

基于集成学习算法的阵列式办公街区太阳能潜力预测模型研究

李高梅¹, 何秋国², 徐桑³

摘要: 人工智能时代背景下采用机器学习算法可以实现城市街区太阳能潜力的快速预测, 然而现有研究大多采用计算机模拟的方法, 本文旨在基于机器学习算法构建城市街区太阳能潜力预测模型。本文基于武汉市真实街区案例构建办公街区参数化模型, 基于集成学习算法构建阵列式办公街区太阳能潜力预测模型。研究结果表明, 集成学习预测模型可以实现太阳能潜力的快速预测, 实现了设计方案太阳能潜力的快速反馈, 显著提升了城市街区设计方案性能评价与设计优化的效率; 屋面和立面太阳能潜力预测模型的 R^2 分别为 0.99 和 0.98, 预测模型具有较高的准确性。本研究旨在为城市办公街区规划设计提供决策依据, 提高设计科学性与合理性, 助力城市办公街区实现碳中和。

关键词: 办公街区; 机器学习算法; 集成学习; 太阳能潜力; 街区形态; 预测模型

Abstract: In the context of the artificial intelligence era, machine learning algorithms can be used to achieve rapid prediction of the solar potential of urban blocks. However, most of the existing studies used computer simulation, this paper aims to construct a prediction model of the solar potential of urban blocks based on machine learning algorithms. This paper constructs a parametric model of office blocks based on real block cases in Wuhan city, and develops an arrayed solar potential prediction model for office blocks based on ensemble learning algorithms. The results showed that the ensemble learning prediction model can achieve rapid prediction of solar potential, achieve rapid feedback of solar potential of the design scheme, and significantly improve the efficiency of the performance evaluation and design optimisation of the urban block design scheme; the R^2 of the solar potential prediction model for the roof and façade are 0.99 and 0.98, respectively, and the prediction model has a high degree of accuracy. This study aims to provide a decision-making basis for the planning and design of urban office blocks, improve the scientific and rational design, and help urban office blocks achieve carbon neutrality.

Keywords: Office block; Machine learning algorithm; Ensemble learning; Solar potential; Block morphology; Prediction model

资助项目情况: 国家自然科学基金(编号: 52378020); 中央高校基本科研业务费专项资金资助(YCJJ20230576)

¹ 李高梅, 华中科技大学建筑与城市规划学院, 湖北省城镇化工程技术研究中心, 博士研究生

² 何秋国, 华中科技大学建筑与城市规划学院, 硕士研究生

³ 徐桑(✉), 华中科技大学建筑与城市规划学院, 湖北省城镇化工程技术研究中心, 教授

引言

为了应对全球气候变化与能源危机，2020年9月，习近平主席在第七十五届联合国大会上郑重承诺，中国二氧化碳排放力争于2030年前达峰，努力争取2060年前实现碳中和^[1]。2023年中国建筑节能年度发展研究报告^[2]指出，2021年建筑运行的总商品能耗为11.1亿tec，约占全国能源消费的21%，碳排放总量为22亿tCO₂，单位面积平均建筑运行碳排放指标为32kg CO₂/m²。30·60双碳战略目标背景下，实现能源转型，采用分布式光伏这一清洁能源是关键环节^[3,4]。建筑光伏一体化可以实现光伏产能的自产自消，为街区建筑提供能源。建筑碳中和的实现需要建筑设计全流程的更新，建筑设计的目标不仅仅需要关注“坚固、实用和美观”，还需要关注设计方案的绿色建筑性能^[5]。传统的建筑方案性能评估采用“设计-模拟-优化设计-再模拟”的流程，采用计算机模拟方法的性能评估过程耗时耗力。目前在设计方案前期对于城市街区太阳能潜力的评估一般采用计算机模拟的方法展开^[6,7]，这一方法难以实现设计方案的实时预测，对于指导建筑师设计方案的修改存在时间滞后性。基于机器学习算法的建筑性能预测模型可以实现建筑方案性能的快速预测，实现了建筑方案的“即绘即模拟”，提高了建筑师的设计决策效率^[8,9]。

因此，在低碳城区建设和太阳能建筑连片化发展的大背景下，本文立足于方案设计前期的建筑师方案设计需求，以武汉市常见的阵列式（点式和板式）办公街区为研究对象，构建办公街区参数化模型，综合考虑城市气象条件、光伏组件性能和街区多尺度设计参数，提出一种基于机器学习算法的办公街区屋面和立面太阳能潜力的快速预测模型。本文旨在实现办公街区建筑方案的太阳能潜力快速预测，为建筑师提供设计辅助工具，提升决策效率。

1 研究方法

本部分提出了基于集成学习算法的阵列式办公街区太阳能潜力预测模型构建流程，如图1所示。步骤1，基于建筑类型学方法对武汉市96个阵列式办公街区进行分类，提取典型组团单元与理想街区场地构建，依托Grasshopper平台，采用Python编程语言生成办公街区参数化模型；步骤2，基于Rhino&Grasshopper参数化平台构建城市办公街区的太阳能潜力评估 workflow；步骤3，提取并计算阵列式办公街区的多尺度设计参数，包括街区-建筑-表皮三个尺度；步骤4，将办公街区的设计参数与太阳能潜力进行整合，构建预测模型的训练集与测试集；步骤5，以办公街区多尺度设计参数、气候参数、光伏组件性能为输入，以屋面和立面太阳能潜力为输出，基于集成学习算法构建办公街区太阳能潜力预测模型。



图1 基于集成学习算法的阵列式办公街区太阳能潜力预测模型构建流程（图片来源：作者自绘）

1.1 办公街区参数化模型构建

1.1.1 建筑组团典型形态提取

对武汉市真实办公街区形态进行实地调研，提取典型的建筑组团形态，以此为基础构建办公街区参数化模型。基于文献研究^[10,11]可知，办公街区形态可按建筑类型学方法分为点式、板式和围合式三种。本文以占比较大的点式和板式街区为研究对象展开研究。采用卫星地图、实地调研、街景地图和 open street map 相结合的方法对武汉市 96 个点式和板式街区进行分析，提取街区建筑组团布局、建筑面宽、进深、层数、层高等设计参数信息，确定点式和板式街区设计参数的区间范围。点式和板式街区在参数化生成逻辑方面具有一致性，因此统称为阵列式。典型的阵列式办公组团案例如图 2 所示。

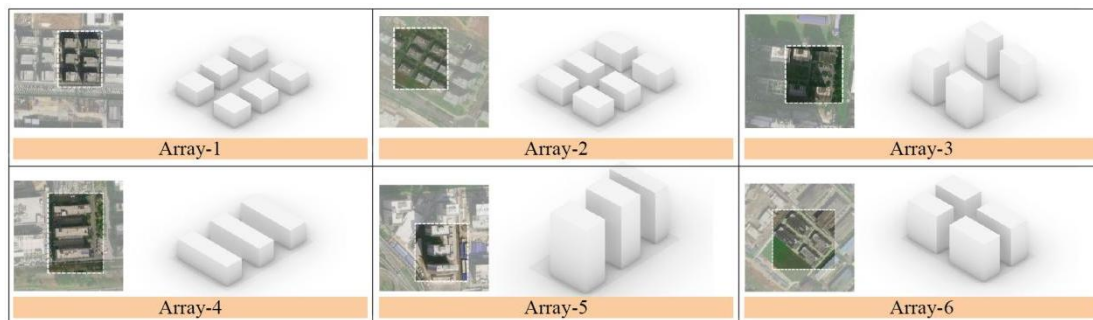


图 2 典型阵列式办公组团案例（图片来源：作者自绘）

1.1.2 办公街区参数化模型构建

办公街区太阳能潜力评估不仅需要考虑到目标建筑，还需要考虑到周边建筑对目标建筑的遮挡，因此在构建办公街区参数化模型时需要考虑研究范围及周边环境。真实街区案例的调研结果表明，办公建筑组团的场地大小尺度为 90-150m，结合建筑组团的场地趋多性与建筑组团形态变化的多样性，本文将办公建筑组团的场地设定为 145m。四个建筑组团构成理想办公街区，办公街区外围为城市支路，城市支路外侧为与研究范围形态一致的周边环境组团。阵列式办公街区参数化模型的示意图如图 3 所示。

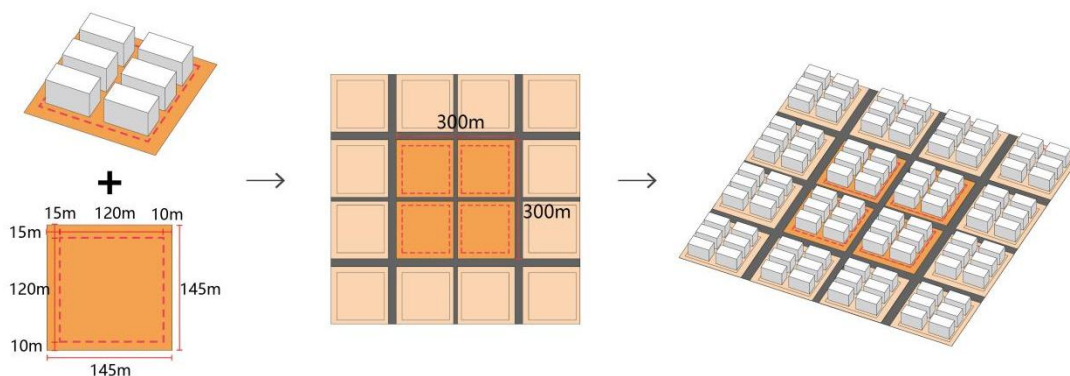


图 3 阵列式办公街区参数化模型示意图（图片来源：作者自绘）

按照建筑师的设计逻辑，办公街区参数化模型生成可以分为以下步骤：首先，根据设计任务书与设计规范确定建筑退城市道路距离，确定建筑设计范围；其次，在设计范围内布置建筑单体，基于标准层面积确定建筑数量；最后为建筑核算，即建筑的南北间距和东西间距应该符合防火规范，当不符时，建筑会自动减少一行或一列。

阵列式办公街区的建筑排布满足以下公式：

$$a + 2(d - NS) + m * BD + (m - 1)BI(NS) + b \leq L_1 \quad (1)$$

$$a + 2(d - EW) + n * BW + (n - 1) + b \leq L_2 \quad (2)$$

式中， a 和 b 为建筑退城市道路距离， $d-NS$ 和 $d-EW$ 为建筑退红线距离， BD 为建筑面宽， BW 为建筑进深， $BI(NS)$ 和 $BI(EW)$ 为建筑南北向距离和建筑东西向距离， m 和 n 为建筑行数和列数。在进行建筑核算时需要满足，当建筑高度大于等于 24m 时，需要满足建筑间距不低于 13m；当建筑高度小于 24m 时，需要满足建筑间距大于等于 6m，小于 13m。

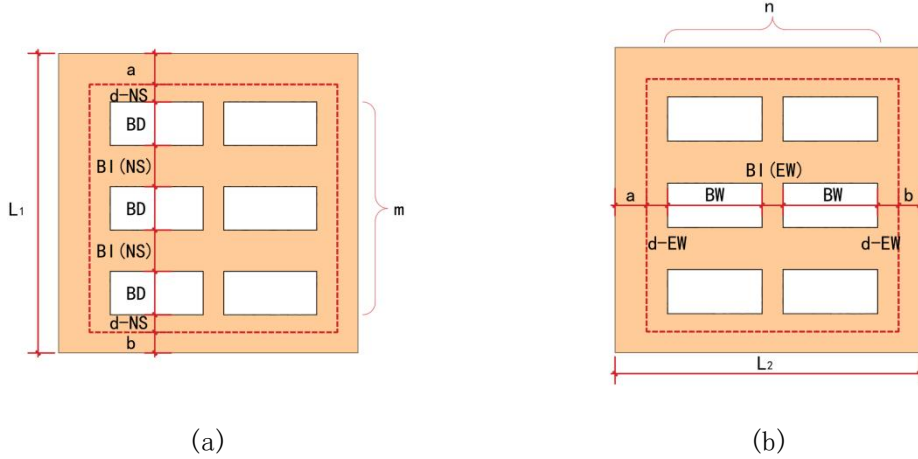


图 4 阵列式办公街区参数化模型生成限制条件（图片来源：作者自绘）

1.2 办公街区太阳能潜力评估

太阳能潜力指的是安装在建筑表面的光伏板将太阳辐射转变为电能的过程。本文基于本课题组街区太阳能潜力评估的相关成果^[6,12]，将城市街区太阳能潜力划分为三个层级：辐射潜力、安装潜力和技术潜力。辐射潜力指的是建筑表面接收到的太阳辐射有效太阳辐射量，它与辐射阈值相关，辐射阈值指的是安装在建筑表面的光伏板在全生命周期内满足光伏系统投入产出平衡的最小辐射值。太阳辐射可以通过基于 Grasshopper 参数化平台调用 Ladybug Tools 的街区太阳能评估工具来实现，该工具已经被 Shen Xu 等学者^[13]证明具有很高的准确性。安装潜力指的是建筑表面适合安装光伏组件的面积，这一指标与街区建筑表面的可安装面积以及辐射阈值密切相关，安装因子通过对真实街区案例的调查获取。技术潜力指的是安装在街区建筑表面的光伏组件的年发电量，它受辐射潜力、安装潜力、光伏组件光电转化率、综合效率、衰减率和生命周期的综合影响。本文选择市场占有率最高的单晶硅光伏组件进行分析，光伏组件生命周期设定为 25 年，辐射阈值为 568 kWh/m²/y。太阳辐射模拟的参数设置如表 1 所示。

表 1 太阳辐射模拟的相关参数设置

相关参数	参数设定
气象参数	CHN_Hubei.Wuhan.574940_CSWD
模拟周期	1月1日 00:00 至 12月31日 24:00
网格大小	1 m * 1m
建筑表面反射率	0.2
地面反射率	0.2

办公街区太阳能潜力指的是技术潜力，计算公式如下：

$$E_p = H_A \times A_{pv} \times \eta_{sys} \times K \times \sum_{i=1}^N (1 - R_1 - (N-1) \times R_d) \quad (3)$$

式中, E_p 指的是技术潜力, H_A 指的是建筑表面接收到的累计太阳辐射, A_{pv} 指的是掩的光伏组件的可安装面积, η_{sys} 指的是光伏组件的光电转化率, 取值为 21.86%, K 指的是综合效率系数, 取值为 85%, N 指的是光伏组件的生命周期, 取值为 25 年, R_1 和 R_d 指的是光伏组件第一年和之后的衰减率, 取值为 1.5% 和 0.48%。为了使光伏产能与建筑能耗相对应, 本文的太阳能潜力 (技术潜力) 指的是单位建筑面积的光伏发电量。

1.3 办公街区设计参数计算

为了与城市街区设计流程“草图设计-总体平面设计-单体方案初步设计-单体方案完善设计”相对应, 本文选取了与设计流程相关的, 描述建设强度、场地布置、平面形态、建筑高度、建筑形体和立面设计的设计参数, 囊括街区-建筑单体-建筑表皮三个尺度, 包括建筑密度 (BCR)、容积率 (FAR)、体形系数 (BSF)、开放空间率 (OSR)、街区朝向 (O)、屋面立面面积比 (R/F)、建筑基底面积 (GBA)、建筑南北间距 (BI(NS))、建筑东西间距 (BI(EW))、标准层面积 (S)、建筑进深 (BD)、建筑层数 (FN)、建筑层高 (FH)、各向窗墙比 (WWR-N、WWR-S、WWR-E、WWR-W)。其中, 开放空间率的定义为办公街区的开放空间面积与总建筑面积的比值。

1.4 办公街区数据集构建

办公街区数据集包括设计参数数据集与太阳能潜力数据集两部分。通过控制性参数生成办公街区参数化模型, 然后对参数化模型进行模拟获取太阳能潜力, 包括屋面太阳能潜力和立面太阳能潜力。设计参数包括控制性参数与描述性参数, 在 17 个设计参数中, 街区朝向、建筑南北间距、建筑东西间距、标准层面积、建筑进深、建筑层数、建筑层高、各向窗墙比为控制性参数, 其余为描述性参数。通过对武汉市阵列式办公街区真实案例进行调研分析, 确定各控制性参数的阈值范围和步长变化, 采用拉丁超立方抽样获取 500 组控制性参数数据, 输入到参数化模型中, 生成相应的 500 个参数化模型, 基于 Grasshopper 参数化平台对太阳辐射进行模拟, 获取屋面和立面的太阳能潜力数据。除此之外, 对 500 各办公街区参数化模型进行描述性参数计算, 将描述性参数与控制性参数进行整合, 生成设计参数数据集。将设计参数数据集和太阳能潜力数据集进行整合构建办公街区数据集。控制性参数的阈值范围和步长变化如表 2 所示。

表 2 阵列式办公街区控制性参数的阈值范围和步长变化

参数类型	阈值范围	步长	参数类型	阈值范围	步长
街区朝向 (O)	-90°— 90°	5°	建筑层高 FH	3.8~5.4m	0.1m
南北建筑间距 BI(NS)	7-52m	1m	东向窗墙比 WWR-E	0.2-0.9	0.1
东西建筑间距 BI(EW)	6-21m	1m	南向窗墙比 WWR-S	0.2-0.9	0.1
标准层面积 GBA	400-2500 m ²	100 m ²	西向窗墙比 WWR-W	0.2-0.9	0.1
建筑进深 BD	14-34m	1m	北向窗墙比 WWR-N	0.2-0.9	0.1
建筑层数 FN	3-25 层	1 层			

1.5 办公街区太阳能潜力预测模型构建

集成学习算法意味着采用多个基础学习器来完成学习任务, 它具有准确性高、过拟合风险低和鲁棒性好的优点, 本文采用 6 种集成学习算法来构建阵列式办公街区太阳能潜力预测模型, 包

括套袋算法 (Bagging algorithm)、超随机树算法 (Extremely randomized trees algorithm)、随机森林算法 (Random forest algorithm)、梯度提升算法 (Gradient Boosting algorithm)、自适应增强算法 (AdaBoosting algorithm) 和梯度提升决策树算法 (XGBoosting algorithm)。套袋算法是一种自适应技术, 通过引入 Bootstrap 抽样方法生成多个子样本集, 并基于每个子样本集构建一个基模型。它的关键优点在于它可以降低模型的过拟合现象。通过集成多个基模型, Bagging 算法能够更好地捕捉到数据的复杂性和不确定性, 从而降低模型的过拟合风险。超随机数算法是一种集成学习方法, 通过构建多棵极度随机化的决策树来提高模型的泛化能力和稳定性。随机森林算法是一种基于决策树的集成学习方法, 它通过将多个决策树组合在一起, 来提高模型的性能。随机森林的核心原理是通过生成多个独立的决策树, 并在训练数据上进行平均, 来减少过拟合。梯度提升算法是一种基于凸优化的集成学习方法, 它通过将多个弱学习器组合在一起, 来提高模型的性能。梯度提升的核心原理是通过在每个弱学习器上进行梯度下降, 来逐步优化模型。自适应增强算法是将弱分类器迭代成强分类器的过程, 充分考虑每个分类器的权重, 每个迭代过程中改变样本权重和分类器权重, 最终结果是每个分类器的权重和。梯度提升决策树算法是经过优化的分布式梯度提升库, 旨在高效、灵活且可移植。

将数据集按照 70% 和 30% 的比例划分为训练集和测试集。将数据集输入到 6 种集成学习算法中进行训练, 采用 R^2 和 MAE 指标进行评价, 选择评价指标好的算法来构建办公街区太阳能潜力预测模型。

2 研究结果

2.1 样本数据集特征

对拉丁超立方抽样的 500 组数据进行参数化模型构建和太阳辐射模拟 (包括屋面和立面), 排除异常结果之后, 有效数据为 465 个, 将其按照 70% 和 30% 的比例划分为训练集和测试集。样本被用来输入到集成学习算法中训练太阳能潜力预测模型。

图 5-6 描述了训练集和测试集的代表性设计参数 (建筑密度、容积率、体形系数、标准层面积、层数和南向窗墙比) 的分布特征, 结果表明设计参数分布均匀, 它证明了数据集划分的合理性和随机性。

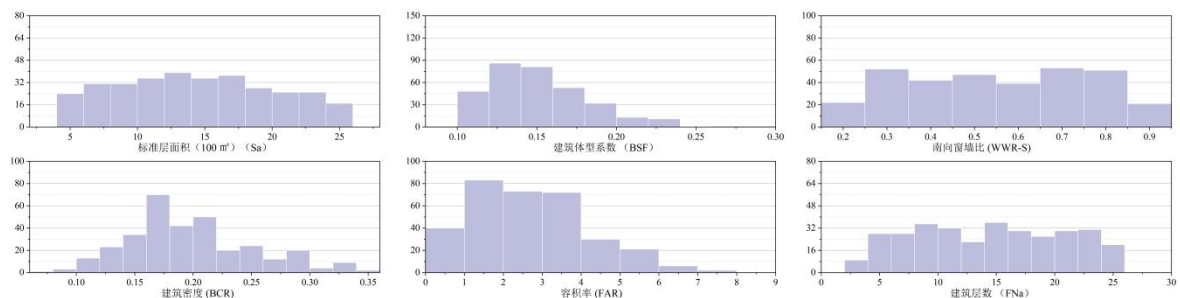


图 5 训练集代表性设计参数的分布特征 (图片来源: 作者自绘)

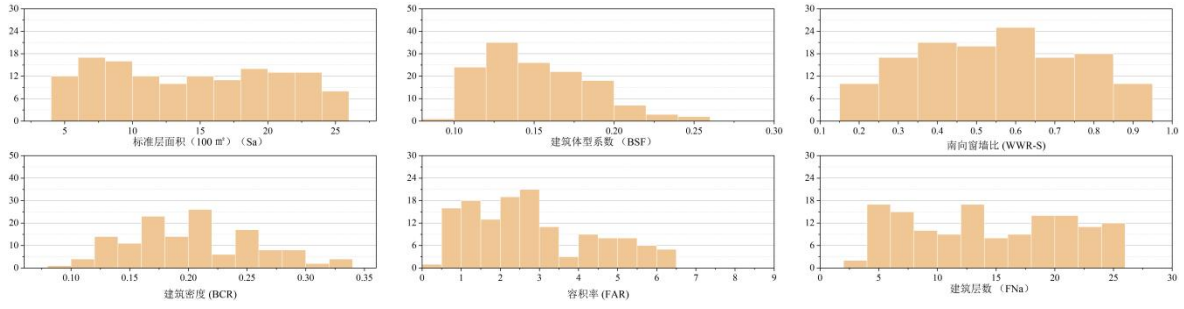


图 6 测试集代表性设计参数的分布特征 (图片来源: 作者自绘)

2.2 样本数据集分析

将 17 个阵列式办公街区多尺度设计参数与屋面太阳能潜力和立面太阳能潜力分别进行相关性分析, Pearson 相关性分析结果见表 3。

表 3 影响因素相关性分析

形态参数	EGI-R 相关系数 (r)	显著性 Sig.	形态参数	EGI-F 相关系数 (r)	显著性 Sig.
BCR	0	0.992	BCR	-0.343**	0.000
FAR	-0.721**	0.000	FAR	-0.608**	0.000
BSF	0.475**	0.000	BSF	0.369**	0.000
OSR	0.827**	0.000	OSR	0.629**	0.000
R/F	0.925**	0.000	R/F	0.472**	0.000
GBA	0	0.992	GBA	-0.343**	0.000
O	0.086	0.063	O	-0.151**	0.001
BI (NS)	0.026	0.581	BI (NS)	0.237**	0.000
BI (EW)	0.031	0.506	BI (EW)	0.093*	0.045
S	-0.038	0.415	S	-0.102*	0.028
BD	0.033	0.479	BD	-0.084	0.072
FN	-0.857**	0.000	FN	-0.549**	0.000
FH	-0.043	0.355	FH	0.009	0.849
WWR-N	0.009	0.847	WWR-N	0.004	0.938
WWR-W	-0.016	0.732	WWR-W	-0.289**	0.000
WWR-S	0.051	0.271	WWR-S	-0.323**	0.000
WWR-E	0.008	0.872	WWR-E	0.052	0.265

其中, 与屋面太阳能潜力相关性最大的设计参数为屋面立面面积比 (R/F), 呈正相关, 这是因为当屋面立面面积比越大时, 说明屋面面积越大, 即屋面太阳能潜力越大。其次是建筑层数, 呈负相关, 然后是开放空间率 (正相关)、容积率 (负相关)、体形系数 (正相关)。对于立面太阳能潜力, 相关性最大的设计参数为开放空间率 (0.629), 这是因为当开放空间率越大时, 说明建筑间距越大, 建筑之间的相互遮挡越小, 立面太阳能潜力越大。紧随其后的是容积率 (-0.608)、建筑层数 (-0.549)、屋面立面面积比 (0.472)、体形系数 (0.369)、建筑密度 (-0.343)、建筑基底面积 (-0.343)、南向窗墙比 (WWR-S)、西向窗墙比 (-0.289)。当这些设计参数与太

阳能潜力呈正相关时，随着这些参数变大，太阳能潜力也变大，反之，当与太阳能潜力呈负相关时，随着这些参数变大，太阳能潜力减小。

2.3 阵列式办公街区集成学习太阳能潜力预测模型

为了消除设计参数的量纲差异对太阳能潜力的影响，对数据集进行归一化处理。将处理后的数据集按照 70%、30%的比例划分为训练集和测试集，将训练集输入到调试好的集成学习算法中进行训练，以拟合优度 (R^2) 和均方误差 (MSE) 作为评价指标，将训练好的模型以测试集检测其泛化性，训练集和测试集的 R^2 和 MSE 同时最优的算法模型被筛选出来进行太阳能潜力预测。6 种集成学习算法的 R^2 和 MSE 指标见表 4 和 5，选择评分最高的算法（随机森林、梯度提升）作为屋面和立面太阳能潜力的预测模型。

表 4 屋面太阳能潜力预测模型的评分结果对比分析

算法类型	R^2		MSE	
	训练集	测试集	训练集	测试集
套袋	0.99	0.99	0.01	0.04
超随机树	1.0	0.99	7.57	0.04
随机森林	0.99	0.99	0.01	0.03
梯度提升	0.99	0.99	0.01	0.04
自适应增强	0.99	0.99	0.21	0.18
梯度提升决策树	0.99	0.99	7.05	0.04

表 5 立面太阳能潜力预测模型的评分结果对比分析

算法类型	R^2		MSE	
	训练集	测试集	训练集	测试集
套袋	0.96	0.66	0.16	0.91
超随机树	1.0	0.71	2.12	0.76
随机森林	0.97	0.70	0.13	0.79
梯度提升	0.98	0.74	0.08	0.68
自适应增强	0.84	0.55	0.64	1.19
梯度提升决策树	0.99	0.70	8.11	0.79

结语

本文提出了一种基于集成学习算法的阵列办公街区太阳能潜力（屋面和立面）预测模型。本研究考虑了城市的气象条件、光伏组件类型和多尺度设计参数来进行单位建筑面积光伏发电量的预测。对 6 种集成学习算法的 R^2 和 MSE 进行对比分析，选择评分最高的算法模型进行办公街区太阳能潜力的预测。与现有的校园宿舍街区基于街区形态参数构建的太阳能潜力的线性预测模型^[14] ($R^2=0.73$) 相比，本文所构建的机器学习预测模型具有较高的精度与准确性，这可能是由于机器学习算法的非线性特点相比于线性模型更有优势。与上海市城市不同功能街区的集成学习算法预测模型^[15] 相比（梯度提升决策树算法， $R^2=0.91$ ），本文所构建的预测模型准确性更好，这可能是由于本文所选择的街区形态参数更能反映太阳能潜力的特征。

未来该模型与多目标优化进行结合可以实现太阳能潜力性能驱动的办公街区设计方案生成。本研究旨在为建筑师进行办公街区太阳能规划设计、零碳街区建设提供有益的启示和参考。

参考文献

- [1] 中华人民共和国商务部. 习近平在第七十五届联合国大会一般性辩论上的讲话（全文）. 2020. Available from: <https://m.mofcom.gov.cn/article/i/jyj/m/202009/20200903003397.shtml>
- [2] 清华大学建筑节能研究中心. 中国建筑节能年度发展研究报告 2023（城市能源系统专题）[M]. 北京: 中国建筑工业出版社, 2023.
- [3] 梅洪元,陈玉婷,史小蕾,等. 寒区建筑碳中和的瓶颈问题与应对策略 [J]. 中国科学基金, 2023,37(03):364-371.
- [4] J. Tian, S. Xu, A morphology-based evaluation on block-scale solar potential for residential area in central China [J]. Solar Energy, 221 (2021) 332-347.
- [5] 朱姝妍,马辰龙. 基于“知情式”设计理念的建筑性能优化方法研究 [J]. 世界建筑, 2022, (03): 84-91.
- [6] 邓丰,谭洪卫. 以近零能耗为导向的上海地区不同住宅类型的能耗及光伏替代率研究 [J]. 建筑科学, 2021,37(04):9-18.
- [7] 朱丽,景一帆,尹宝泉,等. 基于参数化分析的校园光伏潜力评估方法及其应用 [J]. 建筑节能(中英文), 2023,51(03):58-65+138.
- [8] 黄茜,曲大刚,孙澄,等.基于深度学习的自然采光办公空间视觉舒适度预测模型建构[J].建筑学报,2023,(10):50-54.
- [9] 朱姝妍,马辰龙,向科,等.面向方案阶段能耗主导的建筑性能快速优化方法[J].建筑师, 2021,(06):69-76.
- [10] 华东建筑集团股份有限公司. 建筑设计资料集(第三分册) [M]. 北京: 中国建筑工业出版社, 2017.
- [11] Gaomei Li, Cheng Sun, Mengcheng Sang, et al. The synergistic effect of multiple design factors on building energy consumption of office blocks: A case study of Wuhan, China [J]. Journal of Building Engineering, 2024(95), 110200.
- [12] S. Xu, H. Jiang, F. Xiong, C. Zhang, M. Xie, Z. Li, Evaluation for block-scale solar energy potential of industrial block and optimization of application strategies : A case study of Wuhan , China [J]. Sustainable Cities and Society, 2021(72).
- [13] Wei Liao, Yeonsook Heo, Shen Xu. Simplified vector-based model tailored for urban-scale prediction of solar irradiance [J]. Solar Energy, 2019 (188), 483.
- [14] M. Xie, M. Wang, H. Zhong, et al. The impact of urban morphology on the building energy consumption and solar energy generation potential of university dormitory blocks [J], Sustainable Cities and Society, 2023 (96), 104644.
- [15] Q. Jiang, C. Huang, Z. Wu, et al. Predicting building energy consumption in urban neighborhoods using machine learning algorithms [J]. Frontiers of Urban and Rural Planning, 2024, 2(6): 1-21.