



(21) 申请号 202411576832.8

(22) 申请日 2024.11.06

(71) 申请人 深圳市睿智通科技有限公司

地址 518000 广东省深圳市福田区莲花街
道紫荆社区深南大道6006号华丰大厦
606

(72) 发明人 杨宏强 吴伟

(51) Int. Cl.

G06Q 10/04 (2023.01)

G06Q 10/067 (2023.01)

G06Q 50/08 (2012.01)

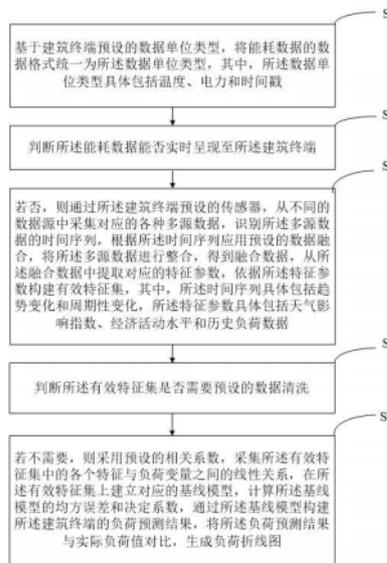
权利要求书3页 说明书24页 附图2页

(54) 发明名称

一种基于负荷曲线的节能优化控制方法及系统

(57) 摘要

本发明提供了一种基于负荷曲线的节能优化控制方法及系统,运用于能耗数据处理领域;本发明通过从不同数据源采集多源数据,可以形成更全面的能耗视图,整合这些数据能够提供更丰富的上下文,帮助识别影响能耗的各种因素,提高节能优化的准确性,同时提取特征参数使得模型能够基于多个维度进行分析,这些特征可以揭示不同因素如何影响能耗,帮助更好地理解 and 预测负荷变化,并且在有效特征集上建立基线模型,并计算均方误差和决定系数,能够量化模型的性能,不仅为能耗预测提供了基准,还能通过与实际数据的比较来不断优化模型参数,提升预测准确性。



1. 一种基于负荷曲线的节能优化控制方法,其特征在于,包括以下步骤:

基于建筑终端预设的数据单位类型,将能耗数据的数据格式统一为所述数据单位类型,其中,所述数据单位类型具体包括温度、电力和时间戳;

判断所述能耗数据能否实时呈现至所述建筑终端;

若否,则通过所述建筑终端预设的传感器,从不同的数据源中采集对应的各种多源数据,识别所述多源数据的时间序列,根据所述时间序列应用预设的数据融合,将所述多源数据进行整合,得到融合数据,从所述融合数据中提取对应的特征参数,依据所述特征参数构建有效特征集,其中,所述时间序列具体包括趋势变化和周期性变化,所述特征参数具体包括天气影响指数、经济活动水平和历史负荷数据;

判断所述有效特征集是否需要预设的数据清洗;

若不需要,则采用预设的相关系数,采集所述有效特征集中的各个特征与负荷变量之间的线性关系,在所述有效特征集上建立对应的基线模型,计算所述基线模型的均方误差和决定系数,通过所述基线模型构建所述建筑终端的负荷预测结果,将所述负荷预测结果与实际负荷值对比,生成负荷折线图。

2. 根据权利要求1所述的基于负荷曲线的节能优化控制方法,其特征在于,所述则通过所述建筑终端预设的传感器,从不同的数据源中采集对应的各种多源数据的步骤前,还包括:

基于所述传感器预设的工作范围,通过所述建筑终端划分所述传感器在建筑内覆盖的采集区域;

判断所述采集区域能否检测到预设的重叠监测边缘;

若能,则根据所述重叠监测边缘构建重叠时间窗口,标识多个传感器在所述重叠时间窗口的采集测量值,依据所述采集测量值生成待融合的数据特征,应用预设的卡尔曼滤波为所述采集测量值进行权重分配,计算所述数据特征的加权平均值,其中,所述数据特征具体包括温度、湿度和能耗。

3. 根据权利要求1所述的基于负荷曲线的节能优化控制方法,其特征在于,所述依据所述特征参数构建有效特征集的步骤中,还包括:

基于预设的虚拟变量,构建所述建筑终端对能耗的季节性特征,计算所述季节性特征之间的比率特征和差值特征,其中,所述虚拟变量具体包括春季、夏季、秋季和冬季,所述比率特征具体为能耗与温度的比率,所述差值特征具体为当前负荷与过去负荷的差值;

判断所述比率特征和所述差值特征是否存在缺失值;

若否,则对所述比率特征和所述差值特征进行标准化处理,确保所述比率特征和所述差值特征始终处于同一尺度上,将所述比率特征和所述差值特征与其他特征进行组合,形成复合特征,根据所述复合特征捕捉季节性变化对应的能耗影响。

4. 根据权利要求1所述的基于负荷曲线的节能优化控制方法,其特征在于,所述在所述有效特征集上建立对应的基线模型的步骤中,还包括:

基于所述建筑终端对能耗的预设图表,从所述图表上绘制X轴为能耗预测值,并从所述图表上绘制Y轴为能耗残差,结合得到能耗散点图;

判断所述能耗散点图上的残差是否随机分布;

若是,则将所述能耗散点图输入至所述基线模型,通过所述基线模型进行交叉验证,获

取所述基线模型在不同数据子集上的波动信息。

5. 根据权利要求1所述的基于负荷曲线的节能优化控制方法,其特征在於,所述判断所述能耗数据能否实时呈现至所述建筑终端的步骤中,还包括:

基于所述建筑终端预设的网络带宽,检测所述能耗数据进行数据传输时的网络延迟信息;

判断所述网络延迟信息是否超出预设的延迟时段;

若是,则在所述能耗数据的数据采集环节和数据展示环节中添加时间戳标签,根据所述时间戳标签计算所述能耗数据的整体延迟。

6. 根据权利要求1所述的基于负荷曲线的节能优化控制方法,其特征在於,所述判断所述有效特征集是否需要预设的数据清洗的步骤中,还包括:

基于所述有效特征集的预统计特征,计算特征缺失值的比例,根据所述比例获取所述特征缺失值的分布,生成对应的缺失值热图;

判断所述缺失值热图是否匹配预设的随机缺失;

若否,则识别所述特征缺失值与所述预统计特征的关联信息,依据所述关联信息采集对应的缺失出处。

7. 根据权利要求1所述的基于负荷曲线的节能优化控制方法,其特征在於,所述基于建筑终端预设的数据单位类型,将能耗数据的数据格式统一为所述数据单位类型的步骤中,还包括:

检测所述能耗数据中的原始单位,基于所述原始单位识别出待转换的不同单位;

判断所述不同单位能否统一转换为所述数据单位类型;

若否,则标记所述不同单位的不一致项,根据所述不一致项对所述不同单位的对应数据逐行进行单位转换,比对转换前后的数据分布和统计特性。

8. 一种基于负荷曲线的节能优化控制系统,其特征在於,包括:

统一模块,用于基于建筑终端预设的数据单位类型,将能耗数据的数据格式统一为所述数据单位类型,其中,所述数据单位类型具体包括温度、电力和时间戳;

判断模块,用于判断所述能耗数据能否实时呈现至所述建筑终端;

执行模块,用于若否,则通过所述建筑终端预设的传感器,从不同的数据源中采集对应的各种多源数据,识别所述多源数据的时间序列,根据所述时间序列应用预设的数据融合,将所述多源数据进行整合,得到融合数据,从所述融合数据中提取对应的特征参数,依据所述特征参数构建有效特征集,其中,所述时间序列具体包括趋势变化和周期性变化,所述特征参数具体包括天气影响指数、经济活动水平和历史负荷数据;

第二判断模块,用于判断所述有效特征集是否需要预设的数据清洗;

第二执行模块,用于若不需要,则采用预设的相关系数,采集所述有效特征集中的各个特征与负荷变量之间的线性关系,在所述有效特征集上建立对应的基线模型,计算所述基线模型的均方误差和决定系数,通过所述基线模型构建所述建筑终端的负荷预测结果,将所述负荷预测结果与实际负荷值对比,生成负荷折线图。

9. 根据权利要求8所述的基于负荷曲线的节能优化控制系统,其特征在於,还包括:

划分模块,用于基于所述传感器预设的工作范围,通过所述建筑终端划分所述传感器在建筑内覆盖的采集区域;

第三判断模块,用于判断所述采集区域能否检测到预设的重叠监测边缘;

第三执行模块,用于若能,则根据所述重叠监测边缘构建重叠时间窗口,标识多个传感器在所述重叠时间窗口的采集测量值,依据所述采集测量值生成待融合的数据特征,应用预设的卡尔曼滤波为所述采集测量值进行权重分配,计算所述数据特征的加权平均值,其中,所述数据特征具体包括温度、湿度和能耗。

10.根据权利要求8所述的基于负荷曲线的节能优化控制系统,其特征在于,所述执行模块还包括:

计算单元,用于基于预设的虚拟变量,构建所述建筑终端对能耗的季节性特征,计算所述季节性特征之间的比率特征和差值特征,其中,所述虚拟变量具体包括春季、夏季、秋季和冬季,所述比率特征具体为能耗与温度的比率,所述差值特征具体为当前负荷与过去负荷的差值;

判断单元,用于判断所述比率特征和所述差值特征是否存在缺失值;

执行单元,用于若否,则对所述比率特征和所述差值特征进行标准化处理,确保所述比率特征和所述差值特征始终处于同一尺度上,将所述比率特征和所述差值特征与其他特征进行组合,形成复合特征,根据所述复合特征捕捉季节性变化对应的能耗影响。

一种基于负荷曲线的节能优化控制方法及系统

技术领域

[0001] 本发明涉及能耗数据处理领域,特别涉及为一种基于负荷曲线的节能优化控制方法及系统。

背景技术

[0002] 随着科技的发展,建筑内设有的智能设备越来越多,最常见的就是仪表(电表、水表、燃气表、能量表、PM2.5仪表等),往往智能仪表不是单独工作的,需要配置数据管理软件来对数据进行分析。

[0003] 由于不同来源的数据(如电力消耗、天气数据、市场需求等)格式和结构各异,需要建立统一的标准和框架进行整合,而多源数据可能存在缺失、错误或不一致,影响融合后的数据可靠性,因此需要有效的数据清洗和验证机制。

发明内容

[0004] 本发明旨在解决如何有效整合来自不同来源的数据,以优化能耗负荷预测的问题,提供一种基于负荷曲线的节能优化控制方法及系统。

[0005] 本发明为解决技术问题采用如下技术手段:

本发明提供一种基于负荷曲线的节能优化控制方法,包括:

基于建筑终端预设的数据单位类型,将能耗数据的数据格式统一为所述数据单位类型,其中,所述数据单位类型具体包括温度、电力和时间戳;

判断所述能耗数据能否实时呈现至所述建筑终端;

若否,则通过所述建筑终端预设的传感器,从不同的数据源中采集对应的各种多源数据,识别所述多源数据的时间序列,根据所述时间序列应用预设的数据融合,将所述多源数据进行整合,得到融合数据,从所述融合数据中提取对应的特征参数,依据所述特征参数构建有效特征集,其中,所述时间序列具体包括趋势变化和周期性变化,所述特征参数具体包括天气影响指数、经济活动水平和历史负荷数据;

判断所述有效特征集是否需要预设的数据清洗;

若不需要,则采用预设的相关系数,采集所述有效特征集中的各个特征与负荷变量之间的线性关系,在所述有效特征集上建立对应的基线模型,计算所述基线模型的均方误差和决定系数,通过所述基线模型构建所述建筑终端的负荷预测结果,将所述负荷预测结果与实际负荷值对比,生成负荷折线图。

[0006] 进一步地,所述则通过所述建筑终端预设的传感器,从不同的数据源中采集对应的各种多源数据的步骤前,还包括:

基于所述传感器预设的工作范围,通过所述建筑终端划分所述传感器在建筑内覆盖的采集区域;

判断所述采集区域能否检测到预设的重叠监测边缘;

若能,则根据所述重叠监测边缘构建重叠时间窗口,标识多个传感器在所述重叠

时间窗口的采集测量值,依据所述采集测量值生成待融合的数据特征,应用预设的卡尔曼滤波为所述采集测量值进行权重分配,计算所述数据特征的加权平均值,其中,所述数据特征具体包括温度、湿度和能耗。

[0007] 进一步地,所述依据所述特征参数构建有效特征集的步骤中,还包括:

基于预设的虚拟变量,构建所述建筑终端对能耗的季节性特征,计算所述季节性特征之间的比率特征和差值特征,其中,所述虚拟变量具体包括春季、夏季、秋季和冬季,所述比率特征具体为能耗与温度的比率,所述差值特征具体为当前负荷与过去负荷的差值;

判断所述比率特征和所述差值特征是否存在缺失值;

若否,则对所述比率特征和所述差值特征进行标准化处理,确保所述比率特征和所述差值特征始终处于同一尺度上,将所述比率特征和所述差值特征与其他特征进行组合,形成复合特征,根据所述复合特征捕捉季节性变化对应的能耗影响。

[0008] 进一步地,所述在所述有效特征集上建立对应的基线模型的步骤中,还包括:

基于所述建筑终端对能耗的预设图表,从所述图表上绘制X轴为能耗预测值,并从所述图表上绘制Y轴为能耗残差,结合得到能耗散点图;

判断所述能耗散点图上的残差是否随机分布;

若是,则将所述能耗散点图输入至所述基线模型,通过所述基线模型进行交叉验证,获取所述基线模型在不同数据子集上的波动信息。

[0009] 进一步地,所述判断所述能耗数据能否实时呈现至所述建筑终端的步骤中,还包括:

基于所述建筑终端预设的网络带宽,检测所述能耗数据进行数据传输时的网络延迟信息;

判断所述网络延迟信息是否超出预设的延迟时段;

若是,则在所述能耗数据的数据采集环节和数据展示环节中添加时间戳标签,根据所述时间戳标签计算所述能耗数据的整体延迟。

[0010] 进一步地,所述判断所述有效特征集是否需要预设的数据清洗的步骤中,还包括:

基于所述有效特征集的预统计特征,计算特征缺失值的比例,根据所述比例获取所述特征缺失值的分布,生成对应的缺失值热图;

判断所述缺失值热图是否匹配预设的随机缺失;

若否,则识别所述特征缺失值与所述预统计特征的关联信息,依据所述关联信息采集对应的缺失出处。

[0011] 进一步地,所述基于建筑终端预设的数据单位类型,将能耗数据的数据格式统一为所述数据单位类型的步骤中,还包括:

检测所述能耗数据中的原始单位,基于所述原始单位识别出待转换的不同单位;

判断所述不同单位能否统一转换为所述数据单位类型;

若否,则标记所述不同单位的不一致项,根据所述不一致项对所述不同单位的对应数据逐行进行单位转换,比对转换前后的数据分布和统计特性。

[0012] 本发明还提供一种基于负荷曲线的节能优化控制系统,包括:

统一模块,用于基于建筑终端预设的数据单位类型,将能耗数据的数据格式统一为所述数据单位类型,其中,所述数据单位类型具体包括温度、电力和时间戳;

判断模块,用于判断所述能耗数据能否实时呈现至所述建筑终端;

执行模块,用于若否,则通过所述建筑终端预设的传感器,从不同的数据源中采集对应的各种多源数据,识别所述多源数据的时间序列,根据所述时间序列应用预设的数据融合,将所述多源数据进行整合,得到融合数据,从所述融合数据中提取对应的特征参数,依据所述特征参数构建有效特征集,其中,所述时间序列具体包括趋势变化和周期性变化,所述特征参数具体包括天气影响指数、经济活动水平和历史负荷数据;

第二判断模块,用于判断所述有效特征集是否需要预设的数据清洗;

第二执行模块,用于若不需要,则采用预设的相关系数,采集所述有效特征集中的各个特征与负荷变量之间的线性关系,在所述有效特征集上建立对应的基线模型,计算所述基线模型的均方误差和决定系数,通过所述基线模型构建所述建筑终端的负荷预测结果,将所述负荷预测结果与实际负荷值对比,生成负荷折线图。

[0013] 进一步地,还包括:

划分模块,用于基于所述传感器预设的工作范围,通过所述建筑终端划分所述传感器在建筑内覆盖的采集区域;

第三判断模块,用于判断所述采集区域能否检测到预设的重叠监测边缘;

第三执行模块,用于若能,则根据所述重叠监测边缘构建重叠时间窗口,标识多个传感器在所述重叠时间窗口的采集测量值,依据所述采集测量值生成待融合的数据特征,应用预设的卡尔曼滤波为所述采集测量值进行权重分配,计算所述数据特征的加权平均值,其中,所述数据特征具体包括温度、湿度和能耗。

[0014] 进一步地,所述执行模块还包括:

计算单元,用于基于预设的虚拟变量,构建所述建筑终端对能耗的季节性特征,计算所述季节性特征之间的比率特征和差值特征,其中,所述虚拟变量具体包括春季、夏季、秋季和冬季,所述比率特征具体为能耗与温度的比率,所述差值特征具体为当前负荷与过去负荷的差值;

判断单元,用于判断所述比率特征和所述差值特征是否存在缺失值;

执行单元,用于若否,则对所述比率特征和所述差值特征进行标准化处理,确保所述比率特征和所述差值特征始终处于同一尺度上,将所述比率特征和所述差值特征与其他特征进行组合,形成复合特征,根据所述复合特征捕捉季节性变化对应的能耗影响。

[0015] 本发明提供了基于负荷曲线的节能优化控制方法及系统,具有以下有益效果:

本发明通过从不同数据源采集多源数据,可以形成更全面的能耗视图,整合这些数据能够提供更丰富的上下文,帮助识别影响能耗的各种因素,提高节能优化的准确性,同时提取特征参数使得模型能够基于多个维度进行分析,这些特征可以揭示不同因素如何影响能耗,帮助更好地理解 and 预测负荷变化,并且在有效特征集上建立基线模型,并计算均方误差和决定系数,能够量化模型的性能,不仅为能耗预测提供了基准,还能通过与实际数据的比较来不断优化模型参数,提升预测准确性。

附图说明

[0016] 图1为本发明基于负荷曲线的节能优化控制方法一个实施例的流程示意图;

图2为本发明基于负荷曲线的节能优化控制系统一个实施例的结构框图。

具体实施方式

[0017] 应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明,本发明为目的的实现、功能特点及优点将结合实施例,参照附图做进一步说明。

[0018] 下面将结合本发明的实施例中的附图,对本发明的实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明的一部分实施例,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作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0019] 参考附图1,为本发明一实施例中的基于负荷曲线的节能优化控制方法,包括:

S1:基于建筑终端预设的数据单位类型,将能耗数据的数据格式统一为所述数据单位类型,其中,所述数据单位类型具体包括温度、电力和时间戳;

S2:判断所述能耗数据能否实时呈现至所述建筑终端;

S3:若否,则通过所述建筑终端预设的传感器,从不同的数据源中采集对应的各种多源数据,识别所述多源数据的时间序列,根据所述时间序列应用预设的数据融合,将所述多源数据进行整合,得到融合数据,从所述融合数据中提取对应的特征参数,依据所述特征参数构建有效特征集,其中,所述时间序列具体包括趋势变化和周期性变化,所述特征参数具体包括天气影响指数、经济活动水平和历史负荷数据;

S4:判断所述有效特征集是否需要预设的数据清洗;

S5:若不需要,则采用预设的相关系数,采集所述有效特征集中的各个特征与负荷变量之间的线性关系,在所述有效特征集上建立对应的基线模型,计算所述基线模型的均方误差和决定系数,通过所述基线模型构建所述建筑终端的负荷预测结果,将所述负荷预测结果与实际负荷值对比,生成负荷折线图。

[0020] 在本实施例中,系统基于建筑终端预先设好的数据单位类型,数据单位类型具体包括温度、电力和时间戳,将建筑内采集得到的能耗数据的数据格式统一为数据单位类型,而后系统判断这些能耗数据能否实时呈现至建筑终端,以执行对应的步骤;例如,当系统判定到采集得到的能耗数据能够实时呈现至建筑终端时,则系统会认为数据在质量上符合要求,能够有效反映实际的能耗情况,系统会在建筑终端上实时显示采集的能耗数据,例如能耗曲线图、饼图或柱状图,以使用户快速获取信息,同时根据历史数据和需求设定报警阈值,当某项能耗超出正常范围时,系统自动发出警报,以便于及时采取措施,并且将实时数据记录到数据库中,以便后续分析和查询,确保数据的完整性和可追溯性,进行动态数据分析,生成趋势图或预测模型,帮助管理者识别能耗模式和潜在的异常情况;例如,当系统判定到采集得到的能耗数据无法实时呈现至建筑终端,此时系统会认为无法反映实际的能耗情况,系统会通过建筑终端预先设好的传感器,从不同的能耗数据源中采集对应的各种多源数据,识别这些多源数据的时间序列,时间序列具体包括趋势变化和周期性变化,根据时间序列应用预先设有的数据融合,将多源数据进行整合,得到融合数据,从融合数据中提取对应的特征参数,特征参数具体包括天气影响指数、经济活动水平和历史负荷数据,依据这些特征参数构建有效特征集;系统通过识别和分析时间序列中的趋势变化和周期性变化,使系统能够捕捉到能耗数据的动态特征,这一过程不仅帮助理解过去的能耗模式,还为未来的能耗预测提供重要依据,通过对时间序列的深入分析,管理者可以识别出季节性波动、经济活动高峰和低谷,从而在特定时段优化能耗策略,同时在融合过程中,系统通过预设的

数据融合算法将不同来源的数据整合为一个统一的数据集,该过程有助于消除数据中的噪声和冗余,提升数据质量,通过数据融合,管理者可以获得更精准的信息,进而更好地支持能耗分析和决策,并且依据提取的特征参数,构建有效特征集为后续的预测模型奠定基础,通过有效特征集,模型能够捕捉到建筑能耗的关键驱动因素,提高预测的准确性;而后系统判断有效特征集是否需要预先设定的数据清洗,以执行对应的步骤;例如,当系统判定到有效特征集需要预先设定的数据清洗时,则系统会认为有效特征集中的数据可能存在错误、异常值或噪声,这些问题会影响后续分析和模型构建的准确性,系统会明确清洗的目标,例如去除重复数据、填补缺失值、处理异常值等,确保所有清洗步骤均与后续分析的需求相符合,同时检查特征集中是否存在缺失值,决定如何处理这些缺失值,并且确保所有特征的单位、格式和类型一致,例如温度单位可能需要统一为摄氏度或华氏度,电力单位需要确保一致性,通过识别并去除重复数据,确保特征集的唯一性,减少冗余对分析的干扰;例如,当系统判定到有效特征集不需要预先设定的数据清洗,此时系统会认为有效特征集中的数据不存在错误、异常值或噪声,系统会采用预先设定的相关系数,采集有效特征集中的各个特征与负荷变量之间的线性关系,在有效特征集上建立对应的基线模型,计算基线模型的均方误差和决定系数,通过基线模型构建建筑终端的负荷预测结果,将负荷预测结果与实际负荷值对比,生成负荷折线图;系统通过直接使用有效特征集可以节省数据清洗所需的时间和资源,加快模型建立的速度,使得系统能够迅速响应变化,及时提供负荷预测结果,同时通过预设的相关系数,系统能够分析有效特征集中的各个特征与负荷变量之间的线性关系,这种分析可以帮助管理者更好地理解影响建筑能耗的关键因素,指导后续的管理决策,并且基于线性关系建立的基线模型通常具有较好的可解释性,管理者可以清晰地看到每个特征对负荷预测的影响程度,从而做出更有针对性的调整,而通过建立基线模型,系统能够计算负荷预测的均方误差和决定系数,这些指标为预测结果的准确性提供了量化的评估,有助于优化能耗管理,将负荷预测结果与实际负荷值对比,可以快速识别预测与实际之间的差距,这一反馈机制使得系统能够及时调整能耗管理策略,提高资源利用效率。

[0021] 需要说明的是,识别所述多源数据的时间序列,根据所述时间序列应用预设的数据融合,将所述多源数据进行整合,得到融合数据,具体示例如下:

假设在一个大型商业建筑中进行能耗预测,数据源包括天气、经济活动和历史能耗数据,将以2024年10月1日为例,详细说明如何通过时间序列进行多源数据融合;

收集数据,假设从不同数据源收集到以下信息:

天气数据(按小时):

08:00 - 温度:10°C,湿度:60%;

09:00 - 温度:12°C,湿度:55%;

10:00 - 温度:14°C,湿度:50%;

经济活动数据(按小时):

08:00 - 客流量:1000人,商业活动指数:0.6;

09:00 - 客流量:1200人,商业活动指数:0.7;

10:00 - 客流量:1500人,商业活动指数:0.8;

历史负荷数据(按小时):

08:00 - 能耗:200kWh;

09:00 - 能耗:220kWh;

10:00 - 能耗:240kWh;

识别时间序列,将每个数据源的时间点对齐,确保所有数据在同一时间范围内,例如,所有数据的时间戳都是2024年10月1日的08:00、09:00和10:00;

应用数据融合,在这一阶段,将所有数据整合到一个统一的数据集中,以下是具体步骤:

对齐时间戳:所有数据已经在同一时间点,无需调整;

创建融合数据表:将各个数据源的特征整合到一个表格中,每行代表一个时间点的

时间戳	温度 (°C)	湿度 (%)	客流量 (人)	商业活动指数	能耗 (kWh)
2024-10-01 08:00	10	60	1000	0.6	200
2024-10-01 09:00	12	55	1200	0.7	220
2024-10-01 10:00	14	50	1500	0.8	240

加权平均:假设认为温度对能耗的影响大于湿度和经济活动指数,可以为不同特征设定权重,例如:

温度:0.5;

湿度:0.2;

客流量:0.2;

商业活动指数:0.1;

可以通过这些权重计算一个加权综合特征,例如:

综合指数=0.5*温度+0.2*湿度+0.2*客流量/1000+0.1*商业活动指数

得到融合数据,系统通过上述步骤,得到了融合数据,而这个融合数据集将为后续的负荷预测分析提供支持;

数据分析与应用,特征提取,根据融合数据,提取相关特征用于建立负荷预测模型,这些特征包括:温度、客流量、商业活动指数;

建模与预测,使用融合数据集训练预测模型,例如线性回归或决策树,以预测未来的能耗,例如使用过去3小时的融合数据预测接下来1小时的能耗;

结果可视化后,生成一个负荷折线图,将实际能耗与模型预测能耗进行对比,以评估模型的效果;

综上所述,系统通过以上步骤,将多源数据有效融合,形成一个完整的分析基础,这样的数据融合方法不仅提高了数据的利用效率,还为能耗预测提供了更高的准确性和可靠性。

[0022] 则采用预设的相关系数,采集所述有效特征集中的各个特征与负荷变量之间的线性关系,具体示例如下:

假设在一个商业建筑中收集了以下能耗相关的数据,以预测其电力消耗,收集的特征包括:

温度(°C);

客流量(人);

商业活动指数(无量纲);

湿度(%)；

能耗(kWh)；

数据集,将以下数据视为输入数据集:

时间戳	温度 (°C)	客流量 (人)	商业活动指数	湿度 (%)	能耗 (kWh)
2024-10-01 08:00	10	1000	0.6	60	200
2024-10-01 09:00	12	1200	0.7	55	220
2024-10-01 10:00	14	1500	0.8	50	240
2024-10-01 11:00	15	1600	0.85	45	260
2024-10-01 12:00	16	1700	0.9	40	280
2024-10-01 13:00	18	1800	0.95	35	300

计算相关系数,假设想要计算这些特征与能耗之间的相关系数,可以使用皮尔逊相关系数进行计算,以下是如何进行计算的详细步骤:

数据准备,首先,需要确保所有数据都是数字格式,以下是用Python进行相关系数计算的伪代码示例:

```
import pandas as pd
# 创建数据框
data = {
    '温度': [10, 12, 14, 15, 16, 18],
    '客流量': [1000, 1200, 1500, 1600, 1700, 1800],
    '商业活动指数': [0.6, 0.7, 0.8, 0.85, 0.9, 0.95],
    '湿度': [60, 55, 50, 45, 40, 35],
    '能耗': [200, 220, 240, 260, 280, 300]
}
df = pd.DataFrame(data)
# 计算相关系数
correlation_matrix = df.corr()
print(correlation_matrix)
```

计算结果,假设计算出的相关系数矩阵如下:

	温度	客流量	商业活动指数	湿度	能耗
温度	1.00	0.98	0.95	-0.97	0.97
客流量	0.98	1.00	0.96	-0.95	0.98
商业活动指数	0.95	0.96	1.00	-0.94	0.96
湿度	-0.97	-0.95	-0.94	1.00	-0.96
能耗	0.97	0.98	0.96	-0.96	1.00

分析相关系数,温度与能耗的相关系数为0.97,说明温度与能耗之间存在非常强的正相关关系,气温上升时,能耗也显著增加;

客流量与能耗的相关系数为0.98,表示客流量的增加与能耗的增加紧密相关;

湿度与能耗的相关系数为-0.96,说明湿度与能耗之间存在强的负相关关系,湿度降低时,能耗增加;

特征选择,根据计算出的相关系数,可以决定哪些特征应该保留在模型中,可以选择与能耗相关性较高的特征,如温度、客流量和商业活动指数,而湿度则可能在模型中不那么重要;

建立基线模型,使用与能耗关系较强的特征,建立线性回归模型,可以使用如下步骤:

模型训练:将温度、客流量和商业活动指数作为输入特征,能耗作为目标变量,训练线性回归模型;

```
from sklearn.linear_model import LinearRegression
```

```
# 划分输入和输出
```

```
X = df[['温度', '客流量', '商业活动指数']]
```

```
y = df['能耗']
```

```
# 创建和训练模型
```

```
model = LinearRegression()
```

```
model.fit(X, y)
```

模型评估,计算模型的均方误差(MSE)和决定系数(R^2),评估模型的性能;

结果可视化,绘制散点图将实际能耗与模型预测值进行比较,生成散点图和负荷折线图,帮助可视化模型的准确性;

综上所述,详细展示了如何计算有效特征集中的各个特征与能耗变量之间的相关系数,并如何基于这些结果进行特征选择和基线模型建立,这种方法能够有效提升能耗预测模型的准确性和可靠性。

[0023] 在所述有效特征集上建立对应的基线模型,计算所述基线模型的均方误差和决定系数,通过所述基线模型构建所述建筑终端的负荷预测结果,具体示例如下:

假设有一个商业建筑的能耗数据集,收集的特征包括温度、客流量和商业活动指数,想用这些特征预测建筑的能耗(以千瓦时kWh为单位),数据集如下:

时间戳	温度 (°C)	客流量 (人)	商业活动指数	能耗 (kWh)
2024-10-01 08:00	10	1000	0.6	200
2024-10-01 09:00	12	1200	0.7	220
2024-10-01 10:00	14	1500	0.8	240
2024-10-01 11:00	15	1600	0.85	260
2024-10-01 12:00	16	1700	0.9	280
2024-10-01 13:00	18	1800	0.95	300

建立基线模型,将使用线性回归模型来预测能耗;

首先,导入必要的库并准备数据:

```
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
import matplotlib.pyplot as plt
```

```
# 创建数据框
```

```
data = {
```

```
    '时间戳': ['2024-10-01 08:00', '2024-10-01 09:00', '2024-10-01
```

```
10:00',  
                '2024-10-01 11:00', '2024-10-01 12:00', '2024-10-01 13:  
00'],
```

```
    '温度': [10, 12, 14, 15, 16, 18],  
    '客流量': [1000, 1200, 1500, 1600, 1700, 1800],  
    '商业活动指数': [0.6, 0.7, 0.8, 0.85, 0.9, 0.95],  
    '能耗': [200, 220, 240, 260, 280, 300]
```

```
}
```

```
df = pd.DataFrame(data)
```

```
df['时间戳'] = pd.to_datetime(df['时间戳']) # 转换为日期时间格式  
而后划分特征和目标变量,将特征(X)和目标变量(y)分开;
```

```
X = df[['温度', '客流量', '商业活动指数']]
```

```
y = df['能耗']
```

然后创建和训练模型,使用线性回归模型进行训练:

```
model = LinearRegression()
```

```
model.fit(X, y)
```

预测能耗,可以使用训练好的模型进行能耗预测:

```
predictions = model.predict(X)
```

计算均方误差和决定系数,使用预测结果计算均方误差 (MSE) 和决定系数 (R^2):

```
mse = mean_squared_error(y, predictions)
```

```
r2 = r2_score(y, predictions)
```

```
print(f"均方误差 (MSE): {mse}") # 输出均方误差  
print(f"决定系数 ( $R^2$ ):  
{r2}") # 输出决定系数
```

示例结果,假设计算结果如下:

均方误差 (MSE): 25.0;

决定系数 (R^2): 0.98;

均方误差 (MSE) 为 25.0, 表示预测值与实际值之间的平均平方差, 数值越小, 模型的预测精度越高;

决定系数 (R^2) 为 0.98, 表明模型可以解释 98% 的能耗变化, 表明模型拟合效果非常好;

最后生成负荷预测结果, 将预测结果与实际能耗值对比, 并生成负荷预测的折线图:

```
plt.figure(figsize=(10, 5))
```

```
plt.plot(df['时间戳'], y, label='实际能耗', marker='o')
```

```
plt.plot(df['时间戳'], predictions, label='预测能耗', marker='x')
```

```
plt.xlabel('时间')
```

```
plt.ylabel('能耗 (kWh)')
```

```
plt.title('能耗预测')
```

```
plt.legend()
```

```
plt.xticks(rotation=45)
plt.grid()
plt.show()
```

综上所述,详细展示了如何在有效特征集上建立基线模型、计算均方误差和决定系数,如何生成负荷预测结果的可视化,这个过程帮助理解模型性能,并为未来的能耗预测提供了数据基础。

[0024] 在本实施例中,则通过所述建筑终端预设的传感器,从不同的数据源中采集对应的各种多源数据的步骤S3前,还包括:

S301:基于所述传感器预设的工作范围,通过所述建筑终端划分所述传感器在建筑内覆盖的采集区域;

S302:判断所述采集区域能否检测到预设的重叠监测边缘;

S303:若能,则根据所述重叠监测边缘构建重叠时间窗口,标识多个传感器在所述重叠时间窗口的采集测量值,依据所述采集测量值生成待融合的数据特征,应用预设的卡尔曼滤波为所述采集测量值进行权重分配,计算所述数据特征的加权平均值,其中,所述数据特征具体包括温度、湿度和能耗。

[0025] 在本实施例中,系统基于传感器在建筑内预先设好的工作范围,通过建筑终端划分传感器在建筑内覆盖的采集区域,而后系统判断这些采集区域能否检测到预先设定的重叠监测边缘,以执行对应的步骤;例如,当系统判定到传感器在建筑内覆盖的采集区域并未检测到预先设定的重叠监测边缘时,则系统会认为传感器的工作范围划分得当,每个传感器覆盖的采集区域均相对独立,避免了重复监测,系统会继续采集各个区域的能耗数据,按预定的频率将数据传输到建筑终端,同时通过监控数据传输的稳定性,确保各个传感器在独立采集区域时的数据能够按时、完整地传输到终端,并且尽管采集区域无重叠,系统也可以定期检查传感器的状态,包括电池寿命、信号强度等,以确保长期稳定运行,而由于传感器覆盖区域未重叠,系统需定期核查是否有区域数据遗漏,如果发现某些区域的数据采集有遗漏或延迟,需要采取修复措施;例如,当系统判定到传感器在建筑内覆盖的采集区域检测到预先设定的重叠监测边缘,此时系统会认为传感器的工作范围存在多次覆盖检测的区域,系统会根据重叠监测边缘构建重叠时间窗口,标识多个传感器在重叠时间窗口的采集测量值,依据不同的采集测量值生成待融合的数据特征,数据特征具体包括温度、湿度和能耗,应用预先设好的卡尔曼滤波为采集测量值进行权重分配,计算数据特征的加权平均值;系统通过在重叠区域内结合来自多个传感器的测量值,能够减少单个传感器可能带来的误差,这种冗余数据可以帮助确认和校正异常值,从而提高整体数据的准确性,同时重叠监测允许系统利用多个来源的数据特征,从而提升数据的可靠性,通过卡尔曼滤波,系统能够在存在噪声和不确定性的情况下,通过加权平均来获得更稳定和可靠的数据特征,并且依据重叠时间窗口内的多个测量值生成待融合的数据特征,确保了数据融合的多样性和全面性,不同传感器在同一时刻的测量结果可以提供更丰富的信息,使得最终融合的数据特征更具代表性,而卡尔曼滤波根据实时测量的变化自动调整不同传感器的权重,使得系统能够根据实际情况动态优化数据处理,这种自适应能力在变化环境中尤为重要,可以实时响应传感器状态的变化。

[0026] 需要说明的是,根据所述重叠监测边缘构建重叠时间窗口,标识多个传感器在所

述重叠时间窗口的采集测量值,依据所述采集测量值生成待融合的数据特征,应用预设的卡尔曼滤波为所述采集测量值进行权重分配,计算所述数据特征的加权平均值,具体示例如下:

假设在一栋办公楼内,有两个温度传感器,传感器A和传感器B,它们的重叠时间窗口为10:00到10:05,在这个时间段内,两个传感器分别记录了温度数据,以下是具体的数据记录:

传感器A的记录:

10:00 - 22°C;

10:01 - 23°C;

10:02 - 24°C;

10:03 - 22°C;

10:04 - 23°C;

10:05 - 24°C;

传感器B的记录:

10:00 - 21°C;

10:01 - 22°C;

10:02 - 23°C;

10:03 - 22°C;

10:04 - 21°C;

10:05 - 23°C;

构建重叠时间窗口,识别到传感器A和B在10:00到10:05的重叠时间段,这个时间窗口将用于整合这两个传感器的数据;

标识采集测量值,在重叠时间窗口内,系统将提取两个传感器的温度测量值,得到以下表格:

时间	传感器A (°C)	传感器B (°C)
10:00	22	21
10:01	23	22
10:02	24	23
10:03	22	22
10:04	23	21
10:05	24	23

生成待融合的数据特征,从表格中提取出待融合的温度特征,例如,对于10:00的时间段,待融合特征为 [22, 21];

应用卡尔曼滤波并设定滤波参数,假设传感器A的测量噪声协方差为0.1,传感器B的测量噪声协方差为0.3,根据噪声的不同,系统会动态调整权重。

[0027] 权重计算:

$$\text{传感器A的权重 } \omega_A = \frac{1}{\text{噪声协方差}} = \frac{1}{0.1} = 10$$

[0028] 传感器B的权重 $\omega_B = \frac{1}{\text{噪声协方差}} = \frac{1}{3} \approx 3.33$

[0029] 总权重 $= \omega_A + \omega_B = 10 + 3.33 = 13.33$

[0030] 归一化权重 $(\omega'_A) = \frac{10}{13.33} \approx 0.75$, $(\omega'_B) = \frac{3.33}{13.33} \approx 0.25$

[0031] 以10:00为例,则计算数据特征的加权平均值:

$$\begin{aligned} &= (\omega'_A \times T_A + \omega'_B \times T_B) \\ &= (0.75 \times 22) + (0.25 \times 21) \\ &= 16.5 + 5.25 \\ &= 21.75^\circ\text{C} \end{aligned}$$

[0032] 对于其他时间点重复上述步骤,得到:

时间	加权平均温度(°C)
10:00	21.75
10:01	$(0.75 \times 23) + (0.25 \times 22) = 22.75$
10:02	$(0.75 \times 24) + (0.25 \times 23) = 23.75$
10:03	$(0.75 \times 22) + (0.25 \times 22) = 22.00$
10:04	$(0.75 \times 23) + (0.25 \times 21) = 22.75$
10:05	$(0.75 \times 24) + (0.25 \times 23) = 23.75$

综上所述,最终得到的加权平均温度数据将更为准确,能够更好地反映实际的环境情况,这种基于重叠监测边缘和卡尔曼滤波的数据融合方法,为智能建筑管理提供了更为精确和可靠的环境监测数据,确保了决策的有效性和科学性。

[0033] 在本实施例中,依据所述特征参数构建有效特征集的步骤S3中,还包括:

S31:基于预设的虚拟变量,构建所述建筑终端对能耗的季节性特征,计算所述季节性特征之间的比率特征和差值特征,其中,所述虚拟变量具体包括春季、夏季、秋季和冬季,所述比率特征具体为能耗与温度的比率,所述差值特征具体为当前负荷与过去负荷的差值;

S32:判断所述比率特征和所述差值特征是否存在缺失值;

S33:若否,则对所述比率特征和所述差值特征进行标准化处理,确保所述比率特征和所述差值特征始终处于同一尺度上,将所述比率特征和所述差值特征与其他特征进行组合,形成复合特征,根据所述复合特征捕捉季节性变化对应的能耗影响。

[0034] 在本实施例中,系统基于预先设好的虚拟变量,虚拟变量具体包括春季、夏季、秋季和冬季,构建建筑终端对能耗的季节性特征,计算季节性特征之间的比率特征和差值特征,比率特征具体为能耗与温度的比率,差值特征具体为当前负荷与过去负荷的差值,而后系统判断这些比率特征和差值特征是否存在缺失值,以执行对应的步骤;例如,当系统判定到季节性特征之间的比率特征和差值特征存在缺失值时,则系统会认为这些特征可能无法准确反映能耗的季节性变化,这可能会影响后续的分析 and 预测,系统会记录缺失值的具体位置(时间点和特征),计算缺失值的比例,以评估数据完整性,同时对数据进行重采样,选择时间段内的代表性数据进行计算,确保数据的时效性和连续性,并且对于无法合理填补

的缺失值,可以选择标记为“缺失”或“异常”,在后续分析中予以剔除或单独处理,在处理完缺失值后,重新计算比率特征和差值特征,评估其对后续模型的影响;例如,当系统判定到季节性特征之间的比率特征和差值特征不存在缺失值,此时系统会认为这些特征能够准确反映能耗的季节性变化,系统会对这些比率特征和差值特征进行标准化处理,确保比率特征和差值特征始终处于同一尺度上,将这些比率特征和差值特征进行组合,形成复合特征,根据复合特征捕捉季节性变化对应的能耗影响;系统通过对比率特征和差值特征进行标准化处理,可以确保所有特征都在同一尺度上进行比较,这种标准化可以防止某些特征因为数值范围较大而主导其他特征,使模型更加平衡,提升预测精度,例如能耗与温度的比率值可能相对较小,而能耗差值可能较大,标准化处理后可以让它们在相同的范围内进行计算,避免特征之间的不平衡,同时复合特征通过组合不同的比率特征和差值特征,能够捕捉到更加复杂的能耗模式。这种组合特征在模型中可以揭示出季节性变化与能耗波动之间更深层次的关联,使得模型不仅仅依赖单一特征,更能够从多个维度全面分析能耗变化的驱动因素,这将有助于提升模型的预测准确性,并且比率特征和差值特征的组合能够精确地捕捉不同季节下能耗的变化趋势,例如冬季的负荷变化与春夏季可能有不同的能耗驱动因素,通过分析这些复合特征,系统能够识别出冬季或夏季中特定的能耗模式,从而优化负荷预测,而复合特征可以为预测模型提供更丰富的信息,使得预测结果更具解释性,决策者可以基于复合特征,明确不同季节对能耗的具体影响因素,从而在实际操作中更好地理解能耗的波动来源,并采取相应的措施。

[0035] 需要说明的是,对所述比率特征和所述差值特征进行标准化处理,确保所述比率特征和所述差值特征始终处于同一尺度上,将所述比率特征和所述差值特征与其他特征进行组合,形成复合特征,根据所述复合特征捕捉季节性变化对应的能耗影响,具体示例如下:

假设系统管理一个大型办公建筑,采集到以下数据:

温度(T):每天的外界温度(单位:°C);

能耗(E):建筑的能耗(单位:kWh),每小时记录;

历史能耗(E_history):过去24小时的能耗记录(单位:kWh);

季节(S):四个虚拟变量,表示当前的季节:春(1)、夏(2)、秋(3)、冬(4);

示例数据,假设在一个夏季的下午,系统采集到以下数据:

当前温度(T):30°C;

当前能耗(E):300 kWh;

过去24小时能耗(E_history):250 kWh;

季节(S):夏季(用2表示);

首先计算比率特征和差值特征,

比率特征(R):

$$\text{计算能耗与温度的比率: } R = \frac{E}{T} = \frac{300}{30} = 10$$

[0036] 差值特征(D):

$$\text{计算当前能耗与过去能耗的差值: } D = E - E_{\text{history}} = 300 - 250 = 50$$

[0037] 假设在整个夏季的能耗数据中,以下是比率特征和差值特征的统计信息:

比率特征的均值 (μ_r) :8

比率特征的标准差 (σ_r) :2

差值特征的均值 (μ_d) :30

差值特征的标准差 (σ_d) :10

则比率特征的标准化: $R_{\text{standardized}} = \frac{R - \mu_r}{\sigma_r} = \frac{10 - 8}{2} = 1$

[0038] 差值特征的标准化: $D_{\text{standardized}} = \frac{D - \mu_d}{\sigma_d} = \frac{50 - 30}{10} = 2$

[0039] 将标准化后的比率特征和差值特征与季节特征组合成复合特征;
假设季节特征 S(夏季)被表示为 2,则可以形成以下复合特征:

$$C_1 = R_{\text{standardized}} \times S = 1 \times 2 = 2$$

[0040] $C_2 = D_{\text{standardized}} \times S = 2 \times 2 = 4$

[0041] 捕捉季节性变化对能耗的影响,现在系统有了复合特征 C1 和 C2,可以用它们来分析能耗的季节性影响;

复合特征 C1 = 2:表明在夏季,能耗对温度的敏感度相对较高;

复合特征 C2 = 4:表明与历史能耗相比,当前能耗增加了显著的50 kWh,且这种变化在夏季是常见的;

综上所述,通过这些复合特征,系统可以使用机器学习模型预测未来几个小时的能耗,考虑到季节性和温度变化的影响,同时如果系统预测到未来几个小时内温度会继续升高,复合特征显示出能耗将显著增加,管理人员可以提前启动节能措施,比如调节空调的温度设定或启动备用能源,并且系统可以生成关于能耗和气候影响的详细报告,帮助建筑管理者做出更科学的决策。

[0042] 在本实施例中,在所述有效特征集上建立对应的基线模型的步骤S5中,还包括:

S51:基于所述建筑终端对能耗的预设图表,从所述图表上绘制X轴为能耗预测值,并从所述图表上绘制Y轴为能耗残差,结合得到能耗散点图;

S52:判断所述能耗散点图上的残差是否随机分布;

S53:若是,则将所述能耗散点图输入至所述基线模型,通过所述基线模型进行交叉验证,获取所述基线模型在不同数据子集上的波动信息。

[0043] 在本实施例中,系统基于建筑终端对能耗预先设置的图表,从图表上绘制X轴为能耗预测值,并在图表上绘制Y轴为能耗残差,结合得到能耗散点图,而后系统判断该能耗散点图上的残差是否随机分布,以执行对应的步骤;例如,当系统判定到能耗散点图上的残差并非随机分布时,则系统会认为基线模型可能存在系统性偏差,模型并未准确捕捉到能耗与相关特征之间的真实关系,系统会检查模型使用的特征是否涵盖了所有可能影响能耗的关键因素,尤其是对天气、经济活动、建筑设备运行状态等外部因素的反映是否充分,同时通过引入更多虚拟变量、调整特征的构造方式或考虑交互特征来增强模型对能耗模式的识别能力,并且再次生成能耗散点图,检查残差分布是否呈现出随机性,确保模型能够较好地反映实际能耗波动;例如,当系统判定到能耗散点图上的残差属于随机分布,此时系统会认为基线模型能够准确捕捉到能耗与相关特征之间的真实关系,系统会将能耗散点图输入至基线模型,通过基线模型进行交叉验证,获取基线模型在不同数据子集上的波动信息;系统

通过在不同数据子集上进行交叉验证,可以评估模型的稳定性和一致性,确保模型在不同场景下都能保持良好的预测性能,同时在不同数据子集上进行交叉验证,可以获取多组模型性能指标(如均方误差、决定系数等),从而全面了解模型的预测能力,并且交叉验证过程中可以调整模型的超参数,以寻找最佳的模型配置,提高模型的预测准确性和鲁棒性,而交叉验证过程中,可以分析各特征对模型预测结果的影响,帮助识别出哪些特征在不同条件下仍然具有重要作用,从而为后续的特征选择和模型改进提供依据。

[0044] 需要说明的是,将所述能耗散点图输入至所述基线模型,通过所述基线模型进行交叉验证,获取所述基线模型在不同数据子集上的波动信息,具体示例如下:

假设有一座办公大楼的能耗数据集,数据包括以下特征:

温度($^{\circ}\text{C}$);

湿度(%);

经济活动水平(指数);

历史能耗(kWh);

当前能耗(kWh)(目标变量);

数据准备,

收集数据:收集一年内的小时能耗数据,假设总共有8,760条记录;

数据清洗:去除缺失值和异常值,确保数据质量;

划分数据集:将数据集随机分为10个子集($K=10$),即每个子集包含876条记录;

交叉验证过程,循环训练和测试:

第1轮:使用子集1作为测试集,子集2到子集10作为训练集;

训练模型,记录模型参数,用子集1的输入特征预测能耗,并计算均方误差(MSE);

第2轮:使用子集2作为测试集,子集1和子集3到10作为训练集;

继续此过程,直到所有子集都被用作测试集一次;

计算性能指标,假设经过10轮交叉验证后,记录的均方误差如下:

子集1:MSE=5.2;

子集2:MSE=4.8;

子集3:MSE=5.0;

子集4:MSE=6.1;

子集5:MSE=5.4;

子集6:MSE=4.9;

子集7:MSE=5.3;

子集8:MSE=5.1;

子集9:MSE=4.7;

子集10:MSE=5.5;

汇总:计算平均MSE和标准差:

平均MSE=(5.2+4.8+5.0+6.1+5.4+4.9+5.3+5.1+4.7+5.5)/10=5.2;

标准差=0.26(假设计算得出);

可视化和分析,

绘制图表:生成图表,横轴为实际能耗,纵轴为预测能耗,添加趋势线,直观显示模

型预测能力;

生成残差图:绘制残差图,查看残差是否随机分布,是否存在模式;

最终模型优化分析结果,发现某些特定时间段(如夏季高温期间)模型的MSE较高,可能是因为温度和能耗之间的关系较为复杂;

模型调整:考虑添加更多特征(如太阳辐射强度、风速等);

综上所述,通过这个详细的交叉验证过程,能够全面评估模型在不同数据集上的表现,确保模型具有良好的泛化能力,若发现模型在某些特定情况下表现不佳,可以根据交叉验证的结果进行针对性调整,从而提升建筑能耗预测的准确性,这样的细致分析和优化过程,有助于实现更有效的能耗管理,提高能效,降低运营成本。

[0045] 在本实施例中,判断所述能耗数据能否实时呈现至所述建筑终端的步骤S2中,还包括:

S21:基于所述建筑终端预设的网络带宽,检测所述能耗数据进行数据传输时的网络延迟信息;

S22:判断所述网络延迟信息是否超出预设的延迟时段;

S23:若是,则在所述能耗数据的数据采集环节和数据展示环节中添加时间戳标签,根据所述时间戳标签计算所述能耗数据的整体延迟。

[0046] 在本实施例中,系统基于建筑终端预先设定的网络带宽,检测能耗数据进行数据传输时的网络延迟信息,而后系统判断这些网络延迟信息是否超出预先设定的延迟时段,以执行对应的步骤;例如,当系统判定到能耗数据进行数据传输时的网络延迟信息并未超出预先设定的延迟时段时,则系统会认为网络连接正常,数据传输顺畅,能够及时反映建筑内的实时能耗情况,系统会保持当前的数据传输状态,确保能耗数据持续上传至建筑终端,同时记录当前的延迟信息,并更新系统监控面板,确保实时监控数据传输状态,并且进行数据完整性检查,确保在传输过程中未出现数据丢失或损坏,生成实时反馈报告,向管理人员展示当前能耗数据的状态及网络传输情况,确保所有相关人员了解系统运行情况;例如,当系统判定到能耗数据进行数据传输时的网络延迟信息超出了预先设定的延迟时段,此时系统会认为数据传输异常,无法及时反映建筑内的实时能耗情况,系统会在能耗数据的数据采集环节和数据展示环节中添加时间戳标签,根据不同的时间戳标签自适应计算能耗数据的整体延迟;系统通过时间戳标签,可以精确记录每个数据点的采集和传输时间,帮助系统更准确地反映能耗情况,减少由于延迟导致的误判,同时通过对不同时间戳标签的分析,系统能够识别延迟的模式和趋势,从而理解在特定条件下延迟增加的原因,便于后续优化网络配置,并且在发生网络延迟时,系统可以快速定位到具体的延迟环节(如数据采集或数据传输),便于技术人员针对性地进行故障排除和修复,而整体延迟的自适应计算可以帮助系统调整数据处理策略,比如在网络条件不佳时,降低数据传输频率或选择更高效的数据压缩方式。

[0047] 在本实施例中,判断所述有效特征集是否需要预设的数据清洗的步骤S4中,还包括:

S41:基于所述有效特征集的预统计特征,计算特征缺失值的比例,根据所述比例获取所述特征缺失值的分布,生成对应的缺失值热图;

S42:判断所述缺失值热图是否匹配预设的随机缺失;

S43:若否,则识别所述特征缺失值与所述预统计特征的关联信息,依据所述关联信息采集对应的缺失出处。

[0048] 在本实施例中,系统基于有效特征集的预先统计特征,计算特征缺失值的比例,根据不同比例获取特征缺失值的分布,生成对应的缺失值热图,而后系统判断该缺失值热图是否匹配预先设定的随机缺失,以执行对应的步骤;例如,当系统判定到缺失值热图能够匹配预先设定的随机缺失时,则系统会认为这些特征的缺失值分布没有受到系统性因素的影响,缺失值是随机产生的,而不是由于某些特定原因导致的,系统会记录缺失值热图的结果及其匹配情况,生成分析报告,以便团队了解缺失数据的状态及其对分析的影响,同时由于缺失值是随机的,系统可以继续评估和验证模型的稳定性和准确性,而不需要对数据进行重大的调整,并且虽然当前缺失值是随机的,系统仍需监控后续数据的缺失情况,以确保数据质量持续符合预期标准,而通过分析缺失值的模式,系统可以改进数据收集策略,以减少未来的缺失值出现,从而提升数据的整体质量;例如,当系统判定到缺失值热图无法匹配预先设定的随机缺失,此时系统会认为这些特征的缺失值分布受到影响,系统会识别特征缺失值与预先统计特征的关联信息,依据不同的关联信息采集对应的缺失出处;系统通过分析缺失值与其他特征的关联,能够识别出导致缺失的潜在原因,如传感器故障、数据采集错误或外部环境因素,有助于针对性地解决问题,同时识别非随机缺失值可以帮助模型设计者理解数据集的局限性,在构建模型时考虑这些因素,增强模型对缺失数据的鲁棒性,并且识别特征缺失的来源可以为建筑管理者提供重要的信息,帮助他们在决策时考虑这些数据缺失对能耗分析和预测的潜在影响。

[0049] 需要说明的是,基于所述有效特征集的预统计特征,计算特征缺失值的比例,根据所述比例获取所述特征缺失值的分布,生成对应的缺失值热图,具体示例如下:

假设有一个建筑能耗监测的数据集,包括以下特征:

温度;

湿度;

能耗;

经济活动水平;

数据集的样本记录如下:

温度 (°C)	湿度 (%)	能耗 (kWh)	经济活动水平
22	60	300	0.8
23	NaN	320	0.9
NaN	65	290	NaN
24	70	NaN	0.7
25	NaN	310	0.85
NaN	68	330	0.9

计算缺失值比例,对于每个特征,计算缺失值的比例。

[0050] 温度:缺失值数量 = 3, 总记录数 = 6, 缺失值比例 = $3/6 = 0.5$ (50%);

湿度:缺失值数量 = 3, 缺失值比例 = $3/6 = 0.5$ (50%);

能耗:缺失值数量 = 2, 缺失值比例 = $2/6 = 0.33$ (33.33%);

经济活动水平:缺失值数量 = 1, 缺失值比例 = $1/6 = 0.17$ (16.67%);

获取缺失值分布,将计算结果整理成表格:

特征	缺失值比例
温度	0.50
湿度	0.50
能耗	0.33
经济活动水平	0.17

生成缺失值热图,生成热图的示例代码如下:

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd

# 创建示例数据
data = {
    '特征': ['温度', '湿度', '能耗', '经济活动水平'],
    '缺失值比例': [0.50, 0.50, 0.33, 0.17]
}
df = pd.DataFrame(data)
# 创建热图
plt.figure(figsize=(8, 4))
heatmap_data = df.set_index('特征')
sns.heatmap(heatmap_data, annot=True, cmap='YlGnBu', cbar=True)
plt.title('缺失值热图')
plt.xlabel('缺失值比例')
plt.ylabel('特征')
plt.show()
```

从热图中可以看到“温度”和“湿度”的缺失值比例都达到了50%,这表明这两个特征的数据质量较差,可能需要进一步的清洗和填补;基于热图的分析,团队可以决定使用插值法或其他填补方法来处理缺失值,或重新审视数据采集过程以减少未来的缺失;

综上所述,通过生成热图,直观展示数据的缺失情况,便于快速识别数据问题,同时根据热图结果,制定更合理的数据清洗和处理策略,提升数据的完整性和准确性,并且定期生成热图,有助于监控数据缺失情况的变化,确保数据收集的持续有效性。

[0051] 在本实施例中,基于建筑终端预设的数据单位类型,将能耗数据的数据格式统一为所述数据单位类型的步骤S1中,还包括:

S11:检测所述能耗数据中的原始单位,基于所述原始单位识别出待转换的不同单位;

S12:判断所述不同单位能否统一转换为所述数据单位类型;

S13:若否,则标记所述不同单位的不一致项,根据所述不一致项对所述不同单位的对应数据逐行进行单位转换,比对转换前后的数据分布和统计特性。

[0052] 在本实施例中,系统通过检测能耗数据中的原始单位,基于各个原始单位识别出待转换的不同单位,而后系统判断这些不同单位能否统一转换为数据单位类型,以执行对应的步骤;例如,当系统判定到待转换的不同单位能够统一转换为数据单位类型时,则系统

会认为这些单位之间具有相同的量纲,可以通过标准的转换公式或比例进行转换,系统会明确各个原始单位与数据单位类型之间的转换关系,同时根据定义好的转换规则,逐项转换待转换的数据,并且将转换后的数据更新到数据集,并记录原始单位和转换后的单位,以便追溯和审计,而一旦数据统一为标准的单位类型,可以进行进一步的数据分析、建模和可视化工作;例如,当系统判定到待转换的不同单位无法统一转换为数据单位类型,此时系统会认为这些单位之间不具备相同量纲,无法统一进行转换,系统会标记不同单位的不一致项,根据这些不一致项对不同单位的对应数据逐行进行单位转换,比对转换前后的数据分布和统计特性;系统通过逐行进行单位转换能够避免全局处理时可能引入的错误或偏差,尤其是在复杂的情况下,如单位之间的量纲差异较大或转换规则不明确,系统根据每一条数据的特性单独处理,确保数据转换的精确性,同时在遇到无法转换的单位时,系统不会简单地丢弃数据,而是通过比对转换前后的数据分布和统计特性,确保保留有意义的信息,这种方式可以在复杂数据环境中保持较高的数据完整性,并且标记不同单位的不一致项,确保在后续分析过程中可以追溯这些不一致项,防止混淆或误用错误单位的数据,这样可以在问题出现时快速定位并修正。

[0053] 参考附图2,为本发明一实施例中基于负荷曲线的节能优化控制系统,包括:

统一模块10,用于基于建筑终端预设的数据单位类型,将能耗数据的数据格式统一为所述数据单位类型,其中,所述数据单位类型具体包括温度、电力和时间戳;

判断模块20,用于判断所述能耗数据能否实时呈现至所述建筑终端;

执行模块30,用于若否,则通过所述建筑终端预设的传感器,从不同的数据源中采集对应的各种多源数据,识别所述多源数据的时间序列,根据所述时间序列应用预设的数据融合,将所述多源数据进行整合,得到融合数据,从所述融合数据中提取对应的特征参数,依据所述特征参数构建有效特征集,其中,所述时间序列具体包括趋势变化和周期性变化,所述特征参数具体包括天气影响指数、经济活动水平和历史负荷数据;

第二判断模块40,用于判断所述有效特征集是否需要预设的数据清洗;

第二执行模块50,用于若不需要,则采用预设的相关系数,采集所述有效特征集中的各个特征与负荷变量之间的线性关系,在所述有效特征集上建立对应的基线模型,计算所述基线模型的均方误差和决定系数,通过所述基线模型构建所述建筑终端的负荷预测结果,将所述负荷预测结果与实际负荷值对比,生成负荷折线图。

[0054] 在本实施例中,统一模块10基于建筑终端预先设好的数据单位类型,数据单位类型具体包括温度、电力和时间戳,将建筑内采集得到的能耗数据的数据格式统一为数据单位类型,而后判断模块20判断这些能耗数据能否实时呈现至建筑终端,以执行对应的步骤;例如,当系统判定到采集得到的能耗数据能够实时呈现至建筑终端时,则系统会认为数据在质量上符合要求,能够有效反映实际的能耗情况,系统会在建筑终端上实时显示采集的能耗数据,例如能耗曲线图、饼图或柱状图,以使用户快速获取信息,同时根据历史数据和需求设定报警阈值,当某项能耗超出正常范围时,系统自动发出警报,以便于及时采取措施,并且将实时数据记录到数据库中,以便后续分析和查询,确保数据的完整性和可追溯性,进行动态数据分析,生成趋势图或预测模型,帮助管理者识别能耗模式和潜在的异常情况;例如,当系统判定到采集得到的能耗数据无法实时呈现至建筑终端,此时执行模块30会认为无法反映实际的能耗情况,系统会通过建筑终端预先设好的传感器,从不同的能耗数

据源中采集对应的各种多源数据,识别这些多源数据的时间序列,时间序列具体包括趋势变化和周期性变化,根据时间序列应用预先设定的数据融合,将多源数据进行整合,得到融合数据,从融合数据中提取对应的特征参数,特征参数具体包括天气影响指数、经济活动水平和历史负荷数据,依据这些特征参数构建有效特征集;系统通过识别和分析时间序列中的趋势变化和周期性变化,使系统能够捕捉到能耗数据的动态特征,这一过程不仅帮助理解过去的能耗模式,还为未来的能耗预测提供重要依据,通过对时间序列的深入分析,管理者可以识别出季节性波动、经济活动高峰和低谷,从而在特定时段优化能耗策略,同时在融合过程中,系统通过预设的数据融合算法将不同来源的数据整合为一个统一的数据集,该过程有助于消除数据中的噪声和冗余,提升数据质量,通过数据融合,管理者可以获得更精准的信息,进而更好地支持能耗分析和决策,并且依据提取的特征参数,构建有效特征集为后续的预测模型奠定基础,通过有效特征集,模型能够捕捉到建筑能耗的关键驱动因素,提高预测的准确性;而后第二判断模块40判断有效特征集是否需要预先设定的数据清洗,以执行对应的步骤;例如,当系统判定到有效特征集需要预先设定的数据清洗时,则系统会认为有效特征集中的数据可能存在错误、异常值或噪声,这些问题会影响后续分析和模型构建的准确性,系统会明确清洗的目标,例如去除重复数据、填补缺失值、处理异常值等,确保所有清洗步骤均与后续分析的需求相符合,同时检查特征集中是否存在缺失值,决定如何处理这些缺失值,并且确保所有特征的单位、格式和类型一致,例如温度单位可能需要统一为摄氏度或华氏度,电力单位需要确保一致性,通过识别并去除重复数据,确保特征集的唯一性,减少冗余对分析的干扰;例如,当系统判定到有效特征集不需要预先设定的数据清洗,此时第二执行模块50会认为有效特征集中的数据不存在错误、异常值或噪声,系统会采用预先设定的相关系数,采集有效特征集中的各个特征与负荷变量之间的线性关系,在有效特征集上建立对应的基线模型,计算基线模型的均方误差和决定系数,通过基线模型构建建筑终端的负荷预测结果,将负荷预测结果与实际负荷值对比,生成负荷折线图;系统通过直接使用有效特征集可以节省数据清洗所需的时间和资源,加快模型建立的速度,使得系统能够迅速响应变化,及时提供负荷预测结果,同时通过预设的相关系数,系统能够分析有效特征集中的各个特征与负荷变量之间的线性关系,这种分析可以帮助管理者更好地理解影响建筑能耗的关键因素,指导后续的管理决策,并且基于线性关系建立的基线模型通常具有较好的可解释性,管理者可以清晰地看到每个特征对负荷预测的影响程度,从而做出更有针对性的调整,而通过建立基线模型,系统能够计算负荷预测的均方误差和决定系数,这些指标为预测结果的准确性提供了量化的评估,有助于优化能耗管理,将负荷预测结果与实际负荷值对比,可以快速识别预测与实际之间的差距,这一反馈机制使得系统能够及时调整能耗管理策略,提高资源利用效率。

[0055] 在本实施例中,还包括:

划分模块,用于基于所述传感器预设的工作范围,通过所述建筑终端划分所述传感器在建筑内覆盖的采集区域;

第三判断模块,用于判断所述采集区域能否检测到预设的重叠监测边缘;

第三执行模块,用于若能,则根据所述重叠监测边缘构建重叠时间窗口,标识多个传感器在所述重叠时间窗口的采集测量值,依据所述采集测量值生成待融合的数据特征,应用预设的卡尔曼滤波为所述采集测量值进行权重分配,计算所述数据特征的加权平均

值,其中,所述数据特征具体包括温度、湿度和能耗。

[0056] 在本实施例中,系统基于传感器在建筑内预先设好的工作范围,通过建筑终端划分传感器在建筑内覆盖的采集区域,而后系统判断这些采集区域能否检测到预先设定的重叠监测边缘,以执行对应的步骤;例如,当系统判定到传感器在建筑内覆盖的采集区域并未检测到预先设定的重叠监测边缘时,则系统会认为传感器的工作范围划分得当,每个传感器覆盖的采集区域均相对独立,避免了重复监测,系统会继续采集各个区域的能耗数据,按预定的频率将数据传输到建筑终端,同时通过监控数据传输的稳定性,确保各个传感器在独立采集区域时的数据能够按时、完整地传输到终端,并且尽管采集区域无重叠,系统也可以定期检查传感器的状态,包括电池寿命、信号强度等,以确保长期稳定运行,而由于传感器覆盖区域未重叠,系统需定期核查是否有区域数据遗漏,如果发现某些区域的数据采集有遗漏或延迟,需要采取修复措施;例如,当系统判定到传感器在建筑内覆盖的采集区域检测到预先设定的重叠监测边缘,此时系统会认为传感器的工作范围存在多次覆盖检测的区域,系统会根据重叠监测边缘构建重叠时间窗口,标识多个传感器在重叠时间窗口的采集测量值,依据不同的采集测量值生成待融合的数据特征,数据特征具体包括温度、湿度和能耗,应用预先设好的卡尔曼滤波为采集测量值进行权重分配,计算数据特征的加权平均值;系统通过在重叠区域内结合来自多个传感器的测量值,能够减少单个传感器可能带来的误差,这种冗余数据可以帮助确认和校正异常值,从而提高整体数据的准确性,同时重叠监测允许系统利用多个来源的数据特征,从而提升数据的可靠性,通过卡尔曼滤波,系统能够在存在噪声和不确定性的情况下,通过加权平均来获得更稳定和可靠的数据特征,并且依据重叠时间窗口内的多个测量值生成待融合的数据特征,确保了数据融合的多样性和全面性,不同传感器在同一时刻的测量结果可以提供更丰富的信息,使得最终融合的数据特征更具代表性,而卡尔曼滤波根据实时测量的变化自动调整不同传感器的权重,使得系统能够根据实际情况动态优化数据处理,这种自适应能力在变化环境中尤为重要,可以实时响应传感器状态的变化。

[0057] 在本实施例中,执行模块还包括:

计算单元,用于基于预设的虚拟变量,构建所述建筑终端对能耗的季节性特征,计算所述季节性特征之间的比率特征和差值特征,其中,所述虚拟变量具体包括春季、夏季、秋季和冬季,所述比率特征具体为能耗与温度的比率,所述差值特征具体为当前负荷与过去负荷的差值;

判断单元,用于判断所述比率特征和所述差值特征是否存在缺失值;

执行单元,用于若否,则对所述比率特征和所述差值特征进行标准化处理,确保所述比率特征和所述差值特征始终处于同一尺度上,将所述比率特征和所述差值特征与其他特征进行组合,形成复合特征,根据所述复合特征捕捉季节性变化对应的能耗影响。

[0058] 在本实施例中,系统基于预先设好的虚拟变量,虚拟变量具体包括春季、夏季、秋季和冬季,构建建筑终端对能耗的季节性特征,计算季节性特征之间的比率特征和差值特征,比率特征具体为能耗与温度的比率,差值特征具体为当前负荷与过去负荷的差值,而后系统判断这些比率特征和差值特征是否存在缺失值,以执行对应的步骤;例如,当系统判定到季节性特征之间的比率特征和差值特征存在缺失值时,则系统会认为这些特征可能无法准确反映能耗的季节性变化,这可能会影响后续的分析 and 预测,系统会记录缺失值的具体

位置(时间点和特征),计算缺失值的比例,以评估数据完整性,同时对数据进行重采样,选择时间段内的代表性数据进行计算,确保数据的时效性和连续性,并且对于无法合理填补的缺失值,可以选择标记为“缺失”或“异常”,在后续分析中予以剔除或单独处理,在处理完缺失值后,重新计算比率特征和差值特征,评估其对后续模型的影响;例如,当系统判定到季节性特征之间的比率特征和差值特征不存在缺失值,此时系统会认为这些特征能够准确反映能耗的季节性变化,系统会对这些比率特征和差值特征进行标准化处理,确保比率特征和差值特征始终处于同一尺度上,将这些比率特征和差值特征进行组合,形成复合特征,根据复合特征捕捉季节性变化对应的能耗影响;系统通过对比率特征和差值特征进行标准化处理,可以确保所有特征都在同一尺度上进行比较,这种标准化可以防止某些特征因为数值范围较大而主导其他特征,使模型更加平衡,提升预测精度,例如能耗与温度的比率值可能相对较小,而能耗差值可能较大,标准化处理后可以让它们在相同的范围内进行计算,避免特征之间的不平衡,同时复合特征通过组合不同的比率特征和差值特征,能够捕捉到更加复杂的能耗模式。这种组合特征在模型中可以揭示出季节性变化与能耗波动之间更深层次的关联,使得模型不仅仅依赖单一特征,更能够从多个维度全面分析能耗变化的驱动因素,这将有助于提升模型的预测准确性,并且比率特征和差值特征的组合能够精确地捕捉不同季节下能耗的变化趋势,例如冬季的负荷变化与春夏季可能有不同的能耗驱动因素,通过分析这些复合特征,系统能够识别出冬季或夏季中特定的能耗模式,从而优化负荷预测,而复合特征可以为预测模型提供更丰富的信息,使得预测结果更具解释性,决策者可以基于复合特征,明确不同季节对能耗的具体影响因素,从而在实际操作中更好地理解能耗的波动来源,并采取相应的措施。

[0059] 在本实施例中,第二执行模块还包括:

绘制单元,用于基于所述建筑终端对能耗的预设图表,从所述图表上绘制X轴为能耗预测值,并从所述图表上绘制Y轴为能耗残差,结合得到能耗散点图;

第二判断单元,用于判断所述能耗散点图上的残差是否随机分布;

第二执行单元,用于若是,则将所述能耗散点图输入至所述基线模型,通过所述基线模型进行交叉验证,获取所述基线模型在不同数据子集上的波动信息。

[0060] 在本实施例中,系统基于建筑终端对能耗预先设置的图表,从图表上绘制X轴为能耗预测值,并在图表上绘制Y轴为能耗残差,结合得到能耗散点图,而后系统判断该能耗散点图上的残差是否随机分布,以执行对应的步骤;例如,当系统判定到能耗散点图上的残差并非随机分布时,则系统会认为基线模型可能存在系统性偏差,模型并未准确捕捉到能耗与相关特征之间的真实关系,系统会检查模型使用的特征是否涵盖了所有可能影响能耗的关键因素,尤其是对天气、经济活动、建筑设备运行状态等外部因素的反映是否充分,同时通过引入更多虚拟变量、调整特征的构造方式或考虑交互特征来增强模型对能耗模式的识别能力,并且再次生成能耗散点图,检查残差分布是否呈现出随机性,确保模型能够较好地反映实际能耗波动;例如,当系统判定到能耗散点图上的残差属于随机分布,此时系统会认为基线模型能够准确捕捉到能耗与相关特征之间的真实关系,系统会将能耗散点图输入至基线模型,通过基线模型进行交叉验证,获取基线模型在不同数据子集上的波动信息;系统通过在不同数据子集上进行交叉验证,可以评估模型的稳定性和一致性,确保模型在不同场景下都能保持良好的预测性能,同时在不同数据子集上进行交叉验证,可以获取多组模

型性能指标(如均方误差、决定系数等),从而全面了解模型的预测能力,并且交叉验证过程中可以调整模型的超参数,以寻找最佳的模型配置,提高模型的预测准确性和鲁棒性,而交叉验证过程中,可以分析各特征对模型预测结果的影响,帮助识别出哪些特征在不同条件下仍然具有重要作用,从而为后续的特征选择和模型改进提供依据。

[0061] 在本实施例中,判断模块还包括:

检测单元,用于基于所述建筑终端预设的网络带宽,检测所述能耗数据进行数据传输时的网络延迟信息;

第三判断单元,用于判断所述网络延迟信息是否超出预设的延迟时段;

第三执行单元,用于若是,则在所述能耗数据的数据采集环节和数据展示环节中添加时间戳标签,根据所述时间戳标签计算所述能耗数据的整体延迟。

[0062] 在本实施例中,系统基于建筑终端预先设定的网络带宽,检测能耗数据进行数据传输时的网络延迟信息,而后系统判断这些网络延迟信息是否超出预先设定的延迟时段,以执行对应的步骤;例如,当系统判定到能耗数据进行数据传输时的网络延迟信息并未超出预先设定的延迟时段时,则系统会认为网络连接正常,数据传输顺畅,能够及时反映建筑内的实时能耗情况,系统会保持当前的数据传输状态,确保能耗数据持续上传至建筑终端,同时记录当前的延迟信息,并更新系统监控面板,确保实时监控数据传输状态,并且进行数据完整性检查,确保在传输过程中未出现数据丢失或损坏,生成实时反馈报告,向管理人员展示当前能耗数据的状态及网络传输情况,确保所有相关人员了解系统运行情况;例如,当系统判定到能耗数据进行数据传输时的网络延迟信息超出了预先设定的延迟时段,此时系统会认为数据传输异常,无法及时反映建筑内的实时能耗情况,系统会在能耗数据的数据采集环节和数据展示环节中添加时间戳标签,根据不同的时间戳标签自适应计算能耗数据的整体延迟;系统通过时间戳标签,可以精确记录每个数据点的采集和传输时间,帮助系统更准确地反映能耗情况,减少由于延迟导致的误判,同时通过对不同时间戳标签的分析,系统能够识别延迟的模式和趋势,从而理解在特定条件下延迟增加的原因,便于后续优化网络配置,并且在发生网络延迟时,系统可以快速定位到具体的延迟环节(如数据采集或数据传输),便于技术人员针对性地进行故障排除和修复,而整体延迟的自适应计算可以帮助系统调整数据处理策略,比如在网络条件不佳时,降低数据传输频率或选择更高效的数据压缩方式。

[0063] 在本实施例中,第二判断模块还包括:

生成单元,用于基于所述有效特征集的预统计特征,计算特征缺失值的比例,根据所述比例获取所述特征缺失值的分布,生成对应的缺失值热图;

第四判断单元,用于判断所述缺失值热图是否匹配预设的随机缺失;

第四执行单元,用于若否,则识别所述特征缺失值与所述预统计特征的关联信息,依据所述关联信息采集对应的缺失出处。

[0064] 在本实施例中,系统基于有效特征集的预先统计特征,计算特征缺失值的比例,根据不同比例获取特征缺失值的分布,生成对应的缺失值热图,而后系统判断该缺失值热图是否匹配预先设定的随机缺失,以执行对应的步骤;例如,当系统判定到缺失值热图能够匹配预先设定的随机缺失时,则系统会认为这些特征的缺失值分布没有受到系统性因素的影响,缺失值是随机产生的,而不是由于某些特定原因导致的,系统会记录缺失值热图的结果

及其匹配情况,生成分析报告,以便团队了解缺失数据的状态及其对分析的影响,同时由于缺失值是随机的,系统可以继续评估和验证模型的稳定性和准确性,而不需要对数据进行重大的调整,并且虽然当前缺失值是随机的,系统仍需监控后续数据的缺失情况,以确保数据质量持续符合预期标准,而通过分析缺失值的模式,系统可以改进数据收集策略,以减少未来的缺失值出现,从而提升数据的整体质量;例如,当系统判定到缺失值热图无法匹配预先设定的随机缺失,此时系统会认为这些特征的缺失值分布受到影响,系统会识别特征缺失值与预先统计特征的关联信息,依据不同的关联信息采集对应的缺失出处;系统通过分析缺失值与其他特征的关联,能够识别出导致缺失的潜在原因,如传感器故障、数据采集错误或外部环境因素,有助于针对性地解决问题,同时识别非随机缺失值可以帮助模型设计者理解数据集的局限性,在构建模型时考虑这些因素,增强模型对缺失数据的鲁棒性,并且识别特征缺失的来源可以为建筑管理者提供重要的信息,帮助他们在决策时考虑这些数据缺失对能耗分析和预测的潜在影响。

[0065] 在本实施例中,统一模块还包括:

识别单元,用于检测所述能耗数据中的原始单位,基于所述原始单位识别出待转换的不同单位;

第五判断单元,用于判断所述不同单位能否统一转换为所述数据单位类型;

第五执行单元,用于若否,则标记所述不同单位的不一致项,根据所述不一致项对所述不同单位的对应数据逐行进行单位转换,比对转换前后的数据分布和统计特性。

[0066] 在本实施例中,系统通过检测能耗数据中的原始单位,基于各个原始单位识别出待转换的不同单位,而后系统判断这些不同单位能否统一转换为数据单位类型,以执行对应的步骤;例如,当系统判定到待转换的不同单位能够统一转换为数据单位类型时,则系统会认为这些单位之间具有相同的量纲,可以通过标准的转换公式或比例进行转换,系统会明确各个原始单位与数据单位类型之间的转换关系,同时根据定义好的转换规则,逐项转换待转换的数据,并且将转换后的数据更新到数据集,并记录原始单位和转换后的单位,以便追溯和审计,而一旦数据统一为标准的单位类型,可以进行进一步的数据分析、建模和可视化工作;例如,当系统判定到待转换的不同单位无法统一转换为数据单位类型,此时系统会认为这些单位之间不具备相同量纲,无法统一进行转换,系统会标记不同单位的不一致项,根据这些不一致项对不同单位的对应数据逐行进行单位转换,比对转换前后的数据分布和统计特性;系统通过逐行进行单位转换能够避免全局处理时可能引入的错误或偏差,尤其是在复杂的情况下,如单位之间的量纲差异较大或转换规则不明确,系统根据每一条数据的特性单独处理,确保数据转换的精确性,同时在遇到无法转换的单位时,系统不会简单地丢弃数据,而是通过比对转换前后的数据分布和统计特性,确保保留有意义的信息,这种方式可以在复杂数据环境中保持较高的数据完整性,并且标记不同单位的不一致项,确保在后续分析过程中可以追溯这些不一致项,防止混淆或误用错误单位的数据,这样可以在问题出现时快速定位并修正。

[0067] 尽管已经示出和描述了本发明的实施例,对于本领域的普通技术人员而言,可以理解在不脱离本发明的原理和精神的情况下可以对这些实施例进行多种变化、修改、替换和变型,本发明的范围由所附权利要求及其等同物限定。

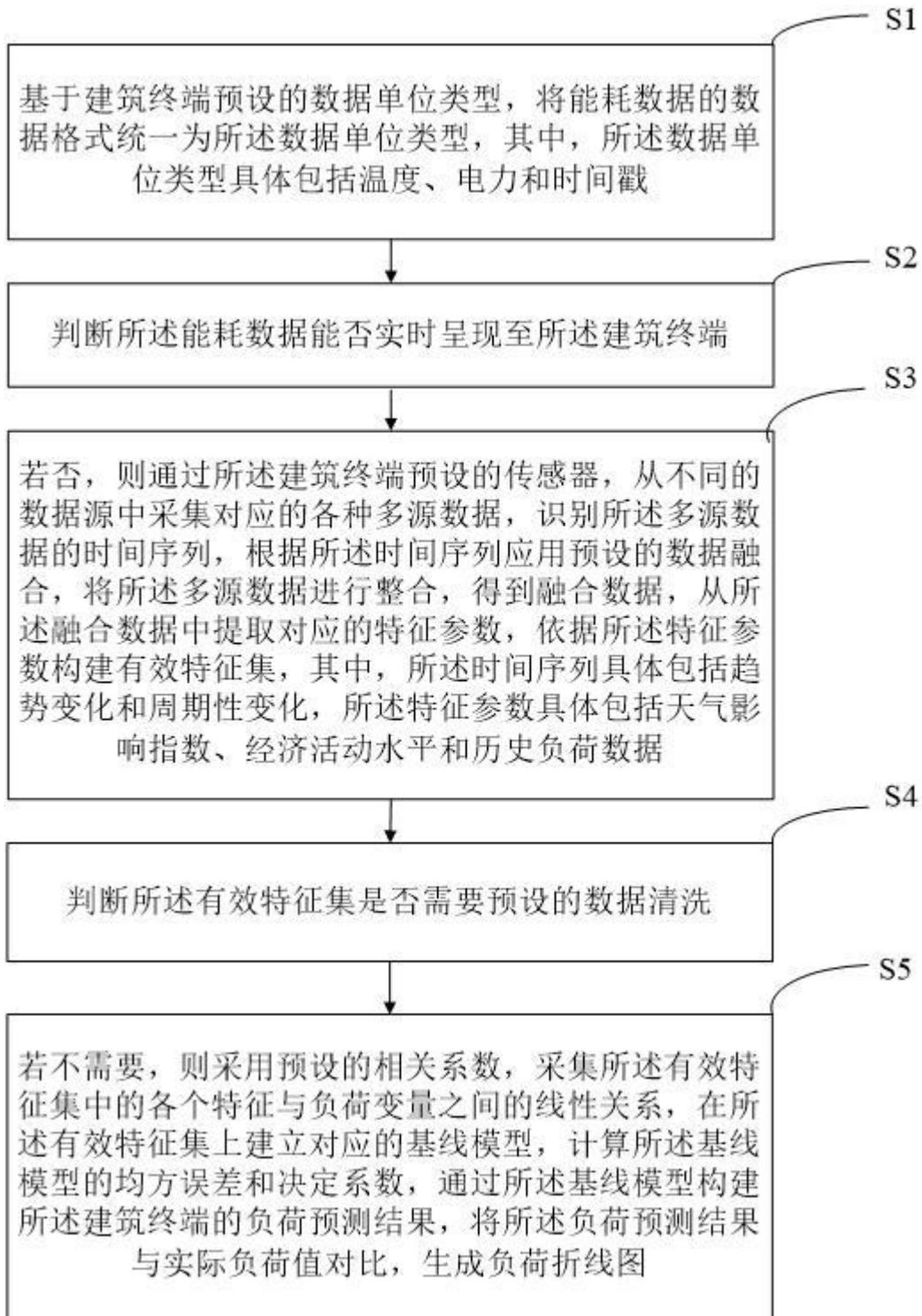


图 1

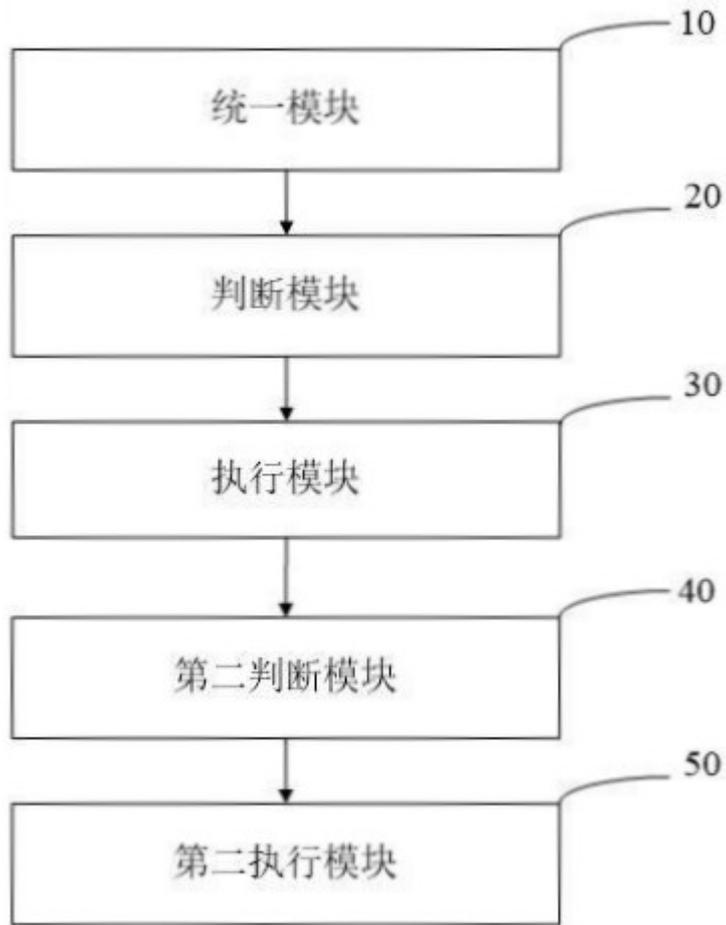


图 2