

基于 PSO-LGBM 的居住建筑碳排放预测模型及分析

摘要

碳中和作为近年来国内的重点战略，目的是加快降低碳排放步伐，大力推进绿色低碳科技创新，提升产业和研究如何经济全球竞争力。本文通过构建居住建筑碳排放指标体系，形成居住建筑碳排放预测模型，通过该模型用以评估和预测居住建筑的碳排放，进而促进绿色环保，低碳的发展，响应国家碳中和号召。

针对问题一：本文对给定条件的单体建筑物通过空调调节温度的年碳排放量进行了计算，首先通过计算建筑物的热损失系数，针对每个月计算室内外温度之差，并乘以各墙体的热损失系数和面积，得到每个月的热损失量，累加得年度热损伤量 8454.8kw/h，通过 COP 与 EER 值计算每个月的电能消耗，累加得年消耗电量 2577.9kw/h，将其转化为碳排放量，即得到通过空调调节温度的年碳排放量为 749.8kg。

针对问题二：本文基于全流程覆盖，可比性等。七项指标构建原则，通过文献研究给出了初始指标体系在经过八位从事生态学或建筑学的专业人员的三轮征询下，筛选确定了指标，形成了三类一级指标 13 项二级指标的建筑生命周期碳排放指标体系，为了对建居住建筑的生命周期碳排放进行综合评价，本文对一，二级指标进行了主客观赋权，其中一级指标结合相关业界研究文献采用了主观赋权，二级指标体系通过 2021 年江苏省 13 个地区市的截面数据采用熵权法进行了客观赋权，整理得到指标体系权重如 6.5.2 所示，研究表明，对于 1 级指标体系运行阶段与建筑阶段的权重分别为 0.7 与 0.25，拆除阶段的权重占比较小，而对于 2 级指标，居住人口数量，生活方式指数以及能源消耗量是排行前三项的关键影响指标。

针对问题三：本文在问题二的权重基础上采用了 TOPSIS 法对江苏省地级市居住建筑碳排放进行了综合评价，研究表明，苏州的居住建筑碳排放综合得分最高，徐州市排行最末。即苏州市居住建筑的碳排放对比其他各地区是更为合理或理想，同时本文还与江苏省 2021 年环境空气质量综合排名进行了对比，发现建筑生命周期碳排放与空气质量排名表现较为吻合，这也验证了模型的可靠性。

针对问题四：本文基于构建 LGBM 与 LSTM 江苏省建筑全过程的碳排放量预测模型，由于 LGBM 模型对比 LSTM 模型在测试集的表现更为优秀，因此我们选取了更为优秀的 LGBM 模型与 PSO 算法进行了超参数调优，再得到最优模型参数后，我们取得了更好的模型预测效果，在测试级的表现上 R2 达到了 0.89，平均百分绝对误差比仅为 3.91%，较前面两个基础模型都有了明显的提升。

最后结合上面四个问题的分析结果我们给出了对江苏省政府对于居住建筑碳排放的建议与措施。

关键字： 指标体系 PSO-LGBM 熵权法 TOPSIS

目录

一、问题重述	3
二、问题分析	4
三、模型假设	4
四、符号说明	5
五、问题 1：建筑物（空调）调节温度的年碳排放量模型建立与求解	6
六、问题 2：建筑生命周期碳排放指标体系的构建	7
6.1 建筑生命周期碳排放指标	7
6.2 斯皮尔曼相关系数分析	9
6.3 通过熵权法进行权重分析	10
七、问题 3：基于 TOPSIS 法的江苏省地级市居住建筑碳排放综合评价	11
八、问题 4：江苏省建筑全过程的碳排放量预测模型建立与求解	12
8.1 定义指标对模型的预测能力进行表达	12
8.2 基于深度学习 LSTM 时间序列预测模型建立与求解	13
8.3 基于粒子群算法（PSO）优化的 LGBM 时序模型建立与求解	14
九、问题 5：总结与建议	16
十、Python 源程序	16
参考文献	21

一、问题重述

“双碳”是指碳达峰和碳中和，中国致力于在 2030 年前实现碳达峰，2060 年前实现碳中和。该战略提倡绿色、环保、低碳的生活方式，加快降低碳排放步伐，大力推进绿色低碳科技创新，以提高产业和经济的全球竞争力。低碳建筑是指在建筑材料与设备制造、施工建造和建筑物使用的整个生命周期内，减少化石能源的使用，提高能效，降低二氧化碳排放量。本文提出以下问题：

问题 1：计算单层平顶单体建筑的年碳排放量

考虑一个长 4 米、宽 3 米、高 3 米的单层平顶单体建筑，墙体为砖混结构，厚度为 30 厘米；屋顶为钢筋混凝土浇筑，厚度为 30 厘米；门窗总面积为 5 平方米；地面为混凝土。该建筑物所处地理位置的月平均温度如表所示。

月份	1	2	3	4	5	6	7	8	9	10	11	12
平均温度	-1	2	6	12	22	28	31	32	26	23	15	2

假设建筑物内部温度需要保持在 18-26 摄氏度范围内，通过电力调节温度时，每消耗 1 度电产生 0.28 千克碳排放。请基于给定条件，利用空调制热性能系数 COP 为 3.5，制冷性能系数 EER 为 2.7，计算该建筑物的年碳排放量。

问题 2：建立居住建筑生命周期碳排放综合评价模型

考虑影响居住建筑生命周期（建造、运行、拆除）碳排放的诸多因素，如建筑设计标准、气候、建材生产运输、地区差异、建造拆除能耗、装修风格、使用能耗、建筑类型等。请查找和分析相关资料，建立数学模型，找出与这些因素相关度较高且易于量化的指标，并基于这些指标对居住建筑整个生命周期的碳排放进行综合评价。

问题 3：评价江苏省地级市居住建筑碳排放及模型有效性验证

在问题 2 的基础上，分别考虑建筑生命周期的三个阶段，查找相关资料，建立数学模型，对 2021 年江苏省 13 个地级市的居住建筑碳排放进行综合评价。同时，对所建立的评价模型的有效性进行验证。

问题 4：预测江苏省建筑全过程碳排放量

建立碳排放预测模型，基于江苏省建筑全过程碳排放的历史数据，预测 2023 年江苏省建筑全过程的碳排放量。准确的碳排放预测将为制定减排政策和优化低碳建筑设计提供重要参考依据。

问题 5：提出江苏省建筑碳减排政策建议

结合前面的讨论，为江苏省建筑碳排放减排提出政策建议。

二、问题分析

对于第一问：根据题意计算出每个月该建筑物需要的制热量和制冷量，以及需要消耗的电量。制热量和制冷量可以通过建筑物的热负荷计算公式计算得出，而需要消耗的电量可以根据空调的制冷性能系数 EER 和制热性能系数 COP 计算得出。然后，可以根据消耗的电量和碳排放系数计算出该建筑物通过空调调节温度的年碳排放量。最后，将所有月份的碳排放量相加即可得到该建筑物的年碳排放量。

对于第二问：我们需要构建建筑生命周期碳排放指标体系，可以先通过梳理相框，文献综述得到一个初始指标体系，基于该指标体系，通过对相关学科专家进行调研，通过德尔菲法。等方法进一步筛选确定该指标体系，为了对居住建筑整个生命周期的碳排放进行综合评价，我们需要对上面确定的指标体系进行权重分析。

对于第三问：我们可以在第二问的基础上收集 2021 年江苏省 13 个地级市的居住建筑碳排放相关数据，进而求解出其综合评价，综合评价模型可以采用 topsis 或 RSR 秩和比等方法，同时为了对评价模型进行有效性验证，我们可以收集江苏省真实的碳排放数据进行对比论证。

对于第四问：需要对 2023 年江苏省建筑全过程的碳排放量进行预测，可以利用时间序列分析方法，根据历史数据建立时间序列模型，预测未来的碳排放量，在建立模型时，需要考虑建筑全过程的不同阶段的碳排放数据，选择合适的时间序列模型，如 ARIMA、LSTM 等，对历史数据进行拟合和验证，然后利用模型预测未来的碳排放量。

对最后一问：结合上面的结论，为江苏省政府提出一些措施与建议。

三、模型假设

针对本文提出的问题，我们做了如下模型假设：

1. 假设题目所给的数据真实可靠。
2. 建筑物的室内温度在制热时保持 18 摄氏度。在制冷时保持 26 摄氏度。
3. 建筑物的墙体、屋顶、门窗和地面的热导系数和厚度保持恒定。
4. 建筑物的墙体、屋顶、门窗和地面的热损失均为线性，即热损失量与室内外温差成正比。
5. 每月的室外平均温度已知且保持恒定。
6. 空调制热性能系数 COP 为 3.5，制冷性能系数 EER 为 2.7，且保持恒定。
7. 1 度电消耗对应的碳排放量为 0.28 千克碳/度电，且保持恒定。
8. 地面混凝土的厚度假设为 1 米。

四、符号说明

本文常用符号见下表，其它符号见文中说明：

符号	意义
U	热损失参数
A	面积
Q	热损失
E	电能损失
$Carbon$	电能损失
ρ	相关系数
w	权重

五、问题 1：建筑物（空调）调节温度的年碳排放量模型建立与求解

为了计算这个单体建筑物通过空调调节温度的年碳排放量，我们首先需要计算在每个月份建筑物通过空调调节温度所需的能量，然后根据空调的性能系数将能量转换为度电，最后将度电转换为碳排放量，其计算步骤如下：

1. 确定各部分的热损失系数（U 值）和面积：

- $U_{wall} = 1W/m^2K$
- $U_{roof} = 0.67W/m^2K$
- $U_{window} = 1.6W/m^2K$
- $U_{ground} = 0.25W/m^2K$
- $A_{wall} = 42m^2$
- $A_{roof} = 12m^2$
- $A_{window} = 5m^2$
- $A_{ground} = 12m^2$

2. 计算每个月的热损失：

假设室内温度在制热时保持 18 摄氏度，在制冷时保持 26 摄氏度。

制热月份（1-4 月、11-12 月）：

$$\Delta T_{wall} = 18 - T_{outside}, \Delta T_{roof} = \Delta T_{ground} = \Delta T_{window} = \Delta T_{wall}$$

制冷月份（5-10 月）：

$$\Delta T_{wall} = T_{outside} - 26, \Delta T_{roof} = \Delta T_{ground} = \Delta T_{window} = \Delta T_{wall}$$

热损失量 Q（每月）：

$$Q_{month} = (\Delta T_{wall} \times U_{wall} \times A_{wall} + \Delta T_{roof} \times U_{roof} \times A_{roof} + \Delta T_{window} \times U_{window} \times A_{window} + \Delta T_{ground} \times U_{ground} \times A_{ground}) \times 24h \times 30days$$

3. 计算每个月的度电消耗：

电能消耗（每月）：

$$E_{month_{heating}} = Q_{month_{heating}}/COP \quad E_{month_{cooling}} = Q_{month_{cooling}}/EER$$

计算年度度电消耗：

将每个月的度电消耗累加，得到年度度电消耗。

$$E_{annual} = \sum E_{month}$$

4. 计算年度碳排放量：

将年度度电消耗转换为碳排放量。

$$Carbon_{emission} = E_{annual} \times 0.28kgCO_2/kWh$$

根据以上计算过程，我们可以得到每个月的度电消耗和年度碳排放量。具体计算结果见附件“data.xlsx”。

年度度电消耗：

$$E_{annual} = \sum E_{month} = 498.3 + 361.9 + 324.8 + 144 + 0 + 74.7 + 178.6 + 214.2 + 0 + 0 + 86.4 + 444 = 2326.9kWh$$

年度碳排放量：

$$Carbon_{emission} = E_{annual} \times 0.28kgCO_2/kWh = 2326.9kWh \times 0.28kgCO_2/kWh = 651.532kgCO_2$$

综上所述，在考虑每个月真实天数后，该建筑物通过空调调节温度的年碳排放量为 651.532 kg。

六、问题 2：建筑生命周期碳排放指标体系的构建

通过文献研究，结合建立建筑生命周期碳排放评价的特殊性和繁杂性，本文主要运用专家问卷法进行碳排放评价的指标体系构建，将此方法的应用原理和适用性说明如下：

德尔菲法：是指调查人员按照规定的流程对专家们发放调查问卷，在专家们对调查人员提出的问题回答完之后，统一将数据归纳整理，并再次向专家征询意见，并进一步归纳整理数据，最终产生结果。这种方法可以有效减少因为某一个人的主观因素所造成的影响。

6.1 建筑生命周期碳排放指标

首先是初始指标的建立。基于指标体系构建的原则与社会稳定相关理论方法，同时参考其他相关文献以及对有经验学者的意见，初步建立一个建筑生命周期碳排放指标体系的框架，如附件“zhibiao.xlsx Sheet1”所示。在征询专家意见后，将一部分指标进行调整或剔除，建立初步指标体系。

建筑生命周期碳排放指标体系是评价建筑物环境性能和可持续性的重要指标之一。建立指标体系需要包括建筑物建造、使用和拆除等阶段的碳排放量，以及建筑材料生产、运输、施工和处理等过程中的碳排放量。以下是结合国内外专家的论文，给出建立建筑生命周期碳排放指标体系相关的理论：

IPCC 指南：《温室气体排放清单手册》是建立建筑生命周期碳排放指标体系的重要参考，它提供了建筑物能源消耗、材料生产和运输、废弃物处理等方面的数据和方法。该手册还提供了不同类型建筑物的参考数据，可用于建立建筑物碳排放清单。

EN15978 标准：欧洲标准 EN15978 规定了建筑物生命周期评价的方法和指标，包括建筑物建造、使用和拆除阶段的环境影响评价和碳足迹评估。该标准提供了建筑物材料的数据来源和评估方法，包括环境产出和材料生命周期分析等，可用于评估建筑物碳排放量。

LEED 认证：LEED 是美国绿色建筑委员会开发的评估建筑物环境性能认证系统，

其中包括了碳排放指标。该认证系统提供了建筑物碳排放量的评估方法和指标，涵盖了建筑物建造、使用和拆除等阶段的碳排放，以及建筑材料的生产和运输等过程中的碳排放。

GB/T 51211-2016 标准：国家标准 GB/T 51211-2016 规定了建筑物碳排放清单的编制方法和指南，该标准提供了建筑物碳排放清单的基本要素、数据来源、计算方法和结果表述等方面的指导，可用于评估建筑物的碳排放量。

国内专家论文：国内专家也对建立建筑生命周期碳排放指标体系进行了研究和探讨，如《建筑生命周期碳排放评估方法研究》、《基于 LCA 的建筑生命周期碳排放评价研究》等论文提出了建筑物生命周期碳排放评价的方法和指标，为建筑物环境性能评估和可持续设计提供了理论支持。综上所述，建立建筑生命周期碳排放指标体系需要考虑建筑物的整个生命周期，包括建筑物的设计、建造、使用和拆除等不同阶段。建筑物的碳排放量不仅包括建筑物使用过程中的能耗排放，还包括建筑材料生产、运输、施工和处理等过程中的碳排放量。因此，在建立建筑生命周期碳排放指标体系时，需要综合考虑不同过程中的碳排放量，并使用适当的方法和指标进行评估和计算。

具体来说，建立建筑生命周期碳排放指标体系需要从以下几个方面考虑：

碳排放清单编制：建立建筑物碳排放清单是建立碳排放指标体系的基础，需要考虑建筑物使用过程中的能耗排放和建筑材料生产、运输、施工和处理等过程中的碳排放量。可根据国际标准和国内规范进行清单编制，如 EN15978 标准、GB/T 51211-2016 标准等。

生命周期评价：生命周期评价是评价建筑物环境性能的一种方法，它可以全面考虑建筑物的碳排放量、能耗、水资源利用等方面的影响。生命周期评价可以通过 LCA 等方法进行，其中需要考虑建筑材料生命周期、能源消耗和废弃物处理等方面的影响。

碳足迹评价：碳足迹评价是评价建筑物碳排放量的一种方法，它可以量化建筑物整个生命周期中的碳排放量。碳足迹评价可以通过建筑物的能耗数据和建筑材料生命周期数据进行计算，可使用专业软件进行模拟和计算。

评价指标体系：建立建筑生命周期碳排放指标体系需要确定评价指标和计算方法。评价指标可以包括建筑物碳排放总量、碳排放强度、碳足迹等指标，计算方法可以包括数据采集、模拟计算、统计分析等方法。

综上所述，建立建筑生命周期碳排放指标体系需要综合考虑建筑物整个生命周期的碳排放量，并使用适当的方法和指标进行评估和计算。该指标体系可以为建筑物环境性能评估和可持续设计提供支持和指导。

在构建建筑生命周期碳排放指标体系时，我们需要为一级、二级指标设置权重。根据国内外相关文献，对于一级指标，由于国内外有较文献描述，因此我们采用文献综述法进行一级指标权重设定：

建造阶段指标：权重 25%；运行阶段指标：权重 70%；拆除阶段指标：权重 5%。

这三个一级指标的权重分配依据了国内外相关文献，分别为建造阶段指标（25%）、运行阶段指标（70%）和拆除阶段指标（5%）。这些权重分配将有助于更准确地评估建筑生命周期碳排放的影响因素，并为减少建筑碳排放提供有力的指导。

为了顺应下文的分析以及确定二级权重，本文就江苏省 13 个地级市的数据展开了收集工作，数据收集是本环节最重要的一步。本文根据上面提出的指标，从国家统计局、江苏省统计局、江苏省 13 个地市统计局、江苏省生态环境厅以及中国建筑节能协会发布的报告中，选取与本文提出指标相近或者能通过直接计算得出的部分年鉴统计数据计算，选择其中几项可以在年鉴中找到原始数据。本文数据主要来源于等针对 2021 年的有关部分。整理 2021 年度年鉴部分统计数据如附件“2021data.xlsx”。

6.2 斯皮尔曼相关系数分析

斯皮尔曼相关系数被定义成等级变量之间的皮尔逊相关系数。对于样本容量为 n 的样本， n 个原始数据被转换成等级数据，相关系数 ρ 为

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

原始数据依据其在总体数据中平均的降序位置，被分配了一个相应的等级。如下表所示：

变量 X_i	降序位置	等级 x_i
0.8	5	5
1.2	4	4
1.2	3	3
2.3	2	2
18	1	1

实际应用中，变量间的连结是无关紧要的，于是可以通过简单的步骤计算。被观测的两个变量的等级的差值，则

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

斯皮尔曼相关系数对 13 个 2 级指标进行相关性分析后，可以得到热力图呈现，如下图所示，可以看到各指标的相关性基本呈现为中等水平，这说明个指标相互之间存在一定的相关性，但又不存在严重的共线性，进一步说明了我们收集的数据较为合理。

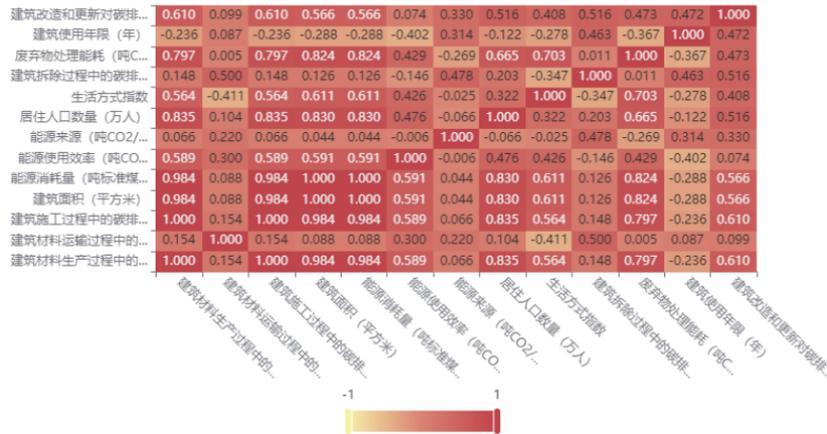


图1 Spearman 相关性

6.3 通过熵权法进行权重分析

熵是信息论中的概念,是对不确定性的一种度量。信息量越大,不确定性越小,熵就越小;信息量越小,不确定性越大,熵也越大。根据信息熵的定义,对于某项指标可用熵值来判断某个指标的离散程度,其熵值越小,指标的离散程度越大,该指标对综合评价的影响(即权重)越大。其步骤为:

(1) 对各个因素按照每个选项的数量进行归一化处理,对于正向指标:

$$x'_{ij} = \frac{X_{ij} - \min(X_{1j}, X_{nj}, \dots, X_{nj})}{\max(X_{1j}, X_{nj}, \dots, X_{nj}) - \min(X_{1j}, X_{nj}, \dots, X_{nj})}$$

对于负向指标:

$$x'_{ij} = \frac{\max(X_{1j}, X_{nj}, \dots, X_{nj}) - X_{ij}}{\max(X_{1j}, X_{nj}, \dots, X_{nj}) - \min(X_{1j}, X_{nj}, \dots, X_{nj})}$$

(2) 计算第 j 项指标的熵值

$$e_j = -k \sum_{i=1}^n p_{ij} \ln(p_{ij}), \quad j = 1, \dots, m$$

其中, $k = 1/\ln(n) > 0$, $e_j \geq 0$ 。

(3) 计算信息熵余度(差异): $d_j = 1 - e_j$, $j = 1, \dots, m$ 。

(4) 计算各项指标的权重: $w_j = \frac{d_j}{\sum_{j=1}^m d_j}$, $j = 1, \dots, m$ 。

据结果对各个指标的权重进行分析: 建筑面积(平方米)的权重为 33.391%、建筑材料生产过程中的碳排放(吨 CO_2)的权重为 19.679%、建筑材料运输过程中的碳排放(吨 CO_2)的权重为 27.605%、建筑施工过程中的碳排放(吨 CO_2)的权重为 19.325%, 其中指标权重最大值为建筑面积(平方米)(33.391%), 最小值为建筑施工过程中的碳排放(吨 CO_2)(19.325%)。

同理可得运行阶段指标权重如下：

熵值法的权重计算结果显示，能源消耗量（吨标准煤）的权重为 19.901%、能源使用效率（吨 CO₂/平方米）的权重为 15.328%、能源来源（吨 CO₂/吨能源）的权重为 8.529%、居住人口数量（万人）的权重为 18.318%、生活方式指数的权重为 24.107%、建筑使用年限（年）的权重为 6.366%、建筑改造和更新对碳排放（吨 CO₂）的权重为 7.45%，其中指标权重最大值为生活方式指数（24.107%），最小值为建筑使用年限（年）（6.366%）。

熵值法的权重计算结果显示，废弃物处理能耗（吨 CO₂）的权重为 57.44%、建筑拆除过程中的碳排放（吨 CO₂）的权重为 42.56%，其中指标权重最大值为废弃物处理能耗（吨 CO₂）（57.44%），最小值为建筑拆除过程中的碳排放（吨 CO₂）（42.56%）。

最终得到指标权重如“zhibiao.xlsx Sheet1”所示。

七、问题 3：基于 TOPSIS 法的江苏省地级市居住建筑碳排放综合评价

TOPSIS 法是一种常用的组内综合评价方法，能充分利用原始数据的信息，其结果能精确地反映各评价方案之间的差距。基本过程为基于归一化后的原始数据矩阵，采用余弦法找出有限方案中的最优方案和最劣方案，然后分别计算各评价对象与最优方案和最劣方案间的距离，获得各评价对象与最优方案的相对接近程度，以此作为评价优劣的依据。该方法对数据分布及样本含量没有严格限制，数据计算简单易行。基本思想是：对原始数据同趋势后构建归一化矩阵，计算评价对象与最优向量和最劣向量的差异，以此测度评价对象的差异。假设有 n 个评价对象， m 个指标，TOPSIS 法的基本步骤为：

步骤 1：原始数据同趋势化区分指标体系中的指标类别（高优或低优），并根据不同类型的指标需要按照不同的公式进行正向化处理。

对于正向指标：

$$x'_{ij} = \frac{X_{ij} - \min(X_{1j}, X_{nj}, \dots, X_{nj})}{\max(X_{1j}, X_{nj}, \dots, X_{nj}) - \min(X_{1j}, X_{nj}, \dots, X_{nj})}$$

对于负向指标：

$$x'_{ij} = \frac{\max(X_{1j}, X_{nj}, \dots, X_{nj}) - X_{ij}}{\max(X_{1j}, X_{nj}, \dots, X_{nj}) - \min(X_{1j}, X_{nj}, \dots, X_{nj})}$$

构建 n 行 m 列的矩阵 X ，矩阵中 X 表示第 i 个对象的第 j 个指标的值。

步骤 2：构建标准化矩阵

$$Z_{ij} = \frac{X_{ij}}{\sqrt{\sum_{k=1}^n (X_{kj})^2}}$$

步骤 3：计算各评价指标与最优及最劣向量之间的差距

$$D_i^+ = \sqrt{\sum_{j=1}^m w_j (Z_j^+ - z_{ij})^2}, \quad D_i^- = \sqrt{\sum_{j=1}^m w_j (Z_j^- - z_{ij})^2}$$

其中 w_j 为第 j 个属性的权重。

步骤 4：评价对象与最优方案的接近程度

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-}$$

越大, 表明评价对象越优。

TOPSIS 法一般结合熵权法一起使用, 整理设置好的加权权重列表如 “zhibiao.xlsx Sheet3” 所示。

江苏省生态厅空气质量 2021 年环境空气质量综合排名如 “zhibiao.xlsx Sheet3” 所示, 环境空气质量综合指数是按照《环境空气质量评价技术规范》(HJ 663-2013) 对《环境空气质量标准》(GB3095-2012) 中的 6 项指标污染情况的综合评估, 指数越大表明综合污染程度越重。

因此, 通过建筑生命周期碳排放和空气质量指标的对比排名, 可以发现, 苏州市在综合排名和空气质量排名中都排名第一, 这说明苏州市在建筑生命周期碳排放方面相对于其他城市的表现比较优异, 同时在空气质量方面也表现得更好。而排名第二的无锡市也在空气质量排名中表现出较好的水平, 这与其在建筑生命周期碳排放方面的表现一致, 而且在江苏省 13 个地级市中, 排名前五位的城市在两个指标中均表现较好, 特别是在空气质量指标中排名前五位的城市也出现在建筑生命周期碳排放排名前五名中, 这说明了我们的综合评价模型与真实情况还是较为相似, 这验证本文所构建的模型可靠。

八、问题 4：江苏省建筑全过程的碳排放量预测模型建立与求解

在模型训练完之后, 就可以预测整个的测试集, 然后将预测的数据和实际数据对比。具体程序可见后附代码。

8.1 定义指标对模型的预测能力进行表达

为了评估模型的性能和预测情况, 本文根据预测值和实际观测值, 选用了均方根误差 (RMSE, Root Mean Squared Error)、平均绝对误差 (MAE, Mean Absolute Error) 和平均绝对百分比误差 (MAPE, Mean Absolute Percentage Error)、 R^2 作为模型的评估指标。它们的计算公式分别如下所示:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum (Y_i - \hat{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

其中, \hat{y}_i 为预测值, y_i 为真实值, N 为测试集的样本个数。当 RMSE、MAE、MAPE 的值越小时, 说明预测值与实际观测值之间的误差越小, 更进一步地说明所用预测模型的性能越好, R^2 则相反, 越接近 1, 说明所用预测模型的性能越好。

8.2 基于深度学习 LSTM 时间序列预测模型建立与求解

根据 LSTM 网络的结构, 每个 LSTM 单元的计算公式为:

$$\begin{aligned} f_t &= \sigma(W_f [h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_C [h_{t-1}, x_t] + b_c) \\ C_t &= f_t C_{t-1} + i_t \tilde{C}_t \\ O_t &= \sigma(W_o [h_{t-1}, x_t] + b_o) \\ h_t &= O_t \tanh(C_t) \end{aligned}$$

式中 f_t 是遗忘门, σ 是 Sigmoid 激活函数, x_t 是当前记忆单元的输入, i_t 是输入门限, \tilde{C}_t 是 t 时刻细胞状态的候选值, C_t 是 cell。

对于 LSTM, 由于神经网络含有多种超参数, 超参数的设置会严重影响实验结果。而合适的模型参数选择往往需要大量的实验验证, 因此, 为了能够与机器学习模型构成平等对比的前提, 本文采用业内默认的三层 LSTM 网络结构进行初次训练, 三层神经元个数分别为为 1024、512、64 作为基本超参数, 采用的参数优化方法是小批量梯度下降方法, batch 设置为 64, 同时还在池化层引入了 dropout 策略, 并将其设置为 0.5, 即每次训练随机舍弃一般的参数, 进一步防止过拟合现象的出现。

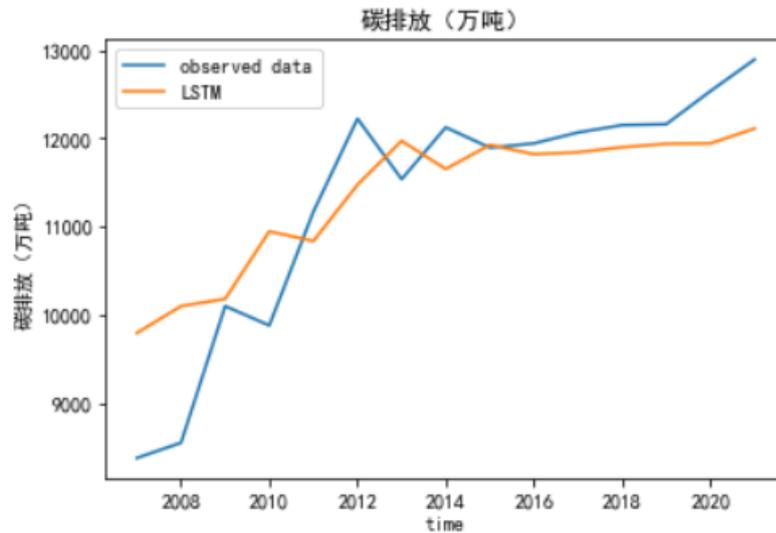


图 2 模型拟合效果

下表为其模型评估结果，我们可以看到 R^2 为 0.74，模型的表现良好。

MAE	RMSE	MAPE	R^2
793.2	871.1	7.309	0.74

8.3 基于粒子群算法 (PSO) 优化的 LGBM 时序模型建立与求解

模型超参数优化一般有六个步骤：

1. 构建相应的模型，确定搜索空间。
2. 从搜索空间中挑选超参数。
3. 将超参数组合应用至模型中，训练数据，并评估其在验证数据上的性能。
4. 根据评估结果选择下一组超参数组合。
5. 重复 2-4 步骤，直至达到迭代次数或者规定时间。
6. 最终评估模型在测试数据上的性能表现。

高效的超参数优化方法对于模型的效果提升具有重要的实际意义。以下本文选择工业界时下最火的模型超参数优化启发式算法 PSO 对中表现最优的回归模型 LGBM 进行优化。

粒子群 (Particle Swarm Optimization, PSO) 算法最初由 Kennedy 和 Eberhart 一起发现的一种寻优算法。其通过对鸟类的群体行为的分析、建模、仿真，借助 Heppner 生物学模型。其模型在群体行为研究层面和大多数模型相似，假设初始阶段，鸟群全部处于无序飞行状态，无目标，无规划路径，当存在某只鸟飞回栖息地时，若设置的栖息阈值比继续跟随鸟群飞行大时，所有鸟均会自主飞向栖息地，形成另一个群体。鸟类确定自己飞行位置和速度只是依靠简单的规则，当某一只飞离种群，返回栖息地，其他也会飞

向栖息地。且一旦发现栖息地，将不会离开，整个种群均留在这里。首先找到栖息地，并停留的鸟，会引起周围种群向其所在地飞行的方法，使整个种群找到栖息地的概率大大增加。此类问题类似于特定问题的最优解求法，因此 Kennedy 和 Eberhart 对 Heppner 的模型修正，使其能够飞向空间，找到最优解。

标准 PSO 算法是一种全局寻优算法，集中了“群体”和“优化”原则，通过微粒的适应值寻优。PSO 算法保留了基于种群的全局搜索策略，将每个个体看作是在 n 维搜索空间里的无重量和体积的微粒，并在搜索空间中以一定的速度飞行，该飞行速度由个体飞行经验和群体的飞行经验进行动态调整。在每次迭代中，每个微粒根据下列来调整其飞行速度和位置。

$$v_{ij}(t+1) = wv_{ij}(t) + c_1r_{1j}[p_{vj}(t) - x_{ij}(t)] + c_2r_{2j}[p_{gj}(t) - x_{ij}(t)]$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1)$$

其中， j 表示粒子第 j 分量； i 表示第 i 个粒子； t 表示第 t 代； c_1, c_2 表示加速常数。由于 LGBM 表现效果最优，因此基于 PSO 进行超参数调优，分析结果采用 5 层交叉检验，其步骤如下：

- 第一步：根据经验对 LGBM 的 PSO 的参数进行初始化；
- 第二步：根据 Bootstrap 算法，随机抽取样本数据中 k 个样本生成决策树；
- 第三步：计算模型的输出结果；
- 第四步：将上述的分类结果作为适应度值，利用 PSO 算法不断地迭代，进行参数寻优，并于历史结果比较，最终输出最优的模型参数；
- 第五步：根据得到的模型参数训练 LGBM，最终得出 MSE 最低即效果最好的超参数。

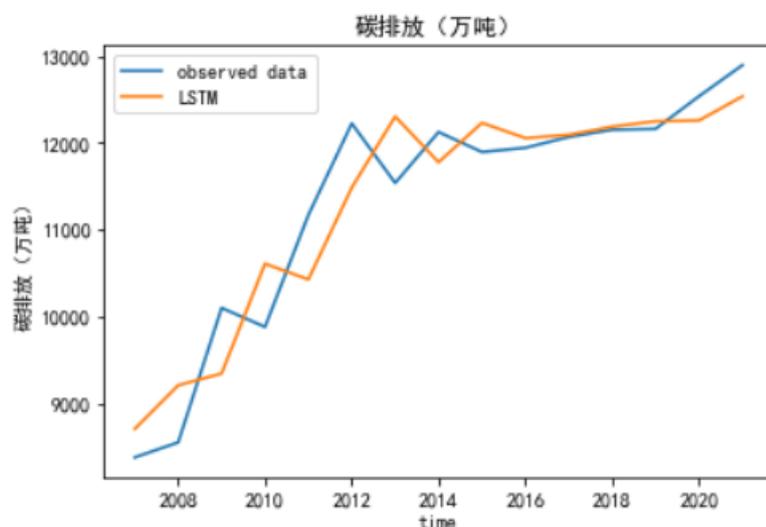


图 3 优化后的模型拟合效果

上图是优化后的模型拟合效果，其中蓝色曲线是真实数据，橙色曲线是 LGBM-PSO 拟合数据可以明显的看出应用了 PSO 算法进行优化后，模型的效果显著提高了，下表为其模型评估结果，我们可以看到 R^2 为 0.89，模型的表现非常优秀。

MAE	RMSE	MAPE	R^2
420.13	502.74	3.91	0.89

基于 LGBM-PSO 模型对未来 1 个时间单位进行了预测，既对未来 1 年进行了预测，2023 年江苏省居住建筑碳排放为 12790.28 万吨。

九、问题 5：总结与建议

结合本文对江苏省居住建筑全流程碳排放的分析，我们给出以下政策建议与措施：

- 鼓励绿色建筑设计和技术创新：政府可通过资金支持、税收优惠等措施，推动绿色建筑设计和低碳技术的研发与应用，从源头降低碳排放。
- 优化能源结构和提高能源使用效率：政府应调整能源结构，提高可再生能源在总能源消费中的比重，同时推广节能建筑材料和设备，提高建筑物的能源使用效率。
- 建立碳排放数据监测与报告制度：政府应建立完善的碳排放数据监测和报告制度，以便准确评估碳排放情况，为政策制定提供科学依据。
- 强化居民碳排放意识与生活方式改变：通过教育宣传等手段，提高居民对碳排放的认识，引导居民改变生活方式，实践绿色低碳生活。

综上所述，政府应结合本研究成果，制定针对性的政策和措施，以降低江苏省居住建筑全流程碳排放，推进碳中和战略的实现。

十、Python 源程序

```
import pandas as pd
data=pd.read_excel('2006-2022data.xlsx')
data
import sys
get_ipython().system('{sys.executable} -m pip install --upgrade pip')
get_ipython().system('{sys.executable} -m pip install tensorflow -i
    https://pypi.tuna.tsinghua.edu.cn/simple')
get_ipython().system('{sys.executable} -m pip install keras')

import pandas as pd
import warnings
```

```

from sklearn.metrics import r2_score
warnings.filterwarnings('ignore')
import numpy
import matplotlib.pyplot as plt
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM
from keras.models import Sequential, load_model

dataset=data['建筑碳排放（万吨）']
dataset = dataset.astype('float32')
def mean_absolute_percentage_error(y_true, y_pred):

return np.mean(np.abs((y_true - y_pred) / y_true)) * 100
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler(feature_range=(0, 1))
dataset = scaler.fit_transform(dataset.values.reshape(-1, 1))

def create_dataset(dataset, look_back):
dataX, dataY = [], []
for i in range(len(dataset)-look_back):
a = dataset[i:(i+look_back)]
dataX.append(a)
dataY.append(dataset[i + look_back])
return numpy.array(dataX),numpy.array(dataY)

look_back = 1
trainX,trainY = create_dataset(dataset,look_back)

trainX = numpy.reshape(trainX, (trainX.shape[0], trainX.shape[1], 1))

model = Sequential()
model.add(LSTM(4, input_shape=(None,1)))
model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='adam')
model.fit(trainX, trainY, epochs=200, batch_size=1, verbose=2)
trainPredict = model.predict(trainX)
trainPredict_ = scaler.inverse_transform(trainPredict)
trainY_ = scaler.inverse_transform(trainY)

import numpy as np
from sklearn.metrics import mean_squared_error,mean_absolute_error
plt.rcParams['font.family'] = 'SimHei'
plt.rcParams['axes.unicode_minus'] = False

```

```

def score(y_true, y_pre):
    print("MAPE :")
    print(mean_absolute_percentage_error(y_true, y_pre))
    print("RMSE :")
    print(np.sqrt(mean_squared_error(y_true, y_pre)))
    print("MAE :")
    print(mean_absolute_error(y_true, y_pre))
    print("R2 :")
    print(np.abs(r2_score(y_true,y_pre)))

score(trainPredict_,trainY_)

plt.plot(data['年份'].values[:-1],trainY_, label='observed data')
plt.plot(data['年份'].values[:-1],trainPredict_, label='LSTM')
plt.xlabel( 'time' )
plt.ylabel( '碳排放 (万吨)' )
plt.title( '碳排放 (万吨)' )
plt.legend()
plt.savefig('./lstm.jpg')
plt.show()

x_input=trainY[:-1]
predict_forword_number=2
predict_list=[]
predict_list.append(x_input)
while len(predict_list) < predict_forword_number:
    x_input = predict_list[-1].reshape((-1, 1, 1))
    yhat = model.predict(x_input, verbose=0)
    predict_list.append(yhat)

scaler.inverse_transform(np.array([ i.reshape(-1,1)[: ,0].tolist() for i in
    predict_list]))[-1]

data['Y']=data['建筑碳排放 (万吨)'].shift(1)
data.dropna(inplace=True)
data
data.to_excel('shift1.xlsx')

import sys
get_ipython().system('{sys.executable} -m pip install sklearn -i
    https://pypi.tuna.tsinghua.edu.cn/simple')
get_ipython().system('{sys.executable} -m pip install xgboost -i
    https://pypi.tuna.tsinghua.edu.cn/simple')
get_ipython().system('{sys.executable} -m pip install lightgbm -i
    https://pypi.tuna.tsinghua.edu.cn/simple')

```

```

import pandas as pd
from numpy import *
import numpy as np
from sklearn.neural_network import MLPRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.linear_model import LinearRegression

import xgboost as xgb
import lightgbm as lgb
import numpy as np

import matplotlib.pyplot as plt
from sklearn.metrics import mean_absolute_error,r2_score

import warnings
warnings.filterwarnings("ignore")
from sklearn.model_selection import train_test_split
def mape(actual, pred):
actual, pred = np.array(actual), np.array(pred)
return np.mean(np.abs((actual - pred) / actual))

tr_x,te_x,tr_y,te_y=train_test_split(data[['建筑碳排放 (万吨)']].values.reshape(-1, 1)
    ,data['Y'].values.reshape(-1, 1),test_size=0.1,random_state=5)

model=MLPRegressor(hidden_layer_sizes=10,max_iter=1000).fit(tr_x,tr_y)

y_pred=model.predict(te_x)
print("神经网络:")
print("训练集平均绝对百分比误差:{:.3f}".format(mape(model.predict(tr_x),tr_y)))
print("测试集平均绝对百分比误差:{:.3f}".format(mape(model.predict(te_x),te_y)))
print("平均绝对误差:",mean_absolute_error(te_y, y_pred))
print("r2_score",r2_score(te_y,y_pred))

print("\n线性回归:")
logreg = LinearRegression()
logreg.fit(tr_x, tr_y)
y_pred = logreg.predict(te_x)
print("训练集平均绝对百分比误差:{:.3f}".format(mape(logreg.predict(tr_x),tr_y)))
print("测试集平均绝对百分比误差:{:.3f}".format(mape(logreg.predict(te_x),te_y)))
print("平均绝对误差:",mean_absolute_error(te_y, y_pred))
print("r2_score",r2_score(te_y,y_pred))

```

```

print("\n随机森林回归:")
rf=RandomForestRegressor(max_depth=20,n_estimators=1000,random_state=0)
rf.fit(tr_x,tr_y)

y_pred = rf.predict(te_x)

print("训练集平均绝对百分比误差:{:.3f}".format(mape(rf.predict(tr_x),tr_y)))
print("测试集平均绝对百分比误差:{:.3f}".format(mape(rf.predict(te_x),te_y)))
print("平均绝对误差:",mean_absolute_error(te_y, y_pred))
print("r2_score",r2_score(te_y,y_pred))

print("\nLGBM回归:")
lgb_model=lgb.LGBMRegressor()
lgb_model.fit(tr_x,tr_y)
y_pred = lgb_model.predict(te_x)
print("训练集平均绝对百分比误差:{:.3f}".format(mape(lgb_model.predict(tr_x),tr_y)))
print("测试集平均绝对百分比误差:{:.3f}".format(mape(lgb_model.predict(te_x),te_y)))
print("平均绝对误差:",mean_absolute_error(te_y, y_pred))
print("r2_score",r2_score(te_y,y_pred))

print("\nXGBOOST回归:")
xgb_model=xgb.XGBRegressor()
xgb_model.fit(tr_x,tr_y)
y_pred = xgb_model.predict(te_x)
print("训练集平均绝对百分比误差:{:.3f}".format(mape(xgb_model.predict(tr_x),tr_y)))
print("测试集平均绝对百分比误差:{:.3f}".format(mape(xgb_model.predict(te_x),te_y)))
print("平均绝对误差:",mean_absolute_error(te_y, y_pred))
print("r2_score",r2_score(te_y,y_pred))

plt.plot(te_y, label='observed data')
plt.plot(y_pred, label='LSTM')
plt.xlabel( 'time')
plt.ylabel( '碳排放 (万吨) ')
plt.title( '碳排放 (万吨) ')
plt.legend()
plt.show()

te_y, y_pred
x_input=tr_y[-1]
predict_forword_number=1
predict_list=[]
predict_list.append(x_input)
while len(predict_list) < predict_forword_number:

```

```
x_input = predict_list[-1]
yhat = xgb_model.predict(x_input)
predict_list
```

参考文献

- [1] 春静, 张智慧. 建筑生命周期碳排放核算 [J]. 工程管理学报, 2010(1):6.
- [2] 春静, 储成龙, 张智慧. 不同结构建筑生命周期的碳排放比较 [J]. 建筑科学, 2011, 27(12):6.
- [3] 燕. 基于全生命周期的建筑碳排放评价模型 [D]. 大连理工大学.
- [4] 骁睿, 屈俊峰, 徐正宏, 等. 一种基于全生命周期碳排放计算的既有建筑改造数字化设计方法:, CN202210114272.9[P]. 2022.
- [5] 依明, 刘念雄. 建筑生命周期评估中碳排放计算的重要意义 [J]. 2020.
- [6] 宇峰, 张道勇, 鲍志东, 等. 利用梯度提升决策树 (GBDT) 预测渗透率——以姬塬油田西部长 4+5 段致密砂岩储层为例 [J]. 地球物理学进展, 2021, 36(2):10.
- [7] 振峰, 汤静远, 常冬霞, 等. 基于 GBDT 的商品分配层次化预测模型 [J]. 北京交通大学学报, 2018, 42(2):6.
- [8] 威威, 李瑞敏, 谢中教. 基于 PCA-GBDT 的城市道路旅行时间预测方法 [J]. 公路工程, 2017, 42(6):6.
- [9] annan D , Jabbour A , Jabbour C . Selecting green suppliers based on GSCM practices: Using fuzzy TOPSIS applied to a Brazilian electronics company[J]. European Journal of Operational Research, 2014, 233(2):432-447.
- [10] f A , Schmidhuber J , Cummins F . Learning to Forget: Continual Prediction with LSTM[C]// Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale. Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale, 1999.
- [11] Imushayti Z . Analyzing the Features Affecting the Performance of Teachers during Covid-19: A Multilevel Feature Selection[J]. Electronics, 2021, 10.