



(12) 发明专利申请

(10) 申请公布号 CN 116028838 A

(43) 申请公布日 2023. 04. 28

(21) 申请号 202310037388.1

G06F 16/27 (2019.01)

(22) 申请日 2023.01.09

(71) 申请人 广东电网有限责任公司

地址 510000 广东省广州市越秀区东风东
路757号

(72) 发明人 彭勃 左婧 龚贤夫 李耀东

李作红 郑嘉鹏 邱健

(74) 专利代理机构 广州三环专利商标代理有限

公司 44202

专利代理师 李妙芬

(51) Int. Cl.

G06F 18/23213 (2023.01)

G06F 18/214 (2023.01)

G06Q 10/04 (2023.01)

G06Q 50/06 (2012.01)

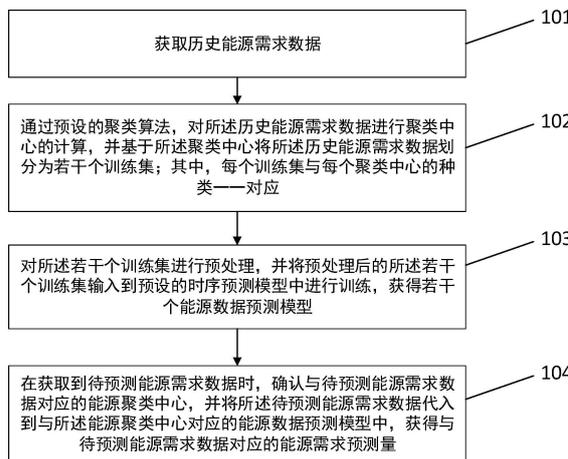
权利要求书2页 说明书11页 附图3页

(54) 发明名称

一种基于聚类算法的能源数据处理方法、装置及终端设备

(57) 摘要

本发明公开了一种基于聚类算法的能源数据处理方法、装置及终端设备,获取历史能源需求数据,通过预设的聚类算法对历史能源需求数据进行聚类中心的计算,并基于聚类中心将历史能源需求数据划分训练集,并将训练集输入到时序预测模型中进行训练,获得能源数据预测模型,并在获得待预测能源需求数据时,确认待预测能源需求数据的能源聚类中心,从而确定与能源聚类中心对应的能源数据预测模型,进而通过模型获得能源需求预测量。本发明通过聚类算法对数据进行聚类分类,使得分类后的数据具有更明显的特征,基于聚类分类后的数据进行多个时序预测模型的训练,有利于提高时序预测模型的预测精度,进而提高能源需求数据预测的准确性。



1. 一种基于聚类算法的能源数据处理方法,其特征在于,包括:

获取历史能源需求数据;

通过预设的聚类算法,对所述历史能源需求数据进行聚类中心的计算,并基于所述聚类中心将所述历史能源需求数据划分为若干个训练集;其中,每个训练集与每个聚类中心的种类一一对应;

对所述若干个训练集进行预处理,并将预处理后的所述若干个训练集输入到预设的时序预测模型中进行训练,获得若干个能源数据预测模型;

在获取到待预测能源需求数据时,确认与待预测能源需求数据对应的能源聚类中心,并将所述待预测能源需求数据代入到与所述能源聚类中心对应的能源数据预测模型中,获得与待预测能源需求数据对应的能源需求预测量。

2. 根据权利要求1所述的基于聚类算法的能源数据处理方法,其特征在于,所述通过预设的聚类算法,对所述历史能源需求数据进行聚类中心的计算,具体为:

在所述历史能源需求数据中,随机选取预设数量个数据点作为若干个初始聚类中心;

将所述若干个初始聚类中心设为若干个待定聚类中心,依次执行欧氏距离计算操作、聚类中心更新操作和聚类中心比较操作,直至更新前的待定聚类中心和更新后的待定聚类中心的欧式距离小于第一门檻值后停止,并将更新后的待定聚类中心作为目标聚类中心;

其中,所述欧氏距离计算操作,具体为:计算每个数据点与所有待定聚类中心的欧氏距离,并选择欧氏距离最小的待定聚类中心作为每个数据点的类别,获得历史能源需求数据对应的聚类数据;所述聚类中心更新操作,具体为:根据欧氏距离计算操作获得的聚类数据,计算每个类别对应的若干数据点的均值,并将均值更新为下一次欧氏距离计算操作所需的待定聚类中心;所述聚类中心比较操作,具体为:对更新前的待定聚类中心和更新后的待定聚类中心进行欧式距离的计算,并根据计算获得的欧氏距离与第一门檻值进行比较,若计算结果大于第一门檻值,则根据更新后的待定聚类中心重新执行欧氏距离计算操作、聚类中心更新操作和聚类中心比较操作。

3. 根据权利要求2所述的基于聚类算法的能源数据处理方法,其特征在于,所述基于所述聚类中心将所述历史能源需求数据划分为若干个训练集,具体为:

计算历史能源需求数据中每个数据点与目标聚类中心的欧氏距离,并选取欧式距离最小的目标聚类中心作为每个数据点的类别标号,将历史能源需求数据所有数据点按照类别标号分类为若干个训练集;其中,所述训练集的个数与所述目标聚类中心的个数一致,且一一对应。

4. 根据权利要求3所述的基于聚类算法的能源数据处理方法,其特征在于,所述对所述若干个训练集进行预处理,具体为:在每个训练集中,按照时间顺序将每个训练集中的数据点进行排序,并通过插值法按照预设的时间颗粒度对每个训练集进行数据补充操作;其中所述数据补充操作具体为:对按照预设的时间颗粒度对每个训练集中的数据点对应的时间进行缺失识别;若数据点的时间缺失,则通过插值法进行数据补充,并完成预处理;若数据点的时间未缺失,则直接完成预处理。

5. 根据权利要求4所述的基于聚类算法的能源数据处理方法,其特征在于,所述将预处理后的所述若干个训练集输入到预设的时序预测模型中进行训练,获得若干个能源数据预测模型,具体为:

构建与训练集个数对应的若干个时序预测模型；其中，每个训练集分别与每个时序预测模型相对应；

将每个训练集分别输入至每个时序预测模型中，通过Xavier算法将每个时序预测模型的参数进行初始化操作，并通过Adam梯度算法、平均绝对误差损失函数对每个输入训练集的时序预测模型进行训练，并验证集进行验证，当所有时序预测模型收敛时，获得若干个能源数据预测模型。

6. 根据权利要求5所述的基于聚类算法的能源数据处理方法，其特征在于，所述确认与待预测能源需求数据对应的能源聚类中心，并将所述待预测能源需求数据代入到与所述能源聚类中心对应的能源数据预测模型中，获得与待预测能源需求数据对应的能源需求预测量，具体为：

计算所述待预测能源需求数据与目标聚类中心的平均距离；

选取平均距离小于或等于第二阈值值的若干个目标聚类中心作为能源聚类中心，将所述待预测能源需求数据分别代入到若干个所述能源聚类中心对应的能源数据预测模型中，获得若干预测结果；

对所述待预测能源需求数据与若干能源聚类中心的平均距离进行归一化处理，并将归一化处理后若干能源聚类中心的平均距离与若干预测结果代入加权计算公式中，计算获得与待预测能源需求数据对应的能源需求预测量。

7. 根据权利要求1所述的基于聚类算法的能源数据处理方法，其特征在于，所述历史能源需求数据，包括：历史气象数据、时间数据、电力能源数据、电网调度数据、社会经济数据和碳排放数据。

8. 一种基于聚类算法的能源数据处理装置，其特征在于，包括：数据获取模块、聚类数据模块、模型训练模块和结果生成模块；

所述数据获取模块，用于获取历史能源需求数据；

所述聚类数据模块，用于通过预设的聚类算法，对所述历史能源需求数据进行聚类中心的计算，并基于所述聚类中心将所述历史能源需求数据划分为若干个训练集；其中，每个训练集与每个聚类中心的种类一一对应；

所述模型训练模块，用于对所述若干个训练集进行预处理，并将预处理后的所述若干个训练集输入到预设的时序预测模型中进行训练，获得若干个能源数据预测模型；

所述结果生成模块，用于在获取到待预测能源需求数据时，确认与待预测能源需求数据对应的能源聚类中心，并将所述待预测能源需求数据代入到与所述能源聚类中心对应的能源数据预测模型中，获得与待预测能源需求数据对应的能源需求预测量。

9. 一种计算机终端设备，其特征在于，包括处理器、存储器以及存储在所述存储器中且被配置为由所述处理器执行的计算机程序，所述处理器执行所述计算机程序时实现如权利要求1至7中任意一项所述的一种基于聚类算法的能源数据处理方法。

10. 一种计算机可读存储介质，其特征在于，所述计算机可读存储介质包括存储的计算机程序，其中，在所述计算机程序运行时控制所述计算机可读存储介质所在设备执行如权利要求1至7中任意一项所述的一种基于聚类算法的能源数据处理方法。

一种基于聚类算法的能源数据处理方法、装置及终端设备

技术领域

[0001] 本发明涉及神经网络领域,尤其涉及一种基于聚类算法的能源数据处理方法、装置及终端设备。

背景技术

[0002] 电力、煤、石油、天然气目前主要预测方法为以时间序列法为代表的传统预测方法和以人工智能技术为代表的人工智能预测法。传统的预测方法对数据质量要求比较严苛,模型相对较为简单,随着人工智能、云计算等信息技术在能源领域的广泛应用,新模式、新业态蓬勃兴起,传统的规划方法、分析模型难以适应研究需要,因此,人工智能预测法得到了众多研究者的关注。人工智能预测法大多使用CNN(卷积神经网络)、LSTM(长短期记忆人工神经网络)等人工智能神经网络通过对于一种或多种能源数据进行训练,提取数据特征,对能源需求进行预测。

[0003] 目前针对能源需求预测的人工智能预测法大多只使用单一模型进行训练与预测。一方面单一模型的损失函数容易陷入到局部最优解当中,无法向全局最优解继续下降;另一方面,对于输入数据质量要求较高,若输入数据不平衡、数据噪声过多,单一模型可能会出现对于训练数据过拟合的现象,鲁棒性较差。因此,现有的对于能源需求数据预测的准确性较低。

[0004] 因此,亟需能源数据处理策略,来解决能源需求数据预测的准确性较低的问题。

发明内容

[0005] 本发明实施例提供一种基于聚类算法的能源数据处理方法、装置及终端设备,以提高能源需求数据预测的准确性。

[0006] 为了解决上述问题,本发明一实施例提供一种基于聚类算法的能源数据处理方法,包括:

[0007] 获取历史能源需求数据;

[0008] 通过预设的聚类算法,对所述历史能源需求数据进行聚类中心的计算,并基于所述聚类中心将所述历史能源需求数据划分为若干个训练集;其中,每个训练集与每个聚类中心的种类一一对应;

[0009] 对所述若干个训练集进行预处理,并将预处理后的所述若干个训练集输入到预设的时序预测模型中进行训练,获得若干个能源数据预测模型;

[0010] 在获取到待预测能源需求数据时,确认与待预测能源需求数据对应的能源聚类中心,并将所述待预测能源需求数据代入到与所述能源聚类中心对应的能源数据预测模型中,获得与待预测能源需求数据对应的能源需求预测量。

[0011] 作为上述方案的改进,所述通过预设的聚类算法,对所述历史能源需求数据进行聚类中心的计算,具体为:

[0012] 在所述历史能源需求数据中,随机选取预设数量个数据点作为若干个初始聚类中

心；

[0013] 将所述若干个初始聚类中心设为若干个待定聚类中心，依次执行欧氏距离计算操作、聚类中心更新操作和聚类中心比较操作，直至更新前的待定聚类中心和更新后的待定聚类中心的欧式距离小于第一门槛值后停止，并将更新后的待定聚类中心作为目标聚类中心；

[0014] 其中，所述欧氏距离计算操作，具体为：计算每个数据点与所有待定聚类中心的欧氏距离，并选择欧氏距离最小的待定聚类中心作为每个数据点的类别，获得历史能源需求数据对应的聚类数据；所述聚类中心更新操作，具体为：根据欧氏距离计算操作获得的聚类数据，计算每个类别对应的若干数据点的均值，并将均值更新为下一次欧氏距离计算操作所需的待定聚类中心；所述聚类中心比较操作，具体为：对更新前的待定聚类中心和更新后的待定聚类中心进行欧式距离的计算，并根据计算获得的欧氏距离与第一门槛值进行比较，若计算结果大于第一门槛值，则根据更新后的待定聚类中心重新执行欧氏距离计算操作、聚类中心更新操作和聚类中心比较操作。

[0015] 作为上述方案的改进，所述基于所述聚类中心将所述历史能源需求数据划分为若干个训练集，具体为：

[0016] 计算历史能源需求数据中每个数据点与目标聚类中心的欧氏距离，并选取欧式距离最小的目标聚类中心作为每个数据点的类别标号，将历史能源需求数据所有数据点按照类别标号分类为若干个训练集；其中，所述训练集的个数与所述目标聚类中心的个数一致，且一一对应。

[0017] 作为上述方案的改进，所述对所述若干个训练集进行预处理，具体为：在每个训练集中，按照时间顺序将每个训练集中的数据点进行排序，并通过插值法按照预设的时间颗粒度对每个训练集进行数据补充操作；其中所述数据补充操作具体为：对按照预设的时间颗粒度对每个训练集中的数据点对应的的时间进行缺失识别；若数据点的时间缺失，则通过插值法进行数据补充，并完成预处理；若数据点的时间未缺失，则直接完成预处理。

[0018] 作为上述方案的改进，所述将预处理后的所述若干个训练集输入到预设的时序预测模型中进行训练，获得若干个能源数据预测模型，具体为：

[0019] 构建与训练集个数对应的若干个时序预测模型；其中，每个训练集分别与每个时序预测模型相对应；

[0020] 将每个训练集分别输入至每个时序预测模型中，通过Xavier算法将每个时序预测模型的参数进行初始化操作，并通过Adam梯度算法、平均绝对误差损失函数对每个输入训练集的时序预测模型进行训练，并验证集进行验证，当所有时序预测模型收敛时，获得若干个能源数据预测模型。

[0021] 作为上述方案的改进，所述确认与待预测能源需求数据对应的能源聚类中心，并将所述待预测能源需求数据代入到与所述能源聚类中心对应的能源数据预测模型中，获得与待预测能源需求数据对应的能源需求预测量，具体为：

[0022] 计算所述待预测能源需求数据与目标聚类中心的平均距离；

[0023] 选取平均距离小于或等于第二门槛值的若干个目标聚类中心作为能源聚类中心，将所述待预测能源需求数据分别代入到若干个所述能源聚类中心对应的能源数据预测模型中，获得若干预测结果；

[0024] 对所述待预测能源需求数据与若干能源聚类中心的平均距离进行归一化处理,并将归一化处理后若干能源聚类中心的平均距离与若干预测结果代入加权计算公式中,计算获得与待预测能源需求数据对应的能源需求预测量。

[0025] 作为上述方案的改进,所述历史能源需求数据,包括:历史气象数据、时间数据、电力能源数据、电网调度数据、社会经济数据和碳排放数据。

[0026] 相应的,本发明一实施例还提供了一种基于聚类算法的能源数据处理装置,包括:数据获取模块、聚类数据模块、模型训练模块和结果生成模块;

[0027] 所述数据获取模块,用于获取历史能源需求数据;

[0028] 所述聚类数据模块,用于通过预设的聚类算法,对所述历史能源需求数据进行聚类中心的计算,并基于所述聚类中心将所述历史能源需求数据划分为若干个训练集;其中,每个训练集与每个聚类中心的种类一一对应;

[0029] 所述模型训练模块,用于对所述若干个训练集进行预处理,并将预处理后的所述若干个训练集输入到预设的时序预测模型中进行训练,获得若干个能源数据预测模型;

[0030] 所述结果生成模块,用于在获取到待预测能源需求数据时,确认与待预测能源需求数据对应的能源聚类中心,并将所述待预测能源需求数据代入到与所述能源聚类中心对应的能源数据预测模型中,获得与待预测能源需求数据对应的能源需求预测量。

[0031] 作为上述方案的改进,所述通过预设的聚类算法,对所述历史能源需求数据进行聚类中心的计算,具体为:

[0032] 在所述历史能源需求数据中,随机选取预设数量个数据点作为若干个初始聚类中心;

[0033] 将所述若干个初始聚类中心设为若干个待定聚类中心,依次执行欧氏距离计算操作、聚类中心更新操作和聚类中心比较操作,直至更新前的待定聚类中心和更新后的待定聚类中心的欧式距离小于第一门檻值后停止,并将更新后的待定聚类中心作为目标聚类中心;

[0034] 其中,所述欧氏距离计算操作,具体为:计算每个数据点与所有待定聚类中心的欧氏距离,并选择欧氏距离最小的待定聚类中心作为每个数据点的类别,获得历史能源需求数据对应的聚类数据;所述聚类中心更新操作,具体为:根据欧氏距离计算操作获得的聚类数据,计算每个类别对应的若干数据点的均值,并将均值更新为下一次欧氏距离计算操作所需的待定聚类中心;所述聚类中心比较操作,具体为:对更新前的待定聚类中心和更新后的待定聚类中心进行欧式距离的计算,并根据计算获得的欧氏距离与第一门檻值进行比较,若计算结果大于第一门檻值,则根据更新后的待定聚类中心重新执行欧氏距离计算操作、聚类中心更新操作和聚类中心比较操作。

[0035] 作为上述方案的改进,所述基于所述聚类中心将所述历史能源需求数据划分为若干个训练集,具体为:

[0036] 计算历史能源需求数据中每个数据点与目标聚类中心的欧氏距离,并选取欧式距离最小的目标聚类中心作为每个数据点的类别标号,将历史能源需求数据所有数据点按照类别标号分类为若干个训练集;其中,所述训练集的个数与所述目标聚类中心的个数一致,且一一对应。

[0037] 作为上述方案的改进,所述对所述若干个训练集进行预处理,具体为:在每个训练

集中,按照时间顺序将每个训练集中的数据点进行排序,并通过插值法按照预设的时间颗粒度对每个训练集进行数据补充操作;其中所述数据补充操作具体为:对按照预设的时间颗粒度对每个训练集中的数据点对应的的时间进行缺失识别;若数据点的时间缺失,则通过插值法进行数据补充,并完成预处理;若数据点的时间未缺失,则直接完成预处理。

[0038] 作为上述方案的改进,所述将预处理后的所述若干个训练集输入到预设的时序预测模型中进行训练,获得若干个能源数据预测模型,具体为:

[0039] 构建与训练集个数对应的若干个时序预测模型;其中,每个训练集分别与每个时序预测模型相对应;

[0040] 将每个训练集分别输入至每个时序预测模型中,通过Xavier算法将每个时序预测模型的参数进行初始化操作,并通过Adam梯度算法、平均绝对误差损失函数对每个输入训练集的时序预测模型进行训练,并验证集进行验证,当所有时序预测模型收敛时,获得若干个能源数据预测模型。

[0041] 作为上述方案的改进,所述确认与待预测能源需求数据对应的能源聚类中心,并将所述待预测能源需求数据代入到与所述能源聚类中心对应的能源数据预测模型中,获得与待预测能源需求数据对应的能源需求预测量,具体为:

[0042] 计算所述待预测能源需求数据与目标聚类中心的平均距离;

[0043] 选取平均距离小于或等于第二门檻值的若干个目标聚类中心作为能源聚类中心,将所述待预测能源需求数据分别代入到若干个所述能源聚类中心对应的能源数据预测模型中,获得若干预测结果;

[0044] 对所述待预测能源需求数据与若干能源聚类中心的平均距离进行归一化处理,并将归一化处理后若干能源聚类中心的平均距离与若干预测结果代入加权计算公式中,计算获得与待预测能源需求数据对应的能源需求预测量。

[0045] 作为上述方案的改进,所述历史能源需求数据,包括:历史气象数据、时间数据、电力能源数据、电网调度数据、社会经济数据和碳排放数据。

[0046] 相应的,本发明一实施例还提供了一种计算机终端设备,包括处理器、存储器以及存储在所述存储器中且被配置为由所述处理器执行的计算机程序,所述处理器执行所述计算机程序时实现如本发明所述的一种基于聚类算法的能源数据处理方法。

[0047] 相应的,本发明一实施例还提供了一种计算机可读存储介质,所述计算机可读存储介质包括存储的计算机程序,其中,在所述计算机程序运行时控制所述计算机可读存储介质所在设备执行如本发明所述的一种基于聚类算法的能源数据处理方法。

[0048] 由上可见,本发明具有如下有益效果:

[0049] 本发明提供了一种基于聚类算法的能源数据处理方法,获取历史能源需求数据,通过预设的聚类算法对历史能源需求数据进行聚类中心的计算,并基于聚类中心将历史能源需求数据划分训练集,并将训练集输入到时序预测模型中进行训练,获得能源数据预测模型,并在获得待预测能源需求数据时,确认待预测能源需求数据的能源聚类中心,从而确定与能源聚类中心对应的能源数据预测模型,进而通过模型获得能源需求预测量。本发明通过聚类算法对数据进行聚类分类,使得分类后的数据具有更明显的特征,基于聚类分类后的数据进行多个时序预测模型的训练,有利于提高时序预测模型的预测精度,进而提高能源需求数据预测的准确性。

[0050] 进一步地,通过输入数据和聚类中心的关系,在获得的预测模型中确定与输入数据对应的若干个时序预测模型,能够通过若干个时序预测模型共同确定能源预测量,有利于提高能源需求数据预测的鲁棒性。

附图说明

[0051] 图1是本发明一实施例提供的基于聚类算法的能源数据处理方法的流程示意图;

[0052] 图2是本发明一实施例提供的基于聚类算法的能源数据处理装置的结构示意图;

[0053] 图3是本发明一实施例提供的一种终端设备结构示意图。

具体实施方式

[0054] 下面将结合本发明实施例中的附图,对本发明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。

[0055] 实施例一

[0056] 参见图1,图1是本发明一实施例提供的一种基于聚类算法的能源数据处理方法的流程示意图,如图1所示,本实施例包括步骤101至步骤104,各步骤具体如下:

[0057] 步骤101:获取历史能源需求数据。

[0058] 在本实施例中,所述历史能源需求数据,包括:历史气象数据、时间数据、电力能源数据、电网调度数据、社会经济数据和碳排放数据。

[0059] 在一具体的实施例中,历史气象数据可包括:最高气温、最低气温、平均气温、平均降水量、平均相对湿度、风速、平均气压等数据中的至少一种;电力能源数据可包括:煤机、核电、气电、水电等并网装机容量、出力数据、风光发电利用小时数、能源消费数据(比如单位GDP电耗、单位生产总值能耗、第一产业能耗、人均煤炭消费量等)等数据中的至少一种;社会经济数据可包括GDP、人均收入、城镇化率、第一产业增长率、第二产业增长率、第三产业增长率、人口数、一二三产业占比、分产业GDP等数据中的至少一种;碳排放数据可包括单位生产总值二氧化碳排放等数据;通过获取包含有多类型的能源相关历史数据,按照时间进行排序,并进行归一化处理,得到历史能源需求数据。

[0060] 为更好地说明历史能源需求数据,提供以下例子进行说明:只考虑GDP和能源消费总量,那么数据是二维的,比如经济数据GDP在2021年是114万亿元,能源消费总量是52.4万吨标准煤,那么这个数据就是(2021,114,52.4),最前面2021是年份,代表时间,不是具体数据值。

[0061] 步骤102:通过预设的聚类算法,对所述历史能源需求数据进行聚类中心的计算,并基于所述聚类中心将所述历史能源需求数据划分为若干个训练集;其中,每个训练集与每个聚类中心的种类一一对应。

[0062] 在本实施例中,所述通过预设的聚类算法,对所述历史能源需求数据进行聚类中心的计算,具体为:

[0063] 在所述历史能源需求数据中,随机选取预设数量个数据点作为若干个初始聚类中心;

[0064] 将所述若干个初始聚类中心设为若干个待定聚类中心,依次执行欧氏距离计算操作、聚类中心更新操作和聚类中心比较操作,直至更新前的待定聚类中心和更新后的待定聚类中心的欧式距离小于第一门槛值后停止,并将更新后的待定聚类中心作为目标聚类中心;

[0065] 其中,所述欧氏距离计算操作,具体为:计算每个数据点与所有待定聚类中心的欧氏距离,并选择欧氏距离最小的待定聚类中心作为每个数据点的类别,获得历史能源需求数据对应的聚类数据;所述聚类中心更新操作,具体为:根据欧氏距离计算操作获得的聚类数据,计算每个类别对应的若干数据点的均值,并将均值更新为下一次欧氏距离计算操作所需的待定聚类中心;所述聚类中心比较操作,具体为:对更新前的待定聚类中心和更新后的待定聚类中心进行欧式距离的计算,并根据计算获得的欧氏距离与第一门槛值进行比较,若计算结果大于第一门槛值,则根据更新后的待定聚类中心重新执行欧氏距离计算操作、聚类中心更新操作和聚类中心比较操作。

[0066] 在本实施例中,所述基于所述聚类中心将所述历史能源需求数据划分为若干个训练集,具体为:

[0067] 计算历史能源需求数据中每个数据点与目标聚类中心的欧氏距离,并选取欧式距离最小的目标聚类中心作为每个数据点的类别标号,将历史能源需求数据所有数据点按照类别标号分类为若干个训练集;其中,所述训练集的个数与所述目标聚类中心的个数一致,且一一对应。

[0068] 在一具体的实施例中,为更好地说明训练集的生成,举出一下例子进行说明,包括步骤S01至S05:

[0069] S01:在历史能源需求数据中随机选择k个(即本发明权要所述的预设数量,由用户自行设置)数据点作为初始聚类中心;

[0070] S02:将k个初始聚类中心设置为待定聚类中心,计算历史能源需求数据中所有数据点与k个待定聚类中心的欧式距离,选择距离最近的一个待定聚类中心作为它的类别标号,将所有数据点都进行标号分类(例如:某一个数据点与第二个待定聚类中心的欧式距离最小,那么该数据点的类别为2);

[0071] S03:所有数据点的k个类别分类完成后,在每个类别内,计算所有数据点的均值,将获得的均值更新待定聚类中心;

[0072] S04:根据更新后的待定聚类中心,重复执行S02和S03,直至更新前后的待定聚类中心之间的欧式距离小于第一门槛值 h_1 (threshold),确认更新后的待定聚类中心为目标聚类中心;

[0073] S05:计算历史能源需求数据中每个数据点分别与k个目标聚类中心的欧式距离,选择距离最近的一个目标聚类中心作为历史能源需求数据中每个数据点的类别标号,将所有数据点都进行标号分类,得到k个训练集。

[0074] 步骤103:对所述若干个训练集进行预处理,并将预处理后的所述若干个训练集输入到预设的时序预测模型中进行训练,获得若干个能源数据预测模型。

[0075] 在本实施例中,所述对所述若干个训练集进行预处理,具体为:在每个训练集中,按照时间顺序将每个训练集中的数据点进行排序,并通过插值法按照预设的时间颗粒度对每个训练集进行数据补充操作;其中所述数据补充操作具体为:对按照预设的时间颗粒度

对每个训练集中的数据点对应的时间进行缺失识别;若数据点的时间缺失,则通过插值法进行数据补充,并完成预处理;若数据点的时间未缺失,则直接完成预处理。

[0076] 在一具体的实施例中,在每个训练集中,将数据点按照时间进行排序。利用插值法按照时间颗粒度将缺失时间数据补充完整,若训练集中数据点的时间数据完整,则无需插值法进行补充,可直接用于训练;

[0077] 提供以下例子进行说明:设时间颗粒度为日,目前数据集中有3个数据,分别是(20220103,30), (20220105,50), (20220107,70),前一项代表日期,后一项代表数据值,利用插值法按照日为时间颗粒度得到(20220104,40)和(20220105,50)两个点,共5个点构成了一个训练集。

[0078] 在本实施例中,所述将预处理后的所述若干个训练集输入到预设的时序预测模型中进行训练,获得若干个能源数据预测模型,具体为:

[0079] 构建与训练集个数对应的若干个时序预测模型;其中,每个训练集分别与每个时序预测模型相对应;

[0080] 将每个训练集分别输入至每个时序预测模型中,通过Xavier算法将每个时序预测模型的参数进行初始化操作,并通过Adam梯度算法、平均绝对误差损失函数对每个输入训练集的时序预测模型进行训练,并验证集进行验证,当所有时序预测模型收敛时,获得若干个能源数据预测模型。

[0081] 在一具体的实施例中,时序预测模型可采用基于LSTM的时序预测模型。

[0082] 在一具体的实施例中,能源数据预测模型的训练具体如下:

[0083] 构建k个基于LSTM的时序预测模型并分别使用k个训练集训练k个基于LSTM的时序预测模型。其中LSTM模型输入数据为序列数据,维度为(N,T,D),N为样本数量(number of samples),T为时间长度(time steps),D为单个输入样本维度(dimension)。LSTM模型可根据当前时间及过去时间(t-n,t-1)输入的序列数据,其中正整数n为步长,输出未来时间(t+1,t+n)的多类型能源需求数据,在指定时间的单个需求数据维度为d,其中d为能源类型数量,其中单一维度的数据为对应能源类型的需求数据;

[0084] 使用Xavier算法将LSTM模型参数进行初始化,应用Adam梯度下降算法,使用平均绝对误差(MAE Loss)作为损失函数对模型进行训练,在每一个epoch使用验证集数据进行验证,直至模型收敛,从而获得能源数据预测模型。

[0085] 步骤104:在获取到待预测能源需求数据时,确认与待预测能源需求数据对应的能源聚类中心,并将所述待预测能源需求数据代入到与所述能源聚类中心对应的能源数据预测模型中,获得与待预测能源需求数据对应的能源需求预测量。

[0086] 在一具体的实施例中,待预测能源需求数据可以为已经获得的资源数据;资源数据包括:气象数据、时间数据、电力能源数据、电网调度数据、社会经济数据和碳排放数据中的其中一项或多项,通过确认待预测能源需求数据与能源聚类中心(即确认待预测能源需求数据的种类),选取与带预测能源需求数据对应的能源数据预测模型进行预测,从而获得能源需求预测量。

[0087] 在本实施例中,所述确认与待预测能源需求数据对应的能源聚类中心,并将所述待预测能源需求数据代入到与所述能源聚类中心对应的能源数据预测模型中,获得与待预测能源需求数据对应的能源需求预测量,具体为:

[0088] 计算所述待预测能源需求数据与目标聚类中心的平均距离；

[0089] 选取平均距离小于或等于第二门槛值的若干个目标聚类中心作为能源聚类中心，将所述待预测能源需求数据分别代入到若干个所述能源聚类中心对应的能源数据预测模型中，获得若干预测结果；

[0090] 对所述待预测能源需求数据与若干能源聚类中心的平均距离进行归一化处理，并将归一化处理后若干能源聚类中心的平均距离与若干预测结果代入加权计算公式中，计算获得与待预测能源需求数据对应的能源需求预测量。

[0091] 在一具体的实施例中，计算待预测能源需求数据与k个目标聚类中心的平均距离 d_i ($i=1, 2, \dots, k$)，若目标聚类中心距离小于或等于第二门槛值 h_2 ，则确定w个目标聚类中心，继而确定w个目标聚类中心的对应的能源需求预测模型，将待预测能源需求数据输入w个目标聚类中心对应的能源需求预测模型中得到w个预测结果 y_j ($j=1, 2, \dots, w$ ，其中 $w \leq k$)。

[0092] 将w个预测结果对应的 d_i ($i=1, 2, \dots, w$)进行归一化得到 d'_i ($i=1, 2, \dots, w$)，使得 $\sum_{i=1}^w d'_i = 1$ 。将预测结果融合得到最终的预测结果 $y = \sum_{i=1}^w d'_i * y_i$ ($w \leq k$) (即本发明权要所述的加权计算公式)。

[0093] 为更好地说明预测结果的生成，举出以下例子：预测下一年度能源消耗总量(单位：亿吨标准煤)时，对能源消耗历史数据进行KMeans聚类，并根据聚类中心获得k个训练集，将k个训练集分别输入到基于LSTM的时序预测模型中进行训练，获得k个能源消耗预测模型(每个能源消耗预测模型与聚类中心一一对应)，计算待预测能源消耗总量与k个聚类中心的平均距离，并确定小于或等于第二门槛值的w个聚类中心以及聚类中更新对应的能源消耗预测模型；

[0094] 假设w为3，也就是说得到了3个能源消耗预测模型以及3个聚类中心，求待预测能源消耗总量与这3个聚类中心的平均距离并进行归一化，比如说得到：0.3, 0.5, 0.2。将待预测能源消耗总量输入这3个模型得到3个输出，比如说：45, 50, 55。那么最终的预测结果就是： $0.3*45+0.5*50+0.2*55=49.5$ 。所以最终的预测结果是49.5亿吨标准煤。

[0095] 本实施例通过获取历史能源需求数据，通过预设的聚类算法对历史能源需求数据进行聚类中心的计算，并基于聚类中心将历史能源需求数据划分训练集，并将训练集输入到时序预测模型中进行训练，获得能源数据预测模型，并在获得待预测能源需求数据时，确认待预测能源需求数据的能源聚类中心，从而确定与能源聚类中心对应的能源数据预测模型，进而通过模型获得能源需求预测量。可根据时间段的主要影响因子数据对多种能源类型需求进行预测。最终的预测结果由多个模型共同决定，最大可能对真实数据进行拟合，可大大提高整体预测结果的准确性以及鲁棒性。

[0096] 实施例二

[0097] 参见图2，图2是本发明一实施例提供的一种基于聚类算法的能源数据处理装置的结构示意图，包括：数据获取模块201、聚类数据模块202、模型训练模块203和结果生成模块204；

[0098] 所述数据获取模块201，用于获取历史能源需求数据；

[0099] 所述聚类数据模块202，用于通过预设的聚类算法，对所述历史能源需求数据进行聚类中心的计算，并基于所述聚类中心将所述历史能源需求数据划分为若干个训练集；其中，每个训练集与每个聚类中心的种类一一对应；

[0100] 所述模型训练模块203,用于对所述若干个训练集进行预处理,并将预处理后的所述若干个训练集输入到预设的时序预测模型中进行训练,获得若干个能源数据预测模型;

[0101] 所述结果生成模块204,用于在获取到待预测能源需求数据时,确认与待预测能源需求数据对应的能源聚类中心,并将所述待预测能源需求数据代入到与所述能源聚类中心对应的能源数据预测模型中,获得与待预测能源需求数据对应的能源需求预测量。

[0102] 作为上述方案的改进,所述通过预设的聚类算法,对所述历史能源需求数据进行聚类中心的计算,具体为:

[0103] 在所述历史能源需求数据中,随机选取预设数量个数据点作为若干个初始聚类中心;

[0104] 将所述若干个初始聚类中心设为若干个待定聚类中心,依次执行欧氏距离计算操作、聚类中心更新操作和聚类中心比较操作,直至更新前的待定聚类中心和更新后的待定聚类中心的欧式距离小于第一阈值后停止,并将更新后的待定聚类中心作为目标聚类中心;

[0105] 其中,所述欧氏距离计算操作,具体为:计算每个数据点与所有待定聚类中心的欧氏距离,并选择欧氏距离最小的待定聚类中心作为每个数据点的类别,获得历史能源需求数据对应的聚类数据;所述聚类中心更新操作,具体为:根据欧氏距离计算操作获得的聚类数据,计算每个类别对应的若干数据点的均值,并将均值更新为下一次欧氏距离计算操作所需的待定聚类中心;所述聚类中心比较操作,具体为:对更新前的待定聚类中心和更新后的待定聚类中心进行欧式距离的计算,并根据计算获得的欧氏距离与第一阈值进行比较,若计算结果大于第一阈值,则根据更新后的待定聚类中心重新执行欧氏距离计算操作、聚类中心更新操作和聚类中心比较操作。

[0106] 作为上述方案的改进,所述基于所述聚类中心将所述历史能源需求数据划分为若干个训练集,具体为:

[0107] 计算历史能源需求数据中每个数据点与目标聚类中心的欧氏距离,并选取欧式距离最小的目标聚类中心作为每个数据点的类别标号,将历史能源需求数据所有数据点按照类别标号分类为若干个训练集;其中,所述训练集的个数与所述目标聚类中心的个数一致,且一一对应。

[0108] 作为上述方案的改进,所述对所述若干个训练集进行预处理,具体为:在每个训练集中,按照时间顺序将每个训练集中的数据点进行排序,并通过插值法按照预设的时间颗粒度对每个训练集进行数据补充操作;其中所述数据补充操作具体为:对按照预设的时间颗粒度对每个训练集中的数据点对应的的时间进行缺失识别;若数据点的时间缺失,则通过插值法进行数据补充,并完成预处理;若数据点的时间未缺失,则直接完成预处理。

[0109] 作为上述方案的改进,所述将预处理后的所述若干个训练集输入到预设的时序预测模型中进行训练,获得若干个能源数据预测模型,具体为:

[0110] 构建与训练集个数对应的若干个时序预测模型;其中,每个训练集分别与每个时序预测模型相对应;

[0111] 将每个训练集分别输入至每个时序预测模型中,通过Xavier算法将每个时序预测模型的参数进行初始化操作,并通过Adam梯度算法、平均绝对误差损失函数对每个输入训练集的时序预测模型进行训练,并验证集进行验证,当所有时序预测模型收敛时,获得若干

个能源数据预测模型。

[0112] 作为上述方案的改进,所述确认与待预测能源需求数据对应的能源聚类中心,并将所述待预测能源需求数据代入到与所述能源聚类中心对应的能源数据预测模型中,获得与待预测能源需求数据对应的能源需求预测量,具体为:

[0113] 计算所述待预测能源需求数据与目标聚类中心的平均距离;

[0114] 选取平均距离小于或等于第二阈值值的若干个目标聚类中心作为能源聚类中心,将所述待预测能源需求数据分别代入到若干个所述能源聚类中心对应的能源数据预测模型中,获得若干预测结果;

[0115] 对所述待预测能源需求数据与若干能源聚类中心的平均距离进行归一化处理,并将归一化处理后若干能源聚类中心的平均距离与若干预测结果代入加权计算公式中,计算获得与待预测能源需求数据对应的能源需求预测量。

[0116] 作为上述方案的改进,所述历史能源需求数据,包括:历史气象数据、时间数据、电力能源数据、电网调度数据、社会经济数据和碳排放数据。

[0117] 本实施例通过数据获取模块获取历史能源需求数据,然后通过聚类数据模块对历史能源需求数据进行训练集的划分,再通过模型训练模块对划分好的训练集进行能源数据预测模型的训练,最后在获取到待预测能源需求数据时,通过结果生成模块将待预测能源需求数据输入到相应的能源数据预测模型中,生成能源需求预测量。本实施例通过聚类算法对数据进行聚类分类,使得分类后的数据具有更明显的特征,基于聚类分类后的数据进行多个时序预测模型的训练,有利于提高时序预测模型的预测精度,进而提高能源需求数据预测的准确性。

[0118] 实施例三

[0119] 参见图3,图3是本发明一实施例提供的终端设备结构示意图。

[0120] 该实施例的一种终端设备包括:处理器301、存储器302以及存储在所述存储器302中并可在所述处理器301上运行的计算机程序。所述处理器301执行所述计算机程序时实现上述各个基于聚类算法的能源数据处理方法在实施例中的步骤,例如图1所示的基于聚类算法的能源数据处理方法的所有步骤。或者,所述处理器执行所述计算机程序时实现上述各装置实施例中各模块的功能,例如:图2所示的基于聚类算法的能源数据处理装置的所有模块。

[0121] 另外,本发明实施例还提供了一种计算机可读存储介质,所述计算机可读存储介质包括存储的计算机程序,其中,在所述计算机程序运行时控制所述计算机可读存储介质所在设备执行如上任一实施例所述的基于聚类算法的能源数据处理方法。

[0122] 本领域技术人员可以理解,所述示意图仅仅是终端设备的示例,并不构成对终端设备的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件,例如所述终端设备还可以包括输入输出设备、网络接入设备、总线等。

[0123] 所称处理器301可以是中央处理单元(Central Processing Unit,CPU),还可以是其他通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现成可编程门阵列(Field-Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理

等,所述处理器301是所述终端设备的控制中心,利用各种接口和线路连接整个终端设备的各个部分。

[0124] 所述存储器302可用于存储所述计算机程序和/或模块,所述处理器301通过运行或执行存储在所述存储器内的计算机程序和/或模块,以及调用存储在存储器302内的数据,实现所述终端设备的各种功能。所述存储器302可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序(比如声音播放功能、图像播放功能等)等;存储数据区可存储根据手机的使用所创建的数据(比如音频数据、电话本等)等。此外,存储器可以包括高速随机存取存储器,还可以包括非易失性存储器,例如硬盘、内存、插接式硬盘,智能存储卡(Smart Media Card,SMC),安全数字(Secure Digital,SD)卡,闪存卡(Flash Card)、至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。

[0125] 其中,所述终端设备集成的模块/单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明实现上述实施例方法中的全部或部分流程,也可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一计算机可读存储介质中,该计算机程序在被处理器执行时,可实现上述各个方法实施例的步骤。其中,所述计算机程序包括计算机程序代码,所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、电载波信号、电信信号以及软件分发介质等。

[0126] 需说明的是,以上所描述的装置实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。另外,本发明提供的装置实施例附图中,模块之间的连接关系表示它们之间具有通信连接,具体可以实现为一条或多条通信总线或信号线。本领域普通技术人员在不付出创造性劳动的情况下,即可以理解并实施。

[0127] 以上所述是本发明的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也视为本发明的保护范围。

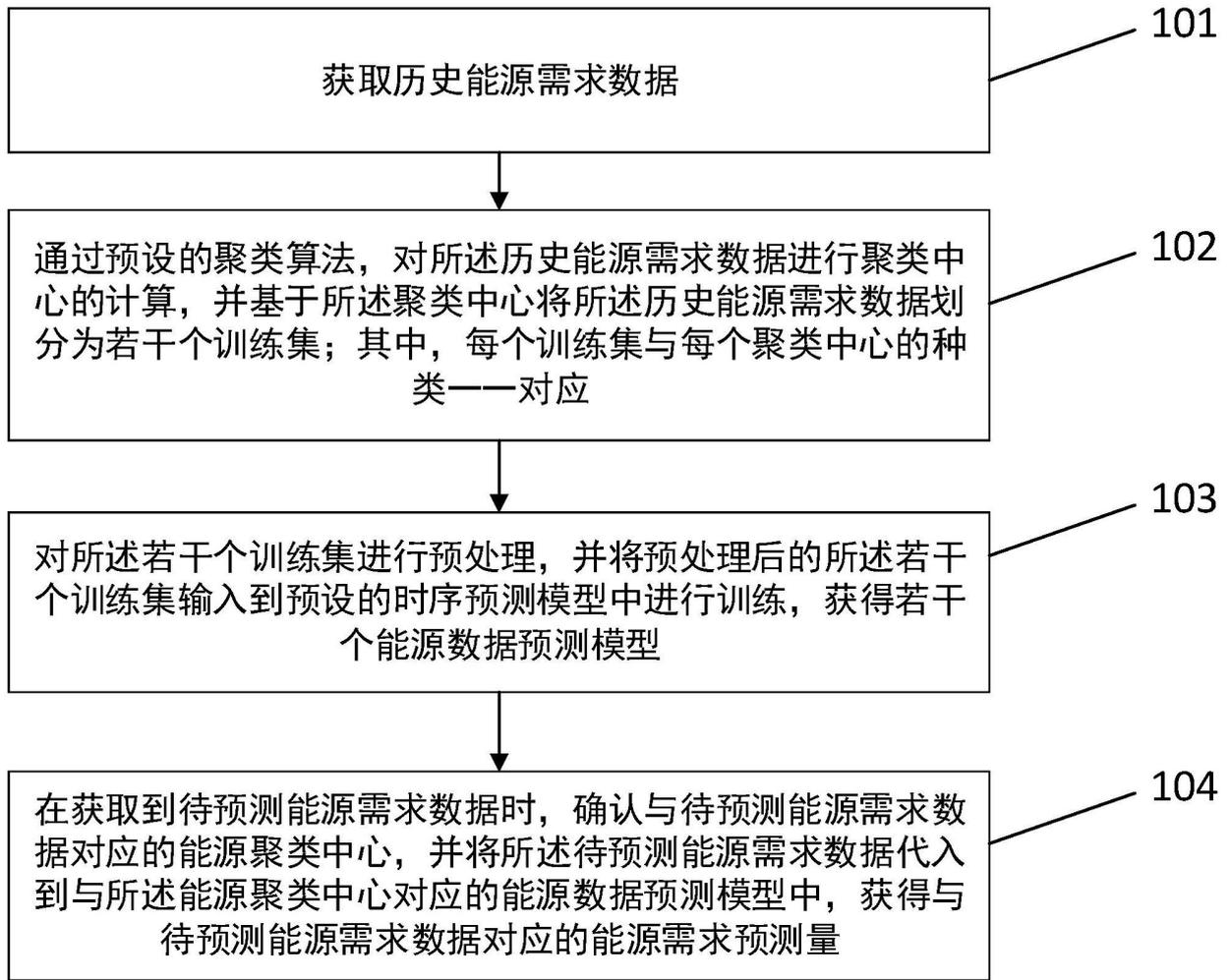


图1



图2

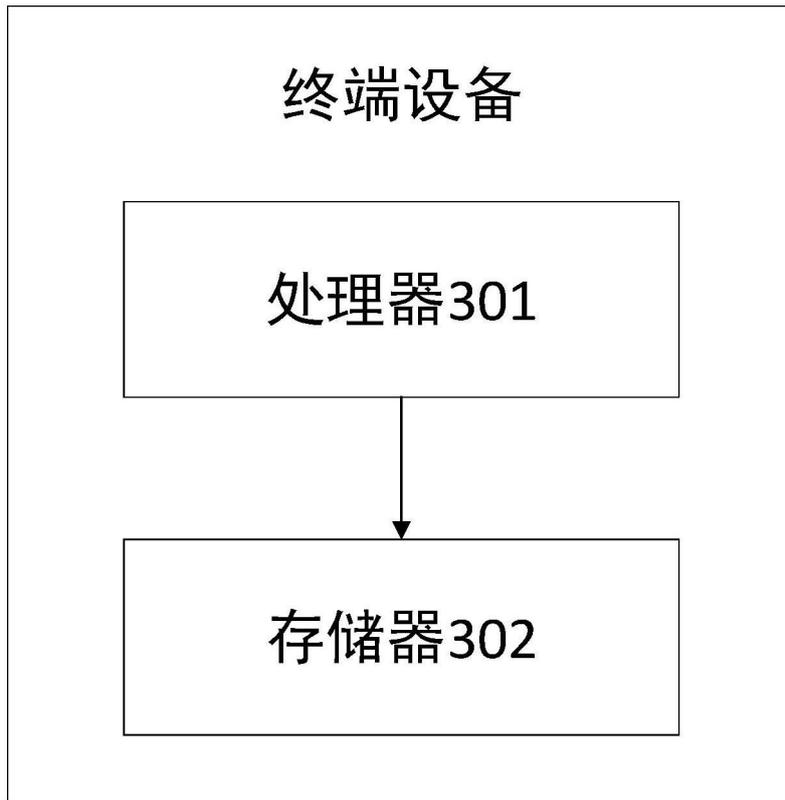


图3