



(21) 申请号 202510390863.2

(22) 申请日 2025.03.31

(71) 申请人 九郡绿建(嘉兴)科技有限公司

地址 314000 浙江省嘉兴市南湖区大桥镇  
亚太路778号6幢301室

(72) 发明人 刘亮俊 刘崇堰 张哲元 魏夏森

(74) 专利代理机构 福州顺升知识产权代理事务  
所(普通合伙) 35242

专利代理师 曹亚辉

(51) Int. Cl.

G06F 18/213 (2023.01)

G06Q 50/06 (2024.01)

G06Q 50/08 (2012.01)

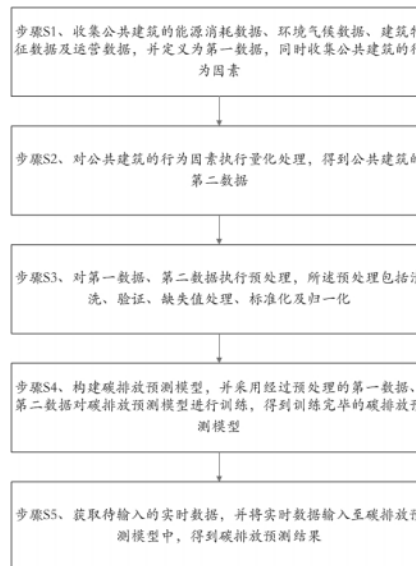
权利要求书2页 说明书8页 附图1页

#### (54) 发明名称

一种基于多源数据处理的公共建筑碳排放  
预测方法

#### (57) 摘要

本发明公开了一种基于多源数据处理的公共建筑碳排放预测方法,涉及碳排放预测技术领域,本发明旨在基于数据驱动的方式有效量化行为因素后,并将其作为预测模型训练所用的额外的一种多源数据,这使得预测模型能够在基于现有静态的建筑特征和环境气候数据等的基础上,进一步学习动态的行为因素,从而能够更全面地考虑建筑物内人员活动和设备使用情况对能耗的影响,进而显著提升预测模型对于公共建筑碳排放预测的准确性。



1. 一种基于多源数据处理的公共建筑碳排放预测方法,其特征在于,包括以下步骤:

步骤S1、收集公共建筑的能源消耗数据、环境气候数据、建筑特征数据及运营数据,并定义为第一数据,同时收集公共建筑的行为因素;

步骤S2、对公共建筑的行为因素执行量化处理,得到公共建筑的第二数据;

步骤S2a、计算公共建筑的每种行为因素的第一得分;

步骤S2b、对每种行为因素按照不同的维度划分为至少两类;

步骤S2c、基于每类行为因素中每种行为因素的第一得分,计算得到每类行为因素的第二得分;

步骤S2d、综合每类行为因素的第二得分,计算得到所有行为因素的综合得分;

步骤S2e、统计每种行为因素的第一得分、每类行为因素的第二得分以及所有行为因素的综合得分,并定义为公共建筑的第二数据;

步骤S3、对第一数据、第二数据执行预处理,所述预处理包括清洗、验证、缺失值处理、标准化及归一化;

步骤S4、构建碳排放预测模型,并采用经过预处理的第一数据、第二数据对碳排放预测模型进行训练,得到训练完毕的碳排放预测模型;

步骤S5、获取待输入的实时数据,并将实时数据输入至碳排放预测模型中,得到碳排放预测结果。

2. 根据权利要求1所述的一种基于多源数据处理的公共建筑碳排放预测方法,其特征在于,所述计算公共建筑的每种行为因素的第一得分,包括:

步骤S2a1、对于每种行为因素,分别定义至少两个基础特征,并基于至少两个基础特征构建每种行为因素的特征向量;

步骤S2a2、基于历史数据,为每种行为因素的特征向量中的每个基础特征定义至少两个影响因子;

步骤S2a3、基于每种行为因素的特征向量中的每个基础特征、每种行为因素的特征向量中的每个基础特征的影响因子,计算得到每种行为因素的特征向量中的每个基础特征的影响值;

步骤S2a4、基于每个基础特征的影响值,采用加权求和的方式计算得到每种行为因素的第一得分。

3. 根据权利要求2所述的一种基于多源数据处理的公共建筑碳排放预测方法,其特征在于,所述每种行为因素的特征向量的表达式为:

$$X_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,k}\};$$

式中, $X_i$ 代表第*i*种行为因素的特征向量; $x_{i,k}$ 代表第*i*种行为因素的第*k*个基础特征;*k*代表第*i*种行为因素中基础特征的数量。

4. 根据权利要求3所述的一种基于多源数据处理的公共建筑碳排放预测方法,其特征在于,所述每种行为因素的特征向量中的每个基础特征的影响因子的表达式为:

$$F_{i,k} = \{f_{i,k,1}, f_{i,k,2}, \dots, f_{i,k,m}\};$$

式中, $F_{i,k}$ 代表第*i*种行为因素的特征向量中第*k*个基础特征的影响因子集合; $f_{i,k,m}$ 代

表第*i*种行为因素的特征向量中第*k*个基础特征的第*m*个影响因子;*m*代表第*i*种行为因素的特征向量中第*k*个基础特征的影响因子数量,且*m*≥2*k*,即每个基础特征至少有两个影响因子。

5.根据权利要求4所述的一种基于多源数据处理的公共建筑碳排放预测方法,其特征在于,所述每种行为因素的特征向量中的每个基础特征的影响值的计算公式为:

$$v_{i,k} = \sum_{j=1}^m f_{i,k,j}(x_{i,k});$$

式中, $v_{i,k}$ 代表第*i*种行为因素的特征向量中的第*k*个基础特征的影响值; $f_{i,k,j}(x_{i,k})$ 代表基于基础特征 $x_{i,k}$ 的第*j*个影响因子的函数;*m*代表第*i*种行为因素的特征向量中第*k*个基础特征的影响因子数量,且*m*≥2*k*,即每个基础特征至少有两个影响因子。

6.根据权利要求5所述的一种基于多源数据处理的公共建筑碳排放预测方法,其特征在于,所述每种行为因素的第一得分的计算公式为:

$$f_{score}(X_i) = \sum_{k=1}^k w_k \cdot v_{i,k};$$

式中, $f_{score}(X_i)$ 代表第*i*种行为因素的第一得分; $w_k$ 代表第*k*个基础特征的权重; $v_{i,k}$ 代表第*i*种行为因素的特征向量中的第*k*个基础特征的影响值。

7.根据权利要求6所述的一种基于多源数据处理的公共建筑碳排放预测方法,其特征在于,所述每类行为因素的第二得分的计算公式为:

$$S_q = \frac{\sum_{i \in q} w_i \cdot f_{score}(X_i)}{\sum_{i \in C_q} w_i};$$

式中, $S_q$ 代表第*q*类行为因素的第二得分; $w_i$ 代表第*i*种行为因素的权重; $f_{score}(X_i)$ 代表第*i*种行为因素的第一得分; $i \in q$ 代表第*i*种行为因素属于第*q*类行为因素。

8.根据权利要求7所述的一种基于多源数据处理的公共建筑碳排放预测方法,其特征在于,所述所有行为因素的综合得分的计算公式为:

$$S_{total} = \frac{\sum_{q=1}^p w_q \cdot S_q}{\sum_{q=1}^p w_q};$$

式中, $S_{total}$ 代表所有行为因素的综合得分; $w_q$ 代表第*q*类行为因素的权重; $S_q$ 代表第*q*类行为因素的第二得分;*p*代表行为因素类别的数量。

## 一种基于多源数据处理的公共建筑碳排放预测方法

### 技术领域

[0001] 本发明涉及碳排放预测技术领域,尤其涉及一种基于多源数据处理的公共建筑碳排放预测方法。

### 背景技术

[0002] 基于多源数据处理的公共建筑碳排放预测方法是一种结合了多种数据来源和技术手段,以实现对公共建筑在使用过程中产生的二氧化碳等温室气体排放量进行准确预估的方法,为了更准确地预测公共建筑的碳排放,现有方法通常遵循一系列具体步骤,以下是这些步骤的详细描述:

[0003] 首先执行数据收集,包括能源消耗数据:通过安装智能仪表或传感器来监测建筑物内不同类型的能源使用情况,包括电力、燃气、热力等,并确保数据能够定时传输至数据中心;环境与气候数据:整合来自本地气象站或第三方API提供的室外温度、湿度、风速、日照时长等信息,为理解建筑能耗提供必要的背景条件;建筑特征数据:记录和整理有关建筑物静态属性的信息,如面积、高度、朝向、建筑材料及其保温性能、建造年代及设计标准等;运营数据:利用楼宇自动化系统(BAS)、物联网(IoT)设备或其他监控手段,捕捉建筑物内的人员活动模式(工作时间、访客流量)、设备运行状态(空调系统、照明系统的启停时刻)等动态信息;

[0004] 随后对收集的数据执行预处理,包括清洗与验证:去除或修正异常值、错误记录以及重复项;确认所有数据点的时间戳正确无误;缺失值处理:采用插补法填补空白数据点,比如平均值填充、最近邻填充或基于模型的估算;标准化与归一化:将各类数据转换成统一格式和单位,便于后续分析;对于数值型变量,可以实施标准化或归一化操作以提高算法效率;

[0005] 接着进行模型选择,往往需要根据问题性质和数据特性,选择适合的建模技术,如线性回归、决策树、随机森林、支持向量机(SVM)、人工神经网络(ANN)、长短期记忆网络(LSTM)等,并在对比不同模型的表现,考量其准确性、计算复杂度、训练时间和解释能力等因素后,选出最适合当前任务的模型,通常采用人工神经网络模型;

[0006] 紧接着进行模型训练,包括划分数据集:将收集的数据分为训练集、验证集和测试集,保证模型既能在已知数据上表现良好,也能在未知数据上有较强的泛化能力;超参数调优:使用网格搜索、随机搜索或贝叶斯优化等方法寻找最优超参数组合,以达到最佳性能;交叉验证:采用K折交叉验证等方法来减少过拟合风险,同时确保模型稳定性和可靠性;

[0007] 最后一旦模型训练完成并通过验证,则部署于实际环境中,用于实时预测未来一段时间内的碳排放趋势。

[0008] 尽管上述方法提供了较为全面的框架来预测公共建筑的碳排放,但在实践中仍存在显著的挑战,特别是以下方面:

[0009] 人的行为习惯如开窗通风、调温设置等对建筑物能耗有着重要影响,然而这部分非结构化信息却被忽视,且并未有效纳入至现有预测模型中,这在一定程度上降低了现有

预测模型对于公共建筑碳排放预测的准确性。

[0010] 因此,现有技术急需一种基于多源数据处理的公共建筑碳排放预测方法的技术方案。

### 发明内容

[0011] 为了解决上述技术问题,本发明提供了一种基于多源数据处理的公共建筑碳排放预测方法,具体包括以下步骤:

[0012] 步骤S1、收集公共建筑的能源消耗数据、环境气候数据、建筑特征数据及运营数据,并定义为第一数据,同时收集公共建筑的行为因素;

[0013] 步骤S2、对公共建筑的行为因素执行量化处理,得到公共建筑的第二数据;

[0014] 步骤S2a、计算公共建筑的每种行为因素的第一得分;

[0015] 步骤S2a1、对于每种行为因素,分别定义至少两个基础特征,并基于至少两个基础特征构建每种行为因素的特征向量;

[0016] 其中,每种行为因素的特征向量的表达式为:

$$X_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,k}\};$$

[0017] 式中, $X_i$ 代表第*i*种行为因素的特征向量; $x_{i,k}$ 代表第*i*种行为因素的第*k*个基础特征;*k*代表第*i*种行为因素中基础特征的数量;

[0018] 步骤S2a2、基于历史数据,为每种行为因素的特征向量中的每个基础特征定义至少两个影响因子;

[0019] 其中,每种行为因素的特征向量中的每个基础特征的影响因子的表达式为:

$$F_{i,k} = \{f_{i,k,1}, f_{i,k,2}, \dots, f_{i,k,m}\};$$

[0020] 式中, $F_{i,k}$ 代表第*i*种行为因素的特征向量中第*k*个基础特征的影响因子集合; $f_{i,k,m}$ 代表第*i*种行为因素的特征向量中第*k*个基础特征的第*m*个影响因子;*m*代表第*i*种行为因素的特征向量中第*k*个基础特征的影响因子数量,且 $m \geq 2k$ ,即每个基础特征至少有两个影响因子;

[0021] 步骤S2a3、基于每种行为因素的特征向量中的每个基础特征、每种行为因素的特征向量中的每个基础特征的影响因子,计算得到每种行为因素的特征向量中的每个基础特征的影响值;

[0022] 其中,每种行为因素的特征向量中的每个基础特征的影响值的计算公式为:

$$v_{i,k} = \sum_{j=1}^m f_{i,k,j}(x_{i,k});$$

[0023] 式中, $v_{i,k}$ 代表第*i*种行为因素的特征向量中的第*k*个基础特征的影响值; $f_{i,k,j}(x_{i,k})$ 代表基于基础特征 $x_{i,k}$ 的第*j*个影响因子的函数;*m*代表第*i*种行为因素的特征向量中第*k*个基础特征的影响因子数量,且 $m \geq 2k$ ,即每个基础特征至少有两个影响因子;

[0024] 步骤S2a4、基于每个基础特征的影响值,采用加权求和的方式计算得到每种行为因素的第一得分;

[0025] 其中,每种行为因素的第一得分的计算公式为:

$$f_{\text{score}}(X_i) = \sum_{k=1}^k W_k \cdot v_{i,k};$$

[0026] 式中,  $f_{\text{score}}(X_i)$ 代表第*i*种行为因素的第一得分;  $W_k$ 代表第*k*个基础特征的权重;  $v_{i,k}$ 代表第*i*种行为因素的特征向量中的第*k*个基础特征的影响值;

[0027] 步骤S2b、对每种行为因素按照不同的维度划分为至少两类;

[0028] 步骤S2c、基于每类行为因素中每种行为因素的第一得分,计算得到每类行为因素的第二得分;

[0029] 其中,每类行为因素的第二得分的计算公式为:

$$S_q = \frac{\sum_{i \in q} W_i \cdot f_{\text{score}}(X_i)}{\sum_{i \in C_q} W_i};$$

[0030] 式中,  $S_q$ 代表第*q*类行为因素的第二得分;  $W_i$ 代表第*i*种行为因素的权重;

$f_{\text{score}}(X_i)$ 代表第*i*种行为因素的第一得分;  $i \in q$ 代表第*i*种行为因素属于第*q*类行为因素;

[0031] 步骤S2d、综合每类行为因素的第二得分,计算得到所有行为因素的综合得分;

[0032] 其中,所有行为因素的综合得分的计算公式为:

$$S_{\text{total}} = \frac{\sum_{q=1}^p W_q \cdot S_q}{\sum_{q=1}^p W_q};$$

[0033] 式中,  $S_{\text{total}}$ 代表所有行为因素的综合得分;  $W_q$ 代表第*q*类行为因素的权重;  $S_q$ 代表第*q*类行为因素的第二得分;  $p$ 代表行为因素类别的数量;

[0034] 步骤S2e、统计每种行为因素的第一得分、每类行为因素的第二得分以及所有行为因素的综合得分,并定义为公共建筑的第二数据;

[0035] 步骤S3、对第一数据、第二数据执行预处理,所述预处理包括清洗、验证、缺失值处理、标准化及归一化;

[0036] 步骤S4、构建碳排放预测模型,并采用经过预处理的第一数据、第二数据对碳排放预测模型进行训练,得到训练完毕的碳排放预测模型;

[0037] 步骤S5、获取待输入的实时数据,并将实时数据输入至碳排放预测模型中,得到碳排放预测结果。

[0038] 本发明实施例具有以下技术效果:

[0039] 本发明旨在基于数据驱动的方式有效量化行为因素后,并将其作为预测模型训练所用的额外的一种多源数据,这使得预测模型能够在基于现有静态的建筑特征和环境气候数据等的基础上,进一步学习动态的行为因素,从而能够更全面地考虑建筑物内人员活动和设备使用情况对能耗的影响,进而显著提升预测模型对于公共建筑碳排放预测的准确性。

## 附图说明

[0040] 为了更清楚地说明本发明具体实施方式或现有技术中的技术方案,下面将对具体实施方式或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图是本发明的一些实施方式,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其它的附图。

[0041] 图1是本发明实施例提供的一种基于多源数据处理的公共建筑碳排放预测方法的流程图。

### 具体实施方式

[0042] 为使本发明的目的、技术方案和优点更加清楚,下面将对本发明的技术方案进行清楚、完整的描述。显然,所描述的实施例仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动的前提下所得到的所有其它实施例,都属于本发明所保护的范围。

[0043] 实施例一:如图1所示,本发明提供了的一种基于多源数据处理的公共建筑碳排放预测方法,包括如下步骤:

[0044] 步骤S1、收集公共建筑的能源消耗数据、环境气候数据、建筑特征数据及运营数据,并定义为第一数据,同时收集公共建筑的行为因素;

[0045] 其中关于采用何种手段进行收集公共建筑的能源消耗数据、环境气候数据、建筑特征数据及运营数据在此不做过多赘述,着重阐述如何收集公共建筑的行为因素,首先可通过在建筑物内部署行为传感器,如红外传感器、摄像头、门禁系统等,记录人员的活动轨迹、开窗通风、调温设置等行为;其次,可定期向建筑物用户发放问卷,了解他们的日常行为模式,如工作时间、休息时间、使用设备的习惯等;亦或者可开发移动应用程序,用户可以通过应用程序报告自己的行为,如开关灯、调整空调温度等。

[0046] 步骤S2、对公共建筑的行为因素执行量化处理,得到公共建筑的第二数据;

[0047] 步骤S2a、计算公共建筑的每种行为因素的第一得分;

[0048] 步骤S2a1、对于每种行为因素,分别定义至少两个基础特征,并基于至少两个基础特征构建每种行为因素的特征向量;

[0049] 值得说明的是,关于定义行为因素的基础特征的手段包括但不限于:一、通过对既有数据采用聚类分析法进行分析,从而识别出与能耗高度相关的特征,以作为行为因素的基础特征;二、采用传统的问卷调查方式,了解建筑物用户的行为模式,以从中提炼出常见的行为特征,并作为行为因素的基础特征。

[0050] 对于公共建筑中的常见行为因素,可以定义以下基础特征但不限于:

[0051] 一、时间戳:记录行为发生的具体时间,包括日期、小时、分钟等,主要作用在于不同时间段的行为对能耗的影响不同,例如工作日与休息日、白天与夜晚等;二、频率:行为发生的频次,如每天几次、每周几次等,主要作用在于高频行为可能比低频行为对能耗的影响更大,例如频繁开关灯或调整空调温度;三、持续时间:每次行为的持续时长,如每次开窗通风的时间、每次使用设备的时间等,主要作用在于长时间的行为通常会消耗更多能源,例如长时间开启空调或照明;四、强度:行为的强度或力度,如空调温度设定值的变化幅度、灯光亮度等,主要作用在于高强度行为可能导致更高的能耗,例如快速调整空调温度或使用高功率设备。

[0052] 其中,每种行为因素的特征向量的表达式为:

$$X_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,k}\};$$

[0053] 式中, $X_i$ 代表第*i*种行为因素的特征向量; $x_{i,k}$ 代表第*i*种行为因素的第*k*个基础特

征; $k$ 代表第*i*种行为因素中基础特征的数量;

[0054] 步骤S2a2、基于历史数据,为每种行为因素的特征向量中的每个基础特征定义至少两个影响因子;

[0055] 值得说明的是,基于历史数据定义影响因子的手段主要通过回归模型分析历史数据,找出每个基础特征与能耗之间的定量关系,从而确定影响因子;

[0056] 对于上述基础特征,可以定义以下影响因子:

[0057] 时间戳的影响因子:影响因子1:一天中的时间段早晨、下午、晚上,因为不同时间段的能耗模式不同;影响因子2:工作日与休息日的区别,因为人们的行为模式在工作日和休息日有较大差异;

[0058] 频率的影响因子:影响因子1:行为发生的频次如每天几次,高频行为可能导致更高的能耗;影响因子2:行为的累积次数,长期来看,累积次数越多,总能耗越高;

[0059] 持续时间的影响因子:影响因子1:单次行为的持续时长,长时间的行为通常消耗更多能源;影响因子2:行为的累积时长,即一段时间内累计的行为时长;

[0060] 强度的影响因子:影响因子1:行为的强度如温度设定值的变化幅度,高强度行为可能导致更高的能耗;影响因子2:行为的剧烈程度如快速调整温度与缓慢调整温度的区别,剧烈行为可能导致更高的瞬时能耗。

[0061] 其中,每种行为因素的特征向量中的每个基础特征的影响因子的表达式为:

$$F_{i,k} = \{f_{i,k,1}, f_{i,k,2}, \dots, f_{i,k,m}\};$$

[0062] 式中, $F_{i,k}$ 代表第*i*种行为因素的特征向量中第*k*个基础特征的影响因子集合; $f_{i,k,m}$ 代表第*i*种行为因素的特征向量中第*k*个基础特征的第*m*个影响因子; $m$ 代表第*i*种行为因素的特征向量中第*k*个基础特征的影响因子数量,且 $m \geq 2k$ ,即每个基础特征至少有两个影响因子;

[0063] 步骤S2a3、基于每种行为因素的特征向量中的每个基础特征、每种行为因素的特征向量中的每个基础特征的影响因子,计算得到每种行为因素的特征向量中的每个基础特征的影响值;

[0064] 其中,每种行为因素的特征向量中的每个基础特征的影响值的计算公式为:

$$v_{i,k} = \sum_{j=1}^m f_{i,k,j}(x_{i,k});$$

[0065] 式中, $v_{i,k}$ 代表第*i*种行为因素的特征向量中的第*k*个基础特征的影响值; $f_{i,k,j}(x_{i,k})$ 代表基于基础特征 $x_{i,k}$ 的第*j*个影响因子的函数; $m$ 代表第*i*种行为因素的特征向量中第*k*个基础特征的影响因子数量,且 $m \geq 2k$ ,即每个基础特征至少有两个影响因子;

[0066] 步骤S2a4、基于每个基础特征的影响值,采用加权求和的方式计算得到每种行为因素的第一得分;

[0067] 其中,每种行为因素的第一得分的计算公式为:

$$f_{\text{score}}(X_i) = \sum_{k=1}^k w_k \cdot v_{i,k};$$

[0068] 式中, $f_{\text{score}}(X_i)$ 代表第*i*种行为因素的第一得分; $w_k$ 代表第*k*个基础特征的权重; $v_{i,k}$ 代表第*i*种行为因素的特征向量中的第*k*个基础特征的影响值;

[0069] 步骤S2b、对每种行为因素按照不同的维度划分为至少两类；

[0070] 值得说明的是，其中，不同的维度可以是根据行为因素的时间性如工作日/休息日、空间性如不同房间或区域、类型如办公、会议、休闲等等维度进行划分。

[0071] 步骤S2c、基于每类行为因素中每种行为因素的第一得分，计算得到每类行为因素的第二得分；

[0072] 其中，每类行为因素的第二得分的计算公式为：

$$S_q = \frac{\sum_{i \in q} w_i \cdot f_{\text{score}}(X_i)}{\sum_{i \in C_q} w_i};$$

[0073] 式中， $S_q$ 代表第 $q$ 类行为因素的第二得分； $w_i$ 代表第 $i$ 种行为因素的权重；

$f_{\text{score}}(X_i)$ 代表第 $i$ 种行为因素的第一得分； $i \in q$ 代表第 $i$ 种行为因素属于第 $q$ 类行为因素；

[0074] 步骤S2d、综合每类行为因素的第二得分，计算得到所有行为因素的综合得分；

[0075] 其中，所有行为因素的综合得分的计算公式为：

$$S_{\text{total}} = \frac{\sum_{q=1}^p W_q \cdot S_q}{\sum_{q=1}^p W_q};$$

[0076] 式中， $S_{\text{total}}$ 代表所有行为因素的综合得分； $W_q$ 代表第 $q$ 类行为因素的权重； $S_q$ 代表第 $q$ 类行为因素的第二得分； $p$ 代表行为因素类别的数量；

[0077] 值得说明的是，上述中，关于每个基础特征的权重、每种行为因素的权重以及每类行为因素的权重的确定方式均相同，以其中第 $k$ 个基础特征的权重的确定方式为例，首先调用第 $k$ 个基础特征的影响值，随后统计并累加第 $i$ 种行为因素的特征向量中所有基础特征的影响值，得到总的影响值，接着再将第 $k$ 个基础特征的影响值于该总的影响值进行比值处理，即可得到该第 $k$ 个基础特征的权重，以此类推，其中每种行为因素的权重以及每类行为因素的权重的确定方式与上述每个基础特征的权重的确定方式一致。

[0078] 步骤S2e、统计每种行为因素的第一得分、每类行为因素的第二得分以及所有行为因素的综合得分，并定义为公共建筑的第二数据；

[0079] 值得说明的是，本申请之所以在计算每种行为因素的第一得分后，还需要对行为因素进行分类，并进一步计算每类行为因素的第二得分和所有行为因素的综合得分，这是由于考虑到行为因素是复杂且多维的，直接将第一得分输入模型可能会丢失一些重要的结构信息，然而通过将行为因素分类并计算第二得分和综合得分，可以为模型提供更丰富的层次化表示，每一层级的数据都反映了不同粒度的行为特征，其中第一得分反映了单个行为因素的影响，第二得分则反映了同一类别内多种行为因素的综合影响，综合得分则是有效反映了所有行为因素的整体影响，这种多层次的数据表示有助于模型更全面地理解行为因素对能耗的影响，从而有效提高碳排放预测的精度；

[0080] 此外，由于公共建筑中的行为因素具有多样性和动态性，直接使用第一得分无法充分捕捉这些特性，然而通过分类和计算第二得分、综合得分，模型可以学习到不同类别行为因素之间的关联和模式，这不仅增强了模型的泛化能力，使其能够更好地应对未见过的数据，还提高了模型在不同场景下的适应性，例如，在工作日和休息日，办公楼内的人员活动模式差异很大，通过分类处理可以确保模型能够区分这两种情况，从而更准确地预测不同时间段的碳排放变化。

[0081] 步骤S3、对第一数据、第二数据执行预处理,所述预处理包括清洗、验证、缺失值处理、标准化及归一化;

[0082] 步骤S4、构建碳排放预测模型,并采用经过预处理的第一数据、第二数据对碳排放预测模型进行训练,得到训练完毕的碳排放预测模型;

[0083] 值得说明的是,在碳排放预测模型的训练阶段中,其训练过程与现有模型的训练过程基本一致,主要区别在于增加了经过量化处理的行为因素,即第二数据,因此关于碳排放预测模型的训练过程在此不做过多赘述;在具体的训练过程中,首先选择长短期记忆网络模型作为主要模型架构,因为它能够有效处理时间序列数据,并且具备良好的动态行为捕捉能力,接着进行数据准备,包括能源消耗数据、环境气候数据、建筑特征数据及运营数据所组成的第一数据,以及基于多源数据处理的方法,对公共建筑的行为因素执行量化处理,得到每种行为因素的第一得分、每类行为因素的第二得分以及所有行为因素的综合得分所组成的第二数据,由于第一、第二数据已经在步骤S3中执行过预处理的一系列操作,接下来需要将第一、第二数据分为三个部分以供模型训练等,三部分包括训练集:第一部分为70%的数据用于训练模型,包括:能源消耗数据、环境气候数据、建筑特征数据、运营数据,以及行为因素数据,即第二数据,第二部分为验证集,即15%的数据用于调整超参数和监控模型性能,数据类型同上,第三部分为测试集,即15%的数据用于最终评估模型性能,数据类型同上;随后进行模型初始化,该过程首先应定义层数与单元数,例如,定义一个包含两层LSTM单元的模型,每层包含64个LSTM单元,接着初始化权重,即使用Xavier或He初始化方法来初始化模型权重,以确保初始状态的良好分布,随后定义用于衡量模型预测值与真实值之间的差异的损失函数,计算公式为:
$$MSE = \frac{1}{N} \sum_{b=1}^N (T_b - P_b)$$
;式中,MSE代表模型预测值与真实值之间的差异值;N代表样本总数,即数据集中的样本数量; $T_b$ 代表第b个样本的真实值; $P_b$ 代表第b个样本的预测值。

[0084] 该损失函数是模型训练的核心目标函数,其值越小,表示模型预测值越接近真实值;

[0085] 随后执行超参数调优,具体方法为采用网格搜索法,尝试不同的超参数组合,如学习率(0.001, 0.01, 0.1)、批次大小(32, 64, 128)、隐藏层单元数(32, 64, 128)等或采用随机搜索法,在大范围内随机选择超参数组合进行测试等;其中超参数调优的目的在于为了找到最优的模型配置,从而提高模型的预测精度和泛化能力,常见的超参数包括学习率、批次大小、隐藏层单元数、正则化参数等;紧接着执行模型训练,具体地,首先将训练数据分批输入模型,逐步更新模型参数,直到达到设定的迭代次数或收敛条件。每个批次大小应根据硬件资源调整,如32或64,随后应定期记录损失值和验证集上的性能指标,确保模型不出现过拟合或欠拟合现象,具体应通过绘制损失曲线和准确率曲线来直观展示模型的学习过程,最后当损失函数值在训练集和验证集上趋于稳定即变化幅度小于某个阈值,如0.001时,则认为模型已经收敛,如果模型未完全收敛但已达到预设的最大迭代次数如1000次,则停止训练,即模型训练完毕,随后还需采用K折交叉验证法,以减少因数据划分导致的偶然性,提高模型的泛化能力和稳定性,具体地,通过将数据集分成K份,每次用K-1份作为训练集,剩下的一份作为验证集,重复K次,平均结果作为最终评估指标,常用的K值为5或10,其中验证时采用的评估指标包括但不限于平均绝对误差、决定系数等;最后,当模型在

验证集和测试集上表现良好时,将其保存为文件格式如.h5或.pkl,以便后续加载和部署。同时这一新增的数据类型为模型提供了更丰富的动态信息,增强了模型的表达能力和泛化能力,减少了噪声和冗余,从而使预测结果更加准确和可靠。

[0086] 步骤S5、获取待输入的实时数据,并将实时数据输入至碳排放预测模型中,得到碳排放预测结果。

[0087] 值得说明的是,在将实时数据输入至碳排放预测模型中时,需要优先对其中的实时行为因素执行量化处理,随后才能够将经过量化处理的实时行为因素同其它实时数据进行组合,形成输入数据,并输入至碳排放预测模型中,并由碳排放预测模型进行预测并得到碳排放预测结果,其中关于其它实时数据指的是新获取的能源消耗数据、环境气候数据等;关于碳排放预测结果可以是未来时间段内的碳排放趋势,如预测未来24小时、7天或一个月内建筑物的二氧化碳排放量变化趋势;以及具体数值预测,即提供具体的碳排放量数值,如每小时、每天或每月的预计排放量。

[0088] 需要说明的是,本发明所用术语仅为了描述特定实施例,而非限制本申请范围。如本发明说明书中所示,除非上下文明确提示例外情形,“一”、“一个”、“一种”和/或“该”等词并非特指单数,也可包括复数。术语“包括”、“包含”或者其任何其它变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法或者设备不仅包括那些要素,而且还包括没有明确列出的其它要素,或者是还包括为这种过程、方法或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法或者设备中还存在另外的相同要素。

[0089] 还需说明的是,术语“中心”、“上”、“下”、“左”、“右”、“竖直”、“水平”、“内”、“外”等指示的方位或位置关系为基于附图所示的方位或位置关系,仅是为了便于描述本发明和简化描述,而不是指示或暗示所指的装置或元件必须具有特定的方位、以特定的方位构造和操作,因此不能理解为对本发明的限制。除非另有明确的规定和限定,术语“安装”、“相连”、“连接”等应做广义理解,例如,可以是固定连接,也可以是可拆卸连接,或一体地连接;可以是机械连接,也可以是电连接;可以是直接相连,也可以通过中间媒介间接相连,可以是两个元件内部的连通。对于本领域的普通技术人员而言,可以具体情况理解上述术语在本发明中的具体含义。

[0090] 最后应说明的是:以上各实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照前述各实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分或者全部技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本发明各实施例技术方案。

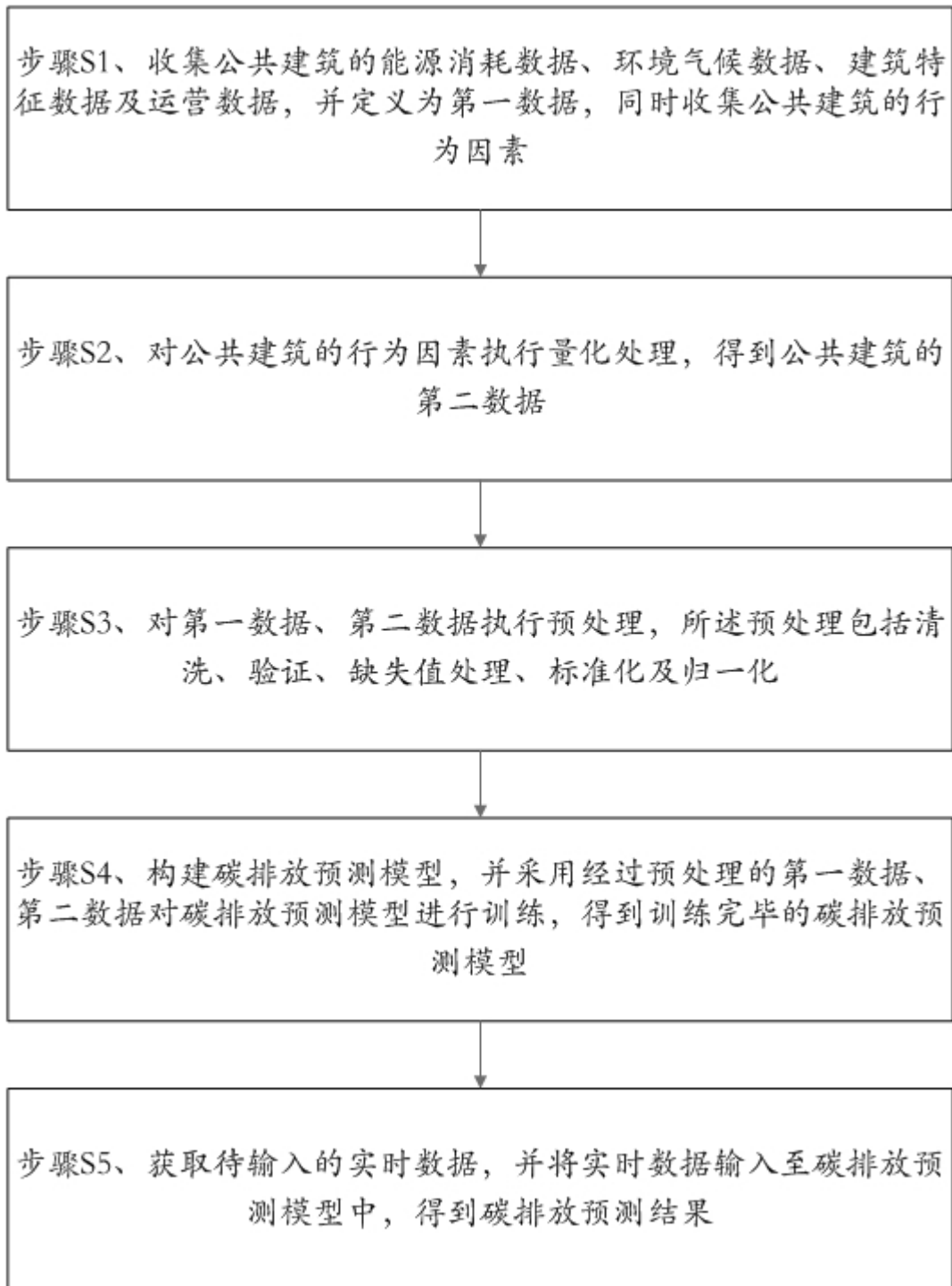


图 1