; N 为训练样本数; $d(k)$ 为样本真实值; $y(k)$ 为样本预测值。该方法主要超参数为学习率、正则化系数、隐藏层数量与神经元数量。

3.1.3 梯度提升回归

梯度提升回归 (gradient boosting regression, GBR) 通过顺次使用多个能力普通的决策树模型(弱学习器),累积修正最终得到一个强大的预测模型(强学习器)。其核心思想是逐步拟合实际值与预测值的差值(残差),使之不断缩小,让预测值向真实值移动,即沿着损失函数(均方误差)下降最快的方向(负梯度方向)不断前进,最终达到损失函数最小化^[21]。其训练过程^[22]如下:

1) 初始化预测:

$$f_0(x_i) = \underset{c}{\operatorname{argmin}} \sum_{i=1}^N L(y_i, c) \quad (9)$$

2) 计算损失函数的(伪残差)负梯度值并据此重新拟合更新模型:

$$r_{i,m} = - \left[\frac{\partial L(y_i, f_{m-1}(x_i))}{\partial f_{m-1}(x_i)} \right] \quad (10)$$

$$f_m(x_i) = f_{m-1}(x_i) + \eta h_m(x_i) \quad (11)$$

3) 输出最终模型:

$$f_M(x_i) = f_0(x_i) + \nu \sum_{m=1}^M h_m(x_i) \quad (12)$$

式中: x_i 为第 i 个输入的样本; $f_0(x_i)$ 为第 i 个样本的初始预测值; y_i 为第 i 个样本的真实值; L 为

模型损失函数; c 为常数值; $r_{i,m}$ 为第 m 轮迭代时第 i 个样本的伪残差; $f_{m-1}(x_i)$ 为前 $m-1$ 轮迭代后对第 i 个样本的累计预测; $\partial L(y_i, f_{m-1}(x_i))$ 为损失函数的导数; $\partial f_{m-1}(x_i)$ 为当前预测值的导数; $f_m(x_i)$ 为经过 m 轮迭代后的对第 i 个样本的完整预测函数; ν 为学习率; $h_m(x_i)$ 为第 m 棵决策树对第 i 个样本的预测; M 为树的总数量(迭代轮数); $f_M(x_i)$ 为经过 M 轮迭代对第 i 个样本的最终预测输出。该方法的主要超参数为树的数量、叶节点、树的深度与学习率。

3.2 模型训练过程

模型训练时应采用相同的数据集和数据划分方式。在 2.2 节已选套型的基础上,再在数据集中抽取二居室与三居室共 120 个套型,并变化高敏感套型参数将套型增扩至 600 套。按照 8:2 的比例划分训练集与测试集,通过网格搜索结合五折交叉验证遍历所有超参数组合,选取最优的超参数组合训练模型(其余参数保持默认设置),当各模型交叉验证均方根误差最低时即可确定最佳超参数组合(表 6),此时输出三种算法的最佳回归模型。

表 6 各模型最优超参数

Tab. 6 Optimal hyperparameters for model

支持向量回归	多层感知回归	梯度提升回归
	正则化参数:0.02	树的数量:600
惩罚因子:2.5	学习率:0.01	叶节点:10
核函数:0.12	隐藏层数量:10	树的深度:3
	神经元数量:7	学习率:0.14

3.3 三个预测模型性能比较

3.3.1 模型预测性能评估指标

机器学习回归模型的性能评价指标一般包括:决定系数(R^2)、平均绝对误差(mean absolute error, MAE)和均方根误差(root mean square error, RMSE)。其中, R^2 用于衡量模型拟合优度,范围为 0 到 1,越接近 1 说明拟合越优;MAE 反映预测值与实际值绝对误差的平均水平,值越小表明预测精度越高;RMSE 则用于评估预测值与实际值之间的均方根误差,值小说明模型预测精度越好。各指标计算公式^[23]详见式(13)~(15):

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (13)$$

$$E_M = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (14)$$

$$E_R = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (15)$$

式中: E_M 为平均绝对误差; E_R 为均方根绝对误差; y_i 为第 i 个样本实际值; $\hat{y}_i$ 为第 i 个样本预测值; $\bar{y}$ 为所有样本实际值的均值; n 为样本数量。

在评估模型预测性能时,一般选择均方根误差为模型训练与评估的主要指标,其余为辅助指标。

3.3.2 模型预测性能评估结果

三个最佳模型的测试集拟合效果见图 6,其均方根误差、平均绝对误差和决定系数对比见图 7,梯度提升回归模型测试集的性能指标整体优于多层感知回归和支持向量回归模型。

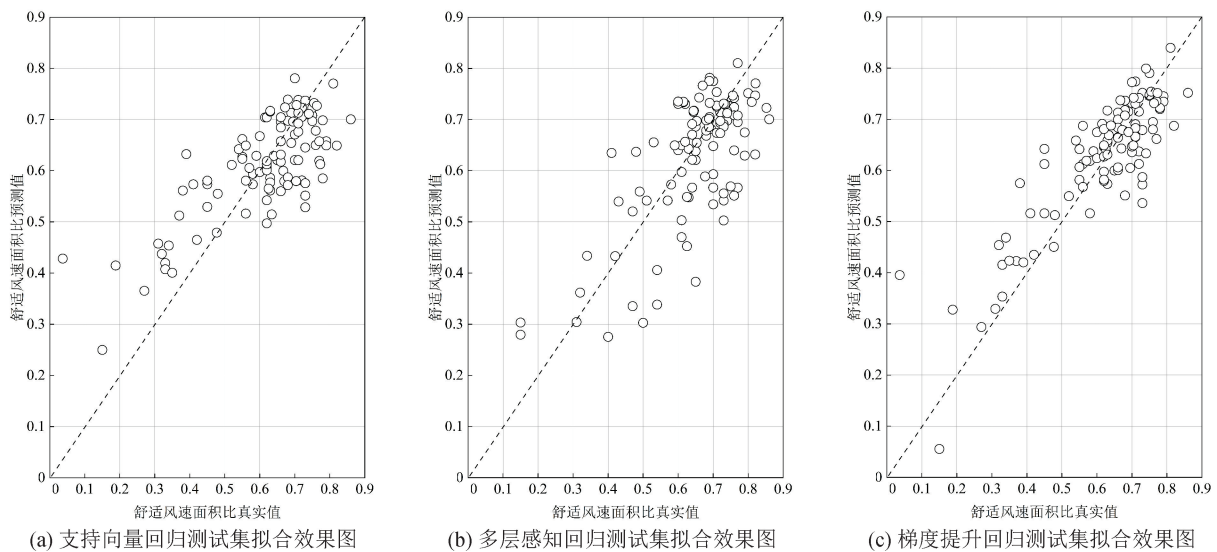


图 6 三种模型测试集性能效果图

Fig. 6 Results of three model test set performance

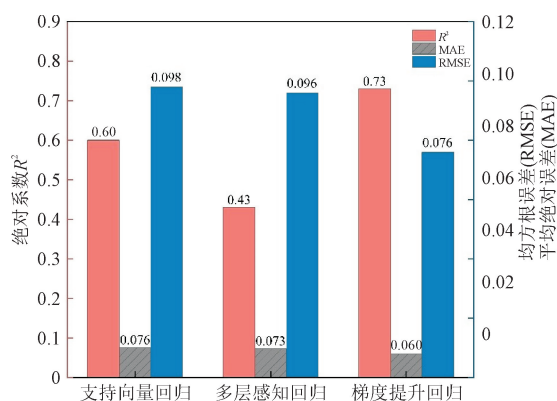


图7 三种模型评估指标对比图

Fig. 7 Comparison chart of evaluation metrics for three models

梯度提升回归模型测试集均方根误差相较支持向量回归模型与多层感知回归模型分别降低了21.1%和17.8%，其平均绝对误差相较支持向量回归模型与多层感知回归模型分别降低了22.4%和20.8%，其决定系数相较于支持向量回归模型与多层感知回归分别提升了21.7%和69.7%。综合其结果，本文选择梯度提升回归模型作为住宅室内风舒适的预测模型。

3.4 预测模型有效性验证

为验证预测模型对未知数据的预测准确性，又随机抽取了30个未参与机器学习模型训练的常见住宅套型进行验证，其CFD数值模拟结果与梯度提升回归模型预测结果对比见图8，30组数值的平均相对误差为8.16%。

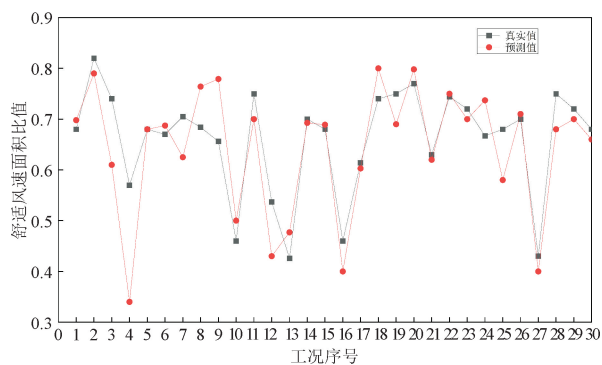


图8 30个验证工况通风预测值与真实值

Fig. 8 Comparison of predicted and actual values or 30 validation conditions

另外，为避免随机抽样的偶然概率和小样本估计的不稳定性，通过非参数统计方法计算出30个套型的平均相对误差有95%概率落在5.61%~11.38%之间，该误差范围在统计学标准的可接受范围内^①，误差验证结果可靠，说明梯度提升回归模型预测具有可靠的准确性。

4 结 论

本文基于流体力学理论假定了六个影响室内通风的套型参数，在合理增扩天津市高层住宅典型套型样本的基础上，应用随机森林重要性分析法进行权重排序。而后利用支持向量回归、多层感知回归、梯度提升回归三种机器学习方法构建出三个高敏感套型参数与室内风舒适面积比的预测模型，并用多种评估指标进行对比择优。

1) 随机森林权重分析结果显示，套型参数对室内舒适风速面积比的影响重要程度排序为：距地高度、进出风口面积比、进风夹角、隔墙密度值、面宽进深比和通风阻碍度。由于前四个参数的重要性评分均值全部大于阈值0.15且总贡献度近80%，因此被认定为高敏感参数，即成为预测模型的自变量。

2) 对比三种机器学习方法的拟合效果发现：梯度提升回归模型比其他两种模型在RMSE、MAE、 R^2 三种评估指标中的综合表现更好，具有更高的预测准确性和更好的模型解释性。而后，应用误差统计检验方法验证了梯度提升回归模型的误差范围有95%的概率落在5.61%~11.38%之间，进一步证明了该模型在统计学意义上的预测稳定性与准确性。

总之，本研究为天津高层住宅的风舒适预测提供了一个快速预测模型，有助于建筑师在住宅初步设计阶段进行设计决策，并为既有住宅的通风舒适优化设计提供依据。

参考文献：

- [1] 李晓旭,黄凯良,冯国会,等.严寒地区超低能耗建筑夜间通风节能潜力分析[J].沈阳建筑大学学报(自然科学版),2022,38(3):520-527.
Li Xiaoxu, Huang Kailiang, Feng Guohui, et al. Analysis of night ventilation energy saving potential for ultra-low energy consumption buildings in severe cold area[J]. Journal of Shenyang Jianzhu University(Natural Science), 2022 38(3): 520-527.
- [2] Mustafa M, Cook M J, McLeod R S. A critical review of ventilation effectiveness in naturally ventilated spaces from the perspective of sustainability and health[J]. Building and Environment, 2025, 270: 112471.
- [3] Hu Hong, Kikumoto H, Ooka R. Effect of wind direction on natural ventilation in a multiple-room house via field measurements and numerical simulations[J].

① 平均相对误差分级标准^[24]中预测良好范围设定为10%~20%，本文据此设定为15%。

- Journal of Wind Engineering and Industrial Aerodynamics, 2024, 248: 105718.
- [4] Zhang Xuelin, Weerasuriya A U, Wang Jiayao, et al. Cross-ventilation of a generic building with various configurations of external and internal openings[J]. Building and Environment, 2022, 207: 108447.
- [5] Gao C F, Lee W L. Evaluating the influence of openings configuration on natural ventilation performance of residential units in Hong Kong[J]. Building and Environment, 2011, 46: 961-969.
- [6] Yin Xin, Muhieddeen M W, Razman R, et al. The potential effects of window configuration and interior layout on natural ventilation buildings: a comprehensive review [J]. Cleaner Engineering and Technology, 2024, 23: 100830.
- [7] Zhang Xingzhao, Yang Luqiao, Luo Ruizhe, et al. Estimating the outdoor environment of workers' villages in East China using machine learning[J]. Building and Environment, 2022, 226: 109738.
- [8] Ding Chao, Lam K P. Data-driven model for cross ventilation potential in high-density cities based on coupled CFD simulation and machine learning [J]. Building and Environment, 2019, 165: 106394.
- [9] Qin Hao, Wang Xiaoxu. A multi-discipline predictive intelligent control method for maintaining the thermal comfort on indoor environment[J]. Applied Soft Computing, 2022, 116: 108299.
- [10] Li Qi, Chen Jiayu, Luo Xiaowei. Estimating omnidirectional urban vertical wind speed with direction-dependent building morphologies[J]. Energy and Buildings, 2024, 303: 113749.
- [11] 王惠民. 流体力学基础[M]. 第 3 版. 北京:清华大学出版社,2015.
- [12] 宋德萱. 建筑节能设计与技术[M]. 北京:中国建筑工业出版社,2003.
- [13] 许江英,陈宏,熊鱼雅. 国内外建筑自然通风研究比较——基于文献计量分析[J]. 南方建筑,2021(6): 114-123.
- Xu Jiangying, Chen Hong, Xiong Yuya. Comparison of natural ventilation of buildings in China and abroad based on bibliometric analyses[J]. South Architecture, 2021(6): 114-123.
- [14] JGJ/T 449—2018 民用建筑绿色性能计算标准[S].
- [15] Wei Pengfei, Lu Zhenzhou, Song Jingwen. Variable importance analysis: a comprehensive review[J]. Reliability Engineering & System Safety, 2015, 142: 399-432.
- [16] 张志宇,甘楠林,刘伟,等. 基于随机森林的自密实再生混凝土抗压强度预测研究[J]. 河南科学,2024,42(12):1773-1783.
- Zhang Zhiyu, Gan Nanlin, Liu Wei, et al. Prediction of compressive strength of self-compacting recycled concrete based on random forest[J]. Henan Science, 2024, 42(12): 1773-1783.
- [17] Zhao Guijie, Pan Xiaoqiang, Yan Huan, et al. Predicting engineering properties of controlled low-strength material made from waste soil using optimized SVR models[J]. Case Studies in Construction Materials, 2024, 20: e03325.
- [18] Huang Zhenqian, Huang Zhen, An Pengtao, et al. Reconstruction and prediction of tunnel surrounding rock deformation data based on PSO optimized LSSVR and GPR models[J]. Results in Engineering, 2024, 24: 103445.
- [19] Zheng Naihao, Zhang Weiping, Zhou Yong, et al. Confinement strength prediction of corroded rectangular concrete columns using BP neural networks and support vector regression [J]. Structures, 2024, 67: 107021.
- [20] Zhang Lidong, Luo Zhenmin, Su Bin, et al. Prediction of explosion hazard of aluminum powder two-phase mixed system using random forest based on K-fold cross-validation[J]. Journal of Loss Prevention in the Process Industries, 2025, 94: 105574.
- [21] Liang Xiguan, Shim J, Anderton O, et al. Low-cost data-driven estimation of indoor occupancy based on carbon dioxide (CO₂) concentration: a multi-scenario case study[J]. Journal of Building Engineering, 2024, 82: 108180.
- [22] Liu Mengyang, Chen Hong, Wei Di, et al. Nonlinear relationship between urban form and street-level PM2.5 and CO based on mobile measurements and gradient boosting decision tree models[J]. Building and Environment, 2021, 205: 108265.
- [23] Li Xueren, Sun Weijie, Qin Chao, et al. Evaluation of supervised machine learning regression models for CFD-based surrogate modelling in indoor airflow field reconstruction[J]. Building and Environment, 2025, 267: 112173.
- [24] Lewis C D. Industrial and business forecasting methods: a practical guide to exponential smoothing and curve fitting [M]. London: Butterworths Scientific, 1982.

(责任编辑 王绪迪)