

DOI: 10.7672/sgjs2025080066

# 基于 BIM-GIS 与机器学习的区域尺度绿色 建筑能耗预测研究\*

杨超杰<sup>1</sup>, 刘琦娟<sup>2</sup>, 霍慧秀<sup>1</sup>, 徐照<sup>1</sup>

(1. 东南大学土木工程学院, 江苏 南京 211189;

2. 深圳顺丰泰森控股(集团)有限公司, 广东 深圳 518000)

[摘要] 建筑节能是实现碳达峰与碳中和目标的关键环节, 而区域尺度建筑能耗研究的重要性随着城市化的发展日益凸显。为此, 构建了一个基于机器学习的区域尺度绿色建筑能耗预测框架。通过 FME 数据转换技术实现从 IFC 到 CityGML 的转换, 集成 BIM 和 GIS, 利用 BIM 的丰富语义信息弥补其在区域尺度时空分析中的不足, 并应用分类算法对建筑类型进行分类以提高预测准确性。通过实例验证, 评估的 3 种机器学习回归模型均能较准确地预测建筑能耗, 且速度显著优于传统能耗模拟方法。该研究为区域尺度绿色建筑能耗预测提供了实用思路, 有助于推动绿色建筑可持续发展。

[关键词] 绿色建筑; 能耗; 建筑信息模型; 地理信息系统; 机器学习; 区域尺度

[中图分类号] TU17

[文献标识码] A

[文章编号] 2097-0897(2025)08-0066-08

## Energy Consumption Prediction of Regional Scale Green Buildings Based on BIM-GIS and Machine Learning

YANG Chaojie<sup>1</sup>, LIU Qijuan<sup>2</sup>, HUO Huixiu<sup>1</sup>, XU Zhao<sup>1</sup>

(1. School of Civil Engineering, Southeast University, Nanjing, Jiangsu 211189, China;

2. Shenzhen S. F. Taisen Holding Co., Ltd., Shenzhen, Guangdong 518000, China)

**Abstract:** Building energy conservation is the key link to achieve the goal of carbon peak and carbon neutralization, and the importance of regional scale building energy consumption research is becoming more and more prominent with the development of urbanization. To this end, a machine learning-based regional-scale green building energy consumption prediction framework is constructed. The conversion from IFC to CityGML is realized by FME data conversion technology, and BIM and GIS are integrated. The rich semantic information of BIM is used to make up for its shortcomings in regional scale spatio-temporal analysis, and the classification algorithm is applied to classify the building types to improve the prediction accuracy. Through example verification, the three machine learning regression models evaluated can accurately predict building energy consumption, and the speed is significantly better than the traditional energy consumption simulation method. This study provides a practical idea for the prediction of energy consumption of green buildings at the regional scale, which is helpful to promote the sustainable development of green buildings.

**Keywords:** green buildings; energy consumption; building information modeling (BIM); geographic information system (GIS); machine learning; regional scale

### 0 引言

中国建筑领域的能源管理是节能减排工作的重点。随着城市数量、规模和人口的持续增长, 区域尺度建筑节能减排研究的重要性日益凸显。区域尺度建筑能耗模拟在节能低碳城区规划设计、旧

\* 江苏省科技计划专项资金(重点研发计划社会发展)项目 (BE2022820)

[作者简介] 杨超杰, 硕士研究生, E-mail: 220221453@seu.edu.cn

[通信作者] 徐照, 教授, E-mail: xuzhao@seu.edu.cn

[收稿日期] 2025-01-20

城节能改造、城市能效管理及区域供暖与制冷系统设计等方面具有重要的实际应用价值。

建筑信息模型 (building information modeling, BIM) 是一种运用数字化技术构建的三维建筑模型,是获取建筑内部细节的关键途径,通过 IFC (industry foundation classes) 实现不同软件平台和应用程序间的信息交换和共享,在单体层面的绿色建筑能耗研究中发挥着重要作用。然而其高精度建模和丰富语义信息的优势并不完全适用于区域尺度绿色建筑能耗研究,数据复杂性及高计算性能要求等问题严重影响了预测的成本、效率与精确度。同时, BIM 的空间分析能力相对较弱,使其在区域尺度建筑能耗研究中的应用受限。

地理信息系统 (geographic information system, GIS) 是集地理信息采集、存储、管理、计算、分析、展示及描述于一体的技术系统,使用 CityGML (city geography markup language) 作为基础数据格式。以 GIS 为基础的地理空间技术的进步,促进了对于区域尺度绿色建筑能耗预测的相关研究。刘晓君等<sup>[1]</sup>引入 GIS 技术呈现省区建筑能耗时空分布情况并进行对比分析。Ma 等<sup>[2]</sup>将 GIS 技术与数据挖掘方法相结合,对纽约市 3 640 栋多户住宅楼的能源使用强度进行建模,根据利用机器学习算法估算的能源使用强度,结合地理空间技术生成全市能耗情况。Davila 等<sup>[3]</sup>使用 EnergyPlus 模型模拟每个建筑物单体的能源使用强度,并结合地理空间技术计算最终的建筑能耗并将其升级至城市水平。

计算机科技的迅猛进步,尤其是机器学习 (machine learning, ML) 领域的显著发展,为解决前述数据复杂性、时间跨度处理需求及高计算性能要求等问题提供了新的解决途径。于军琪等<sup>[4]</sup>提出树种算法优化的 RBF 神经网络结合 LSTM 神经网络的混合预测模型 TSA-RBF-LSTM,用于建筑逐时电力能耗预测。滕文龙等<sup>[5]</sup>使用基于能耗监管平台输出的历史能耗数据训练的 BP 神经网络与 MEA-BP 神经网络模型对建筑能耗进行预测。Dong 等<sup>[6]</sup>提出一种基于集成学习和能耗模式分类的建筑能耗预测策略。

GIS 模型拥有强大的空间分析能力,但其语义信息相对较少,通过 BIM, GIS 的有效集成,可更好地实现对区域尺度绿色建筑能耗的大规模预测。因此,本研究构建了一个框架,在数据获取阶段利用 FME (feature manipulate engine) 实现 IFC 到 CityGML 格式的转换以集成 BIM 和 GIS;考虑到建筑使用模式对能源消耗需求的显著影响因建筑类

型而异,基于分类算法实现区域绿色建筑中住宅和非住宅建筑的分类,并针对不同建筑类型设置不同模拟参数;在此基础上,利用机器学习方法构建回归模型,对建筑能耗进行预测,并通过实例验证该框架的可行性。

## 1 基础数据集成与获取

区域尺度绿色建筑能耗预测所需基础数据包括天气数据、几何数据和非几何数据。

### 1.1 天气数据获取

采用典型气象年 (typical meteorological year, TMY) 天气数据和未来天气数据作为天气数据源。

TMY 天气数据的获取采用直接获取法<sup>[7]</sup>。使用 CBE Clima Tool<sup>[8]</sup> 获取 EPW (energy plus weather) 格式数据,数据源采用中国建筑热环境分析所依据的专用气象数据集 CSWD (Chinese standard weather data)<sup>[9]</sup>。

未来天气数据的获取采用 Morphing 方法<sup>[10]</sup>,以 NASA 提供的 NEX-GDDP-CMIP6<sup>[11]</sup> 数据集为基础,以当前的天气数据为输入,通过数学方法将已有的气象参数时间序列与气候变化进行融合,生成全新的气象参数文件。

### 1.2 IFC 向 CityGML 转换

分析 IFC 与 CityGML 的差异性,IFC 向 CityGML 的转换过程如图 1 所示。

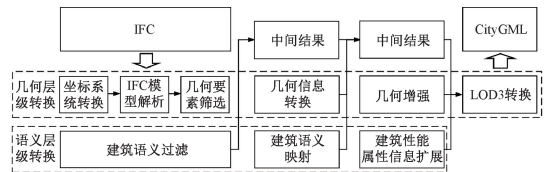


图 1 IFC 向 CityGML 的转换过程

Fig. 1 Conversion process from IFC to CityGML

CityGML 通过多层次细节 (LOD) 的方法对几何模型进行分级描述。本研究旨在建立区域绿色建筑能耗预测模型。LOD4 虽然包含详尽的信息,但会增加计算成本;较低等级的 LOD 更适合不需要详细细节的模型或研究规模较大的模型<sup>[12]</sup>; LOD2 对建筑外立面开口的描述信息不足,因此本研究重点实现 LOD3 层次的转换。

#### 1.2.1 几何层级转换

##### 1.2.1.1 坐标系统转换

FME 软件中定义了与坐标转换相关的转换器,如 LocalCoordinateSystemSetter, CsmmapReprojector, CoordinateSystemDescriptionConverter 等。通过其参数设定,可实现 IFC 相对坐标系统坐标向 CityGML 绝对坐标系统坐标的转换,确保转换过程中坐标的

准确性和一致性。

### 1.2.1.2 IFC 模型解析

利用 FME 软件对 IFC 模型进行深入解析,提取其要素内容。

### 1.2.1.3 几何重构

在几何信息转换前,应根据 LOD3 层级的要求筛选几何要素。在此过程中,可以忽略那些主要用于表示要素间关系或定义属性但不包含直接几何信息的类。几何信息转换的主要工作是将 IFC 模型中的扫描实体(sweep)和构造实体几何(CSG)表达方式,转换为 CityGML 常用的边界表示(B-rep)几何表达方式。

扫描实体转换为边界表示的过程,可通过扫描体横截面的坐标和扫描线(扫描方向)生成边界表达的几何<sup>[13]</sup>;而构造实体几何的转换,则需从 shape 到 point 逐层遍历,提取点坐标及其关联关系,最终绘制出表面模型。

对于复杂实体,可以使用 FME 及其中内置的转换器构建几何实体转换为边界表示的流程。通过转换器的搭配和参数设置,可以构建自定义转换器 ConvertGeometry,并预先设定标准化逻辑,实现对各类几何实体转换为表面几何的重复性操作。转换流程如图 2 所示。

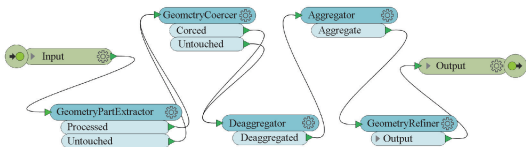


图 2 FME 中自定义几何转换器 ConvertGeometry  
Fig. 2 Custom geometry converter ConvertGeometry in FME

转换完成后,应用 AttributeRenamer 转换器设置统一标识符, CityGMLGeometrySetter 转换器设置预输出 CityGML 模型,输出为 CityGML 要素。

经过转换的模型中各要素间的对应关系发生改变,为保持数据准确性和逻辑严密性,需参照原有对应关系,对转换后模型中的要素关系进行重新梳理和确立。

### 1.2.1.4 CityGML 模型 LOD 转换

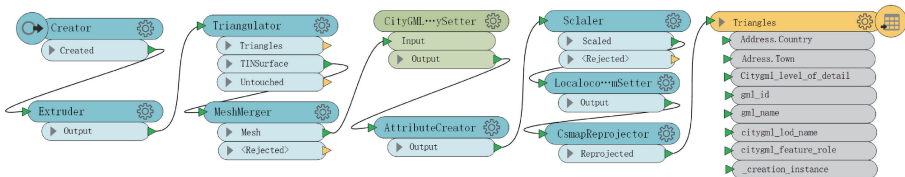


图 3 FME 中 LOD3 几何层级转换  
Fig. 3 LOD3 geometric level conversion in FME

转换采用增量算法,从最简单的 LOD 逐步细化至 LOD3 层次,同时编辑和关联几何要素属性。在 FME 中 LOD3 几何层级转换流程如图 3 所示。

### 1.2.2 语义层级转换

#### 1.2.2.1 建筑语义过滤

在 FME 软件中使用 FeatureTypeFilter 过滤器进行建筑语义过滤。对 IFC 模型中的每个建筑构件 IfcObject 进行三重判断的流程如图 4 所示。

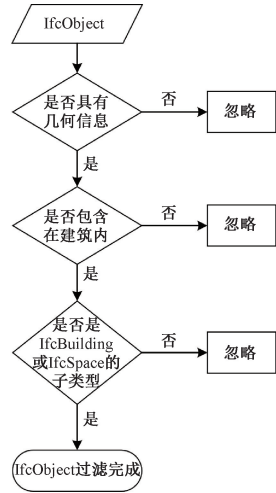


图 4 语义过滤流程  
Fig. 4 Semantic filtering process

#### 1.2.2.2 语义映射

根据映射对象特性,常见的 IFC 和 CityGML 的语义映射对应如表 1 所示。除将 IFC 类映射到 CityGML 实体外,还需确保属性与关系的正确映射。利用 FME 实现属性映射的流程如下:①解析 IFC 模型获取其包含的属性与对应属性值,应用 AttributeRenamer 转换器对应 IFC 模型 ID 与 CityGML 模型 ID;②应用自定义转换器 GetParentID,设置 IFC 模型与 CityGML 模型的父类 ID 逐一对应,转换流程如图 5 所示;③以 CityGML 的标准对要素属性进行重新组织和整理,将其纳入 CityGML 模型中;④设置预输出 CityGML 模型,应用 AttributeCreator 转换器给要素添加属性;应用 GeometryPropertySetter 转换器根据要素属性设置相应的几何名称和特征。属性的转换可与几何模型

转换同步进行,并最终汇集至 CityGML 模型的输出模块。

表 1 常见 IFC 和 CityGML 的语义映射对应关系

Table 1 Common semantic mapping correspondence between IFC and CityGML

IFC 实体	CityGML 语义	映射方式	
IfcBuilding	Buiding		
IfcAnnotation	Annotation	一对一	IFC 直接映射到 CityGML
IfcDoor	Door	一对一	映射
IfcRoof	RoofSurface		
IfcSlab	OuterFloorSurface/ WallSurface/ OutCeilingSurface	一对多	1 个 IFC 类可以映射到多个 CityGML 类
IfcBeam/IfcPile/ IfcRamp/IfcColumn/ IfcStair/IfcRailing	Building Installation	多对一	多个 IFC 类可以映射到单个 CityGML 实体
IfcBuildingStorey			IFC 信息无法通过直接映射转化为 CityGML 信息,而需经过一定的运算才能实现
IfcSpace	Room	间接映射	
IfcCovering			

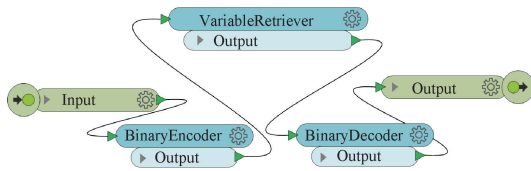


图 5 FME 中自定义几何转换器 GetParentID

Fig. 5 Custom geometry converter GetParentID in FME

1.2.2.3 建筑性能属性信息扩展

对于语义映射后仍缺失的与建筑能耗相关的部分材质属性信息,采用 UML 建模方式实现扩展。流程为:①根据 CityGML 标准通过 Enterprise Architect 平台设计 UML 类图,扩展了原 CityGML 标准中表面类(如墙面、屋顶表面等)的参数信息,定义围护结构中的新属性;②创建 ADE 模型,依据 UML 类图编写 XML Schema 定义(XSD)文件,定义新的元素和类型并确保新的类和属性符合 CityGML 的命名空间和结构要求;③从 UML 类图自动生成包含所有必要定义的 XML Schema 文件,以便在 CityGML 文档中使用新的元素和类型;④更新 CityGML 的 cityGML.xsd 文件,将生成的 XML Schema 文件集成至 CityGML 模型中,确保 CityGML 解析器和应用程序能识别和使用新的 ADE 元素。

1.3 建筑基础数据的获取与修正

1.3.1 几何数据获取与修正

1.3.1.1 建筑轮廓数据获取

建筑轮廓数据通过 LocaSpace Viewer 内置的测量工具从 Google Earth 卫星图像<sup>[14]</sup>中提取,包括轮廓面积、周长、长和宽等参数,如图 6 所示。



图 6 建筑轮廓数据测取

Fig. 6 Measurement of building contour data

1.3.1.2 建筑轮廓数据修正

为在保持建筑基本特征不变的前提下,简化模型结构并提升模拟的效率和准确性,对原始 GIS 文件中的建筑轮廓进行优化处理。构建一个抽象的长方体作为替代模型,并确保原有建筑的轮廓长度、轮廓面积及层数 3 个决定体形系数唯一性的核心参数不变。转化规则如式(1),(2)所示。

$$LS_i = \frac{C_i}{4} + \left( \frac{C_i^2}{16} - S_i \right)^{0.5} \quad (1)$$

$$SS_i = \frac{C_i}{2} - LS_i \quad (2)$$

式中:  $LS_i$  为长方形长边边长;  $SS_i$  为长方形短边边长;  $C_i$  为建筑轮廓长;  $S_i$  为建筑轮廓面积。

1.3.1.3 建筑高度数据获取

建筑高度数据采用基于 Google Earth 卫星影像的阴影线方法<sup>[14]</sup>提取,步骤如下:①在 LocaSpace Viewer 平台上测量建筑垂直边长和阴影边长,记录建筑物地理位置信息(包括经度和纬度)、拍摄时刻太阳方位角和卫星方位角等核心参数,如图 7 所示。综合考虑影像拍摄时间,根据式(3)反推出建筑高度。②根据理论假设<sup>[14]</sup>,在同一卫星影像的特定区域内,建筑物高度与其影像垂直投影长度之比可视作常数。通过计算该比值的平均数,构建建筑高度推算模型,进而获取高度数据。

$$H = \tanh_s [ L_{A_2B} \cos(\alpha_s - \alpha_{SA}) + \sqrt{L_{A_1A_2}^2 - L_{A_2B}^2 \sin^2(\alpha_s - \alpha_{SA})} ] \quad (3)$$

式中:  $H$  为建筑高度;  $h_s$  为太阳高度角;  $L_{A_1A_2}$  为建筑檐口至其投影点的长度;  $L_{A_2B}$  为建筑的垂直边长;  $\alpha_s$  为太阳方位角;  $\alpha_{SA}$  为卫星方位角。

1.3.2 建筑非几何数据获取

建筑非几何数据根据国家和工业设计标准制



图 7 反推建筑高度所需数据测取

Fig. 7 Data acquisition required for reverse calculation of building height

定。以实例所在的江苏省南京市为例,住宅建筑的建筑特征依据 JGJ 134—2010《夏热冬冷地区居住建筑节能设计标准》描述,非住宅建筑特征依据 GB 50189—2015《公共建筑节能设计标准》描述,特征设置如表 2 所示。

表 2 住宅和非住宅建筑的详细建筑特征

Table 2 Detailed architectural features of residential and non-residential buildings

特征	住宅建筑	非住宅建筑
建筑物包括的热物理性能		
墙	0.83	0.5
屋顶	0.8	0.69
板	1.31	0.7
窗户(传热系数或太阳能得热系数)	2.67/0.34	2.50/0.34
渗透换气次数 ACH/(次·h <sup>-1</sup> )	1	0.15
暖通空调新鲜空气供应/(m <sup>3</sup> ·(s·人)) <sup>-1</sup>	0	0.008
HVAC 采暖或制冷设置点/℃	18/26	20/26
系统 采暖或制冷能源效率比	1.9/2.3	0.9/4.8
内部荷载		
居住密度/(人·m <sup>-2</sup> )	0.03	0.1
设备密度/(W·m <sup>-2</sup> )	4.3	15
照明密度/(W·m <sup>-2</sup> )	6	9

## 2 基于分类算法的建筑类型分类

### 2.1 数据集初始化

本研究实例位于南京市江宁区,面积约 12.5 km<sup>2</sup>。该区域的卫星影像图和通过卫星影像图获取的建筑足迹分别如图 8,9 所示。



图 8 卫星影像

Fig. 8 Satellite images

采用 GeoPandas 对数据集进行初始化操作。对建筑能耗字段数据进行筛选,过滤极端值的数据流程如下:①导入 Pandas, geopandas, matplotlib, pyplot



图 9 建筑足迹

Fig. 9 Building footprints

等第三方库并读取建筑相关地理数据集;②添加相关分析指标列,如表 3 所示;③通过设定面积>5m<sup>2</sup>及体形系数<1.5 的筛选条件剔除不符合要求的极端数据。筛选后研究区域的数据量从原先的 3 349 条减少至 3 134 条。

表 3 数据集相关分析指标列

Table 3 Correlation analysis index column of data set

指标名称	指标对应参数名称	计算方法
建筑占地面积	area_floor	直接获得
建筑层数	floor	直接获得
建筑几何形状	geometry	直接获得
建筑体积	volume	建筑占地面积×建筑高度
周长面积比	perimeter_area_ratio	建筑周长/建筑面积
建筑周长	perimeter	(建筑长+建筑宽)×2
建筑高度	height	直接获得
建筑表面积	surface_area	建筑周长×建筑高度
体形系数	shape_factor	建筑表面积/建筑体积
建筑面积	costruction_area	建筑占地面积×建筑层数

### 2.2 分类模型构建

分类模型构建过程如下。

1) 从构建的数据集中选取部分作为训练集,人工添加建筑类型标签;0 表示未标记状态,1 表示非住宅建筑类型,2 表示住宅建筑类型。

2) 利用带标签的训练集数据建立分类模型,自变量为表 3 所示建筑轮廓的属性,因变量为建筑类型,训练并对模型性能进行评估,选择表现最佳的算法和参数组合作为最终分类模型。

3) 利用该模型对剩余测试集数据进行建筑类型预测。

此外,为提高分类精确性,引入地图中的信息点(point of interest, POI)作为分类的一个判别特征,并作为机器学习模型的输入数据。从高德地图中提取与选定研究区建筑群匹配的 POI,并将其表示为四元组格式:(POI ID, x 坐标, y 坐标, POI 类型代码),其中,“POI ID”为每个 POI 的通用标识符;“x 坐标”和“y 坐标”分别为高德地图坐标系中 POI 的

纬度和经度坐标,需转换为与用于将 POI 与建筑进行匹配的 GIS 文件相同的坐标系。“POI 类型代码”是指高德地图中每个 POI 的类别编号。高德地图中的 POI 分为大类、中类和小类,依次为包含关系。选取与建筑相关的 12 种大类,如表 4 所示。基于其坐标使用 ArcGIS 的 Spatial Join 功能将提取的 POI 映射到建筑物上。试验设置参数为:Join Operation = JOIN\_ONE\_TO\_MANY; Match Option = Within\_A\_Distance; Search Radius = 0.000 2 十进制度。过滤掉匹配值为 0 的建筑物。

表 4 高德地图中与建筑相关的 POI 类型代码

Table 4 POI type codes related to architecture in

Gaode map			
POI 类型代码	POI 含义	POI 类型代码	POI 含义
010000	汽车服务	080000	体育休闲服务
020000	汽车销售	090000	医疗保健服务
030000	汽车维修	100000	住宿服务
050000	餐饮服务	120000	商务住宅
060000	购物服务	140000	科教文化服务
070000	生活服务	170000	公司企业

### 2.3 分类算法选择和分类结果分析

选择并测试朴素贝叶斯分类算法(NB)、支持向量机分类算法(SVM)、决策树分类算法(DT)、随机森林分类算法(RF)、梯度提升决策树分类算法(GBDT)和 K 近邻算法(KNN)6 种机器学习分类算法。使用精确度(Precision)、召回率(Recall)、准确率(Accuracy)和 F1 分数对其分类性能进行评估。根据表 5 所示的性能评估结果,提出的分类方法在建筑类型分类方面表现有效,其中梯度提升决策树分类算法(GBDT)整体性能最优。

表 5 6 种机器学习分类算法性能评估

Table 5 Performance evaluation of six machine learning

classification algorithms				%
分类算法	精确度	召回率	准确率	F1 分数
NB	67.9	68.8	62.2	68.3
SVM	78.1	83.4	78.3	80.4
DT	83.3	82.0	78.1	82.6
RF	86.6	82.1	82.5	84.3
GBDT	89.2	87.1	83.8	88.1
KNN	79.2	88.3	80.1	83.2

## 3 基于回归算法的绿色建筑能耗模拟

### 3.1 建筑能耗数据集构建

从分类得到的住宅和非住宅建筑数据集中随机抽取 600 条样本,通过 Grasshopper 平台的 Ladybug+Honeybee 插件(以 EnergyPlus 为模拟引擎)进行能耗模拟,生成建筑能耗数据集。Grasshopper 模型的调整可通过修改参数直接实现。

在 Grasshopper 平台上进行建筑能耗建模和模拟的过程、所需运算器及功能如表 6 所示。

表 6 在 Grasshopper 平台上进行建筑能耗建模和模拟的过程

Table 6 Process of building energy consumption modeling and simulation on Grasshopper platform

序号	步骤	运算器名称	输入信息/参数	功能
1	气候信息输入	LB_Openweatherfile LB_importEPW	EPW 气候文件	气候分析
2	构建建筑几何模型	HB_Split2Floors HB_Split2Zone	由 BIM-GIS 相关文件获取的几何信息	将建筑几何模型转换为建筑性能模型
3	进行建筑能耗模拟	HB_Mass2Zone HB_bldgProgram HB_setZoneLoad HB_setEPZone Thresholds HB_glaingCreator HB_setEPZoneCnstr	区域功能的设定,设备照明能耗密度、人员密度、采暖制冷温度、窗墙比、围护结构的构造参数(传热系数、太阳能得热系数等)	用于 Energy Plus 能耗模拟引擎计算
4	模拟数据导出	HB_Massaddition HB_Division HB_Excelwrite	—	导出模拟数据供后续研究使用

### 3.2 能耗回归模型构建

#### 3.2.1 回归模型选择

选用线性回归模型、极端梯度提升回归模型(XGBoost)和人工神经网络模型(ANN)3 种经典机器学习回归模型进行预测。常用的线性回归模型包括普通最小二乘线性回归(OLS)、岭回归(Ridge)、最小绝对收缩和选择算子模型(LASSO)和弹性网络模型(Elastic Net)。

#### 3.2.2 数据预处理与数据集划分

在模型训练前,对所有预测变量进行中心化和缩放,作为数据预处理步骤,以避免较高数值范围内的属性占据主导地位,并提高数据稳定性。

预处理后,将所有可用数据随机分为两部分:20%作为测试集,80%作为训练集和验证集,并保持 2 组数据集中住宅和非住宅建筑的比例相等。

#### 3.2.3 模型性能评价

采用十折交叉验证(10-fold cross-validation)方法对不同机器学习回归模型进行模型评估。将数据集均匀划分为 10 个子集,进行 10 次迭代,每次迭代中使用 1 个子集用于验证数据,剩余 9 个子集用于训练。比较 10 次迭代的平均性能并确定最佳的调整参数后,使用整个训练集训练最终的模型,使用未参与任何训练或参数调整过程的测试集评估

最终模型的预测准确性。

选用归一化平均绝对误差 NMAE、正规化均方根误差 NRMSE 和相对误差对模型进行评价。

1) NMAE 可消除不同数据集或不同任务间由于规模或单位不同而导致的误差值差异,使其之间性能更加公平准确。其计算如式(4)所示。

$$\text{NMAE} = \frac{\sum_{k=1}^n |y_k - \hat{y}_k|}{\sum_{k=1}^n y_k} \quad (4)$$

2) NRMSE 能比较不同单位或范围的数据集的预测准确性,同时帮助评估模型的预测精度及比较不同模型在同一数据集中的表现。计算如式(5)所示。

$$\text{NRMSE} = \frac{\sqrt{\frac{\sum_{k=1}^n (y_k - \hat{y}_k)^2}{n}}}{\sum_{k=1}^n y_k} \quad (5)$$

3) 为评估机器学习模型对整体存量、住宅存量和非住宅存量能耗预测的准确性,使用式(6)估计测试集中全部建筑、所有住宅建筑和所有非住宅建筑的总供暖和制冷能耗的相对误差  $\delta_s$ 。相对误差绝对值越小,说明预测值与实际值越接近,模型的预测性能越好,计算如式(6)所示。

$$\delta_s = \frac{\sum_{k=1}^m (\hat{y}_k F_k) - \sum_{k=1}^m (y_k F_k)}{\sum_{k=1}^m (y_k F_k)} \times 100\% \quad (6)$$

式中:  $y_k$  为通过 EnergyPlus 模拟得到的第  $k$  栋建筑的供暖或制冷能耗;  $\hat{y}_k$  为通过机器学习回归模型得到的第  $k$  栋建筑的供暖或制冷能耗;  $n$  为测试集中建筑总数;  $m$  为测试集中属于特定建筑存量的建筑总数;  $F_k$  为第  $k$  栋建筑总面积。

### 3.3 回归结果分析

应用不同机器学习回归模型的 NMAE 和 NRMSE 结果如表 7 所示。一般来说,在 NMAE 和 NRMSE 较小的情况下,制冷能耗的预测精度优于供暖能耗。整体来看人工神经网络在建筑单体级别上的精度最好。

机器学习回归模型在建筑存量供暖和制冷能耗预测方面的表现如表 8 所示。对于全部存量,除人工神经网络模型供暖预测的相对误差为 -9.7%

表 7 应用不同机器学习回归模型的 NMAE, NRMSE 结果

Table 7 NMAE and NRMSE results using different machine learning regression models

模型	供暖能耗		制冷能耗	
	NMAE	NRMSE	NMAE	NRMSE
OLS	0.172	0.357	0.062	0.103
Ridge	0.158	0.321	0.063	0.097
LASSO	0.172	0.354	0.059	0.093
Elastic Net	0.158	0.321	0.064	0.101
XGBoost	0.133	0.312	0.060	0.134
ANN	0.138	0.298	0.061	0.089

外,其他预测的相对误差均在  $\pm 4\%$  以内。岭回归表现最好,其次是弹性网络模型。人工神经网络模型虽然对于供暖预测具有较高的相对误差,但在制冷预测方面表现好。对于住宅建筑存量,极端梯度提升回归模型在供暖和制冷预测方面表现均最好。对于非住宅建筑存量,极端梯度提升回归模型和人工神经网络模型分别在供暖和制冷预测方面表现最好。

表 8 不同机器学习回归模型在建筑存量能耗预测方面的表现

Table 8 Performance of different machine learning regression models in building stock energy

模型	consumption prediction						%
	全部建筑存量		住宅建筑存量		非住宅建筑存量		
	供暖	制冷	供暖	制冷	供暖	制冷	
OLS	-3.8	1.7	-7.6	3.2	8.2	-2.1	
Ridge	-2.3	0.5	-5.7	1.4	8.8	-1.6	
LASSO	-3.8	1.7	-7.6	3.2	8.2	-2.1	
Elastic Net	-2.3	0.6	-5.7	1.8	8.8	-2.6	
XGBoost	2.4	-1.0	5.5	0.1	-7.7	-3.8	
ANN	-9.7	0.2	-9.0	0.7	-11.7	-1.2	

3 种机器学习回归模型均能较准确地预测建筑能耗,预测 600 栋建筑供暖和制冷能耗的运算时间均在 5s 内,而对 1 栋建筑的供暖和制冷能耗进行能耗模拟需  $> 10s$ 。机器学习回归模型可使建筑供暖和制冷能耗的预测速度提高 10 000 倍以上,大大缩短预测所需时间,有利于大规模建筑存量能耗预测。

## 4 结语

本研究构建了一个基于机器学习的预测框架,集成 BIM 和 GIS 用于区域尺度绿色建筑能耗预测。首先利用 FME 实现了从 IFC 到 CityGML 的完整转换流程,应用分类算法对区域尺度绿色建筑的建筑类型进行分类,并构建基于机器学习的能耗模拟模型。通过实例验证,提出的预测方法可行且较为精确,能够通过减少物理模拟的建筑数量显著缩短处理时间。生成的机器学习回归模型可用于快速预测存量建筑的供暖和制冷能耗,并评估不同改造方

案的节能潜力。

本研究为区域尺度绿色建筑能耗预测提供了一种新的思路和方法,但研究仍存在一定不足。一方面,通过 FME 实现的 IFC 到 CityGML 转换以手动操作为主,自动化程度较低。另一方面,由于建筑周围环境的影响因素难以量化并标准化输出,研究只考虑了区域气候和建筑自身特性对能耗的影响,未将微气候的影响纳入考量。后续研究可尝试应用更先进的计算机技术实现转换流程的自动化,探索并总结建筑环境的定量描述手段,也可结合先进机器学习算法和数据融合技术,提高预测精度和泛化能力。

#### 参考文献:

- [ 1 ] 刘晓君,胡升凯,李玲燕. 中国省区建筑能耗时空分布与影响因素分析[J]. 数学的实践与认识,2020,50(6):74-85.  
LIU X J, HU S K, LI L Y. Analysis of spatial and temporal distribution and influencing factors of building energy consumption in Chinese provinces [J]. Mathematics in practice and theory, 2020,50(6):74-85.
- [ 2 ] MA J, CHENG J C. Estimation of the building energy use intensity in the urban scale by integrating GIS and big data technology [J]. Applied energy, 2016, 183: 182-92.
- [ 3 ] DAVILA C C, REINHART C F, BEMIS J L. Modeling Boston: a workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets [J]. Energy, 2016, 117: 237-50.
- [ 4 ] 于军琪,杨思远,赵安军,等. 基于神经网络的建筑能耗混合预测模型[J]. 浙江大学学报(工学版),2022,56(6):1220-1231.  
YU J Q, YANG S Y, ZHAO A J, et al. A hybrid prediction model of building energy consumption based on neural network [J]. Journal of Zhejiang University (engineering science), 2022,56(6):1220-1231.
- [ 5 ] 滕文龙,丛炳虎,商云坤,等. 基于 MEA-BP 神经网络的建筑能耗预测模型[J]. 吉林大学学报(工学版),2021,51(5):1857-1865.  
TENG W L, CONG B H, SHANG Y K, et al. Building energy consumption prediction model based on MEA-BP neural network [J]. Journal of Jilin University (engineering and technology edition), 2021,51(5):1857-1865.
- [ 6 ] DONG Z, LIU J, LIU B, et al. Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification [J]. Energy and buildings, 2021, 241: 110929.
- [ 7 ] 付显曦,李红莲,王赏玉,等. 建筑节能分析用典型年数据的获取方法[J]. 西安建筑科技大学学报(自然科学版),2021,53(1):147-154.  
FU Y X, LI H L, WANG S Y, et al. Acquisition method of typical year data for building energy conservation analysis [J]. Journal of Xi'an University of Architecture & Technology(natural science edition), 2021,53(1):147-154.
- [ 8 ] BETTI G, TARTARINI F, NGUYEN C, et al. CBE clima tool: a free and open-source web application for climate analysis tailored to sustainable building design[J]. Building simulation, 2024, 17(3): 493-508.
- [ 9 ] 宋芳婷,诸群飞,吴如宏,等. 中国建筑热环境分析专用气象数据集 [C]//全国暖通空调制冷 2006 学术年会资料集,2006.  
SONG F T, ZHU Q F, WU R H, et al. Special meteorological data set for thermal environment analysis of Chinese buildings [C]//National HVAC and Refrigeration 2006 Academic Annual Meeting, 2006.
- [ 10 ] BELCHER S E, HACKER J N, POWELL D S. Constructing design weather data for future climates [J]. Building services engineering research and technology, 2005, 26(1): 49-61.
- [ 11 ] THRASHER B, WANG W, MICHAELIS A, et al. NASA global daily downscaled projections, CMIP6[J]. Scientific data, 2022, 9(1): 262.
- [ 12 ] NAGELER P, ZÄHRER G, HEIMRATH R, et al. Novel validated method for GIS based automated dynamic urban building energy simulations [J]. Energy, 2017, 139: 142-54.
- [ 13 ] 张文元,刘润桦. 古建筑三维语义模型转换方法研究[J]. 测绘科学,2023,48(2):227-236.  
ZHANG W Y, LIU R H. Research on the conversion method of 3D semantic model of ancient buildings [J]. Surveying and mapping science, 2023,48(2): 227-236.
- [ 14 ] QI F, ZHAI J Z, DANG G. Building height estimation using Google Earth [J]. Energy and buildings, 2016, 118: 123-132.