



(12) 发明专利

(10) 授权公告号 CN 113485986 B

(45) 授权公告日 2024. 08. 02

(21) 申请号 202110717117.1

(22) 申请日 2021.06.25

(65) 同一申请的已公布的文献号
申请公布号 CN 113485986 A

(43) 申请公布日 2021.10.08

(73) 专利权人 国网江苏省电力有限公司信息通信分公司

地址 210003 江苏省南京市鼓楼区南瑞路8号

专利权人 南京南瑞信息科技有限公司

(72) 发明人 夏飞 汤铭 王鹏飞 邹昊东
宋浒 胡游君 刘军 邱玉祥
张磊 刘赛 高雪 晁凯 杨颢

(74) 专利代理机构 南京纵横知识产权代理有限公司 32224

专利代理师 母秋松

(51) Int. Cl.
G06F 16/215 (2019.01)
G06N 3/0464 (2023.01)
G06N 3/08 (2023.01)
G06N 3/044 (2023.01)

(56) 对比文件
WO 2015158198 A1, 2015.10.22
CN 110097920 A, 2019.08.06

审查员 谢晓琦

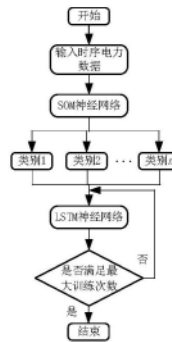
权利要求书2页 说明书5页 附图1页

(54) 发明名称

一种电力数据修复方法

(57) 摘要

本发明公开了一种电力数据修复方法,利用SOM神经网络对历史电力数据集中电力数据进行分类处理;利用Pearson相关系数理论,获得电力数据类型满足关联性阈值的影响因素;将缺失数据的影响因素输入训练好的LSTM神经网络,获得缺失数据的电力数据类型;根据缺失数据的电力数据类型采用不同的方法对数据进行修复。本发明考虑到电力数据的复杂非线性,利用神经网络的学习能力强,能够处理非线性问题的特性,实现电力数据的修复,可有效提升分类效率和准确率。



1. 一种电力数据修复方法,其特征在于:包括如下步骤:

步骤S1、获取历史电力数据集,利用SOM神经网络对历史电力数据集中电力数据进行分类处理,得到电力数据类型;

步骤S2、利用Pearson相关系数理论,对电力数据类型与影响因素进行相关性分析,获得电力数据类型满足关联性阈值的影响因素,将电力数据类型对应的影响因素作为特征值;

步骤S3、将电力数据类型与对应的特征值作为训练样本,对LSTM神经网络进行训练,获得训练好的LSTM神经网络;

步骤S4、将缺失数据的影响因素输入训练好的LSTM神经网络,获得缺失数据的电力数据类型;

步骤S5、根据缺失数据的电力数据类型采用不同的方法对数据进行修复。

2. 根据权利要求1所述的一种电力数据修复方法,其特征在于:所述SOM神经网络由输入层和输出层两层组成,输入层与输出层之间采用全连接;在对输入电力数据进行分类时,输入层的各神经元协同工作,分别竞争输入电力数据的响应机会,从而得到输出的神经元;通过对权值的更新,输出层的神经元四周的权值都会进行相应的调整,电力数据经过调整后输入层的每一个神经元对特定类别的学习,实现输出层对输入电力数据的学习分类。

3. 根据权利要求2所述的一种电力数据修复方法,其特征在于:所述权值取值范围为 $[0,1]$ 。

4. 根据权利要求1所述的一种电力数据修复方法,其特征在于:所述Pearson相关系数理论,计算公式如下:

$$\rho_{x,y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 (Y_i - \bar{Y})^2}}$$

式中: $\rho_{x,y}$ 为相关性系数,范围为 $[0,1]$;X,Y分别为连续变量;n为连续变量样本的个数; \bar{X} 和 \bar{Y} 分别为连续变量的平均值; $\rho_{x,y}=0$,表示变量间无关; $\rho_{x,y}>0$,表示变量间正相关; $\rho_{x,y}<0$,表示变量间负相关; $|\rho_{x,y}|$ 越大说明变量间关联性越强,反之越弱。

5. 根据权利要求1所述的一种电力数据修复方法,其特征在于:所述LSTM神经网络包括3个门结构实现对电力数据类型的预测,分别是遗忘门,输入门和输出门。

6. 根据权利要求5所述的一种电力数据修复方法,其特征在于:所述遗忘门计算公式如下:

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + B_f)$$

式中: f_t 为遗忘门的输出; W_{xf} 和 W_{hf} 为遗忘门需要学习的网络参数; B_f 为遗忘门的偏置; σ 为sigmoid函数:

$$\sigma(x) = \frac{1}{1 + \exp(-x)}$$

所述输入门计算公式如下:

$$i_t = \delta(W_{xi}x_t + W_{hi}h_{t-1} + B_i)$$

$$C_t' = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + B_c)$$

$$C_t = f_t \times C_{t-1} + i_t \times C_t'$$

式中： i_t 为输入门的输出； W_{xi} 和 W_{hi} 为输入门需要学习的网络参数； B_i 为输入门的偏置； C_t' 为tanh函数更新后的输入； W_{xc} 和 W_{hc} 为细胞状态需要学习的网络参数； B_c 为细胞状态的偏置； C_t 为细胞状态；

所述输出门计算公式如下：

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + B_o)$$

$$h_t = o_t \times \tanh(C_t)$$

式中： o_t 为输出门的输出； h_t 为输出的隐藏单元； W_{xo} 和 W_{ho} 为输出门的需要学习的网络参数； B_o 为输出门的偏置。

7. 根据权利要求1所述的一种电力数据修复方法,其特征在于:所述根据缺失数据的电力数据类型采用不同的方法对数据进行修复,包括如下步骤:

当缺失数据的电力数据类型为电能数据时,采用分层均值填补法,将电能数据按每年的相同月份进行分层,取每个层的平均值代替缺失值;

当缺失数据的电力数据类型为负荷数据时,采用最大期望算法,对缺失数据进行最大似然估计,并不断迭代,最终输出值代替缺失值;

当缺失数据的电力数据类型为电流数据、电压数据或频率数据时,采用均值填补法,对前后两个观测值的平均值或者是非缺失数据的均值来代替缺失值。

一种电力数据修复方法

技术领域

[0001] 本发明涉及一种电力数据修复方法,属于电力数据检测修复技术领域。

背景技术

[0002] 数字化技术不断发展进步,使得电力系统中产生了大量的电力数据。但是,电力数据常常因为外界干扰、传输错误、设备异常、网络延迟等原因而出现数据缺失的情况,这样会影响电力系统内处理数据的正确性与及时性。

[0003] 现有修复方法多采用传统机器学习方法,但是,面对目前复杂的日益结构复杂的电网结构,尤其是新能源发电系统与电动汽车的接入、需求响应机制的应用,导致传统机器学习方法无法应对高随机性的电力数据的修复。

[0004] 为了保证智能化电力系统的优化稳定运行,需要有完整正确的电力数据提供支持,避免影响电力系统安全稳定运行。本领域技术人员急需需要对电力系统的数据进行更高精度的修复。

发明内容

[0005] 目的:为了克服现有技术中存在的不足,本发明提供一种电力数据修复方法。

[0006] 技术方案:为解决上述技术问题,本发明采用的技术方案为:

[0007] 一种电力数据修复方法,包括如下步骤:

[0008] 步骤S1、获取历史电力数据集,利用SOM神经网络对历史电力数据集中电力数据进行分类处理,得到电力数据类型。

[0009] 步骤S2、利用Pearson相关系数理论,对电力数据类型与影响因素进行相关性分析,获得电力数据类型满足关联性阈值的影响因素,将电力数据类型对应的影响因素作为特征值。

[0010] 步骤S3、将电力数据类型与对应的特征值作为训练样本,对LSTM神经网络进行训练,获得训练好的LSTM神经网络。

[0011] 步骤S4、将缺失数据的影响因素输入训练好的LSTM神经网络,获得缺失数据的电力数据类型。

[0012] 步骤S5、根据缺失数据的电力数据类型采用不同的方法对数据进行修复。

[0013] 作为优选方案,所述SOM神经网络由输入层和输出层两层组成,输入层与输出层之间采用全连接。在对输入电力数据进行分类时,输入层的各神经元协同工作,分别竞争输入电力数据的响应机会,从而得到输出的神经元。通过对权值的更新,输出层的神经元四周的权值都会进行相应的调整,电力数据经过调整后输入层的每一个神经元对特定类别的学习,实现输出层对输入电力数据的学习分类。权值取值范围通常为[0,1]。

[0014] 作为优选方案,所述Pearson相关系数理论,计算公式如下:

$$[0015] \quad \rho_{x,y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 (Y_i - \bar{Y})^2}}$$

[0016] 式中： $\rho_{x,y}$ 为相关性系数，范围为 $[0, 1]$ ； X, Y 分别为连续变量； n 为连续变量样本的个数； \bar{X} 和 \bar{Y} 分别为连续变量的平均值； $\rho_{x,y}=0$ ，表示变量间无关； $\rho_{x,y}>0$ ，表示变量间正相关； $\rho_{x,y}<0$ ，表示变量间负相关； $|\rho_{x,y}|$ 越大说明变量间关联性越强，反之越弱。

[0017] 作为优选方案，所述LSTM神经网络构建如下所示：

[0018] LSTM神经网络包括3个门结构实现对电力数据类型的预测，分别是遗忘门，输入门和输出门。

[0019] (1) 遗忘门。主要用于对输入数据筛选，计算数据的保留程度，通过sigmoid神经层处理后输出一个 $[0, 1]$ 范围内的值，数值越大表示保留成分越多，反之越少。遗忘门计算公式如下：

$$[0020] \quad f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + B_f)$$

[0021] 式中： f_t 为遗忘门的输出； W_{xf} 和 W_{hf} 为遗忘门需要学习的网络参数； B_f 为遗忘门的偏置； σ 为sigmoid函数：

$$[0022] \quad \sigma(x) = \frac{1}{1 + \exp(-x)}$$

[0023] (2) 输入门。主要用于对数据状态的更新，主要包括两部分信息，一部分是sigmoid函数选取需要保存的数据；另一部分是tanh函数把当前输入 x_t 生成的新信息， C'_t 将两部分信息结合起来产生新的记忆状态，计算公式如下：

$$[0024] \quad i_t = \delta(W_{xi}x_t + W_{hi}h_{t-1} + B_i)$$

$$[0025] \quad C'_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + B_c)$$

$$[0026] \quad C_t = f_t \times C_{t-1} + i_t \times C'_t$$

[0027] 式中： i_t 为输入门的输出； W_{xi} 和 W_{hi} 为输入门需要学习的网络参数； B_i 为输入门的偏置； C'_t 为tanh函数更新后的输入； W_{xc} 和 W_{hc} 为细胞状态需要学习的网络参数； B_c 为细胞状态的偏置； C_t 为细胞状态。

[0028] (3) 输出门。主要用于计算输入信息 x_t 的输出程度。通过细胞状态中输出的信息经过tanh函数后与sigmoid函数做乘得到最终的隐藏单元 h_t ，计算公式如下：

$$[0029] \quad o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + B_o)$$

$$[0030] \quad h_t = o_t \times \tanh(C_t)$$

[0031] 式中： o_t 为输出门的输出； h_t 为输出的隐藏单元； W_{xo} 和 W_{ho} 为输出门的需要学习的网络参数； B_o 为输出门的偏置。

[0032] 作为优选方案，所述根据缺失数据的电力数据类型采用不同的方法对数据进行修复，包括如下步骤：

[0033] 当缺失数据的电力数据类型为电能数据时，采用分层均值填补法，将电能数据按每年的相同月份进行分层，取每个层的平均值代替缺失值。

[0034] 当缺失数据的电力数据类型为负荷数据时，采用最大期望算法，对缺失数据进行最大似然估计，并不断迭代，最终输出值代替缺失值。

[0035] 当缺失数据的电力数据类型为电流数据、电压数据或频率数据时，采用均值填补

法,对前后两个观测值的平均值或者是非缺失数据的均值来代替缺失值。

[0036] 有益效果:本发明提供了一种电力数据修复方法,本方法考虑到电力数据的复杂非线性,利用神经网络的学习能力强,能够处理非线性问题的特性,实现电力数据的修复,可有效提升分类效率和准确率。

附图说明

[0037] 图1是本发明的方法流程示意图。

[0038] 图2是本发明一实施例提供的分类模型SOM-LSTM神经网络的性能准确率对比图。

[0039] 图3是本发明另一实施例提供的分类模型SOM-LSTM神经网络的性能准确率对比图。

具体实施方式

[0040] 下面结合具体实施例对本发明作更进一步的说明。

[0041] 以下结合附图和具体实施方式对本发明提出的一种电力数据修复方法作进一步详细说明。根据下面说明,本发明的优点和特征将更清楚。需要说明的是,附图采用非常简化的形式且均使用非精准的比例,仅用以方便、明晰地辅助说明本发明实施方式的目的。为了使本发明的目的、特征和优点能够更加明显易懂,请参阅附图。须知,本说明书所附图式所绘示的结构、比例、大小等,均仅用以配合说明书所揭示的内容,以供熟悉此技术的人士了解与阅读,并非用以限定本发明实施的限定条件,故不具技术上的实质意义,任何结构的修饰、比例关系的改变或大小的调整,在不影响本发明所能产生的功效及所能达成的目的下,均应仍落在本发明所揭示的技术内容能涵盖的范围内。

[0042] 需要说明的是,在本文中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他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、物品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他要素,或者是还包括为这种过程、方法、物品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、物品或者设备中还存在另外的相同要素。

[0043] 结合附图1所示,本实施例提供了一种电力数据修复方法,包括如下步骤:

[0044] 步骤S1、获取历史电力数据集,利用SOM神经网络对历史电力数据集中电力数据进行分类处理,得到电力数据类型。

[0045] 步骤S2、利用Pearson相关系数理论,对电力数据类型与影响因素进行相关性分析,获得电力数据类型满足关联性阈值的影响因素,将电力数据类型对应的影响因素作为特征值。

[0046] 步骤S3、将电力数据类型与对应的特征值作为训练样本,对LSTM神经网络进行训练,获得训练好的LSTM神经网络。

[0047] 步骤S4、将缺失数据的影响因素输入训练好的LSTM神经网络,获得缺失数据的电力数据类型。

[0048] 步骤S5、根据缺失数据的电力数据类型采用不同的方法对数据进行修复。

[0049] 所述SOM神经网络由输入层和输出层两层组成,输入层与输出层之间采用全连接。在对输入电力数据进行分类时,输入层的各神经元协同工作,分别竞争输入电力数据的响应机会,从而得到输出的神经元。通过对权值的更新,输出层的神经元四周的权值都会进行相应的调整,电力数据经过调整后输入层的每一个神经元对特定类别的学习,实现输出层对输入电力数据的学习分类。权值取值范围通常为[0,1]。

[0050] 所述Pearson相关系数理论,计算公式如下:

$$[0051] \quad \rho_{X,Y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 (Y_i - \bar{Y})^2}}$$

[0052] 式中: $\rho_{X,Y}$ 为相关性系数,范围为[0,1]; X,Y 分别为连续变量; n 为连续变量样本的个数; \bar{X} 和 \bar{Y} 分别为连续变量的平均值; $\rho_{X,Y}=0$,表示变量间无关; $\rho_{X,Y}>0$,表示变量间正相关; $\rho_{X,Y}<0$,表示变量间负相关; $|\rho_{X,Y}|$ 越大说明变量间关联性越强,反之越弱。

[0053] 所述LSTM神经网络构建如下所示:

[0054] LSTM神经网络包括3个门结构实现对电力数据类型的预测,分别是遗忘门,输入门和输出门。

[0055] (1) 遗忘门。主要用于对输入数据筛选,计算数据的保留程度,通过sigmoid神经层处理后输出一个[0,1]范围内的值,数值越大表示保留成分越多,反之越少。遗忘门计算公式如下:

$$[0056] \quad f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + B_f)$$

[0057] 式中: f_t 为遗忘门的输出; W_{xf} 和 W_{hf} 为遗忘门需要学习的网络参数; B_f 为遗忘门的偏置; σ 为sigmoid函数:

$$[0058] \quad \sigma(x) = \frac{1}{1 + \exp(-x)}$$

[0059] (2) 输入门。主要用于对数据状态的更新,主要包括两部分信息,一部分是sigmoid函数选取需要保存的数据;另一部分是tanh函数把当前输入 x_t 生成的新信息, C'_t 将两部分信息结合起来产生新的记忆状态,计算公式如下:

$$[0060] \quad i_t = \delta(W_{xi}x_t + W_{hi}h_{t-1} + B_i)$$

$$[0061] \quad C'_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + B_c)$$

$$[0062] \quad C_t = f_t \times C_{t-1} + i_t \times C'_t$$

[0063] 式中: i_t 为输入门的输出; W_{xi} 和 W_{hi} 为输入门需要学习的网络参数; B_i 为输入门的偏置; C'_t 为tanh函数更新后的输入; W_{xc} 和 W_{hc} 为细胞状态需要学习的网络参数; B_c 为细胞状态的偏置; C_t 为细胞状态。

[0064] (3) 输出门。主要用于计算输入信息 x_t 的输出程度。通过细胞状态中输出的信息经过tanh函数后与sigmoid函数做乘得到最终的隐藏单元 h_t ,计算公式如下:

$$[0065] \quad o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + B_o)$$

$$[0066] \quad h_t = o_t \times \tanh(C_t)$$

[0067] 式中: o_t 为输出门的输出; h_t 为输出的隐藏单元; W_{xo} 和 W_{ho} 为输出门的需要学习的网络参数; B_o 为输出门的偏置。

[0068] 所述根据缺失数据的电力数据类型采用不同的方法对数据进行修复,包括如下步

骤:

[0069] 当缺失数据的电力数据类型为电能数据时,采用分层均值填补法,将电能数据按每年的相同月份进行分层,取每个层的平均值代替缺失值。

[0070] 当缺失数据的电力数据类型为负荷数据时,采用最大期望算法,对缺失数据进行最大似然估计,并不断迭代,最终输出值代替缺失值。

[0071] 当缺失数据的电力数据类型为电流数据、电压数据或频率数据时,采用均值填补法,对前后两个观测值的平均值或者是非缺失数据的均值来代替缺失值。

[0072] 如图2,3所示,为了验证所提的基于SOM-LSMT电力数据修复方法的有效性,本文实验仿真的计算机配置为Inter Intel (R)Core (TM) i5-8300H CPU@2.30GHz2.30GHz和NVIDIA GTX 1050Ti 4G显存,在MATLAB R2019a平台上进行仿真计算。

[0073] 本文选取2018年8月中国某地区的家庭电力数据以及相应气象站所采集到的温度数据,数据采集间隔为15分钟。选用平均绝对误差((mean absolute error,MAE)和均方根误差(root mean square error, RMSE)评价指标对电力数据修复结果进行分析。

[0074] 为了对所提出的基于SOM-LSTM模型的准确性进行分析,本文分别对典型工作日和非工作日的连续缺失电力数据SOM-LSTM模型与极限学习机(Extreme Learning Machine, ELM)模型、LSTM模型的修复结果绘制曲线进行对比,结果如图2和图3所示。

[0075] 由图2和图3可知,针对典型工作日和非工作日的用户电压数据连续缺失进行修复时,LSTM神经网络的误差最大,且随着数据缺失的增加,精确度显著降低,说明该神经网络的拟合修复效果不足。对比LSTM模型和ELM模型,经过SOM神经网络分类后的数据进行预测修复的SOM-LSTM神经网络能够有效改善连续缺失数据修复准确率低的问题,并且在数据缺失较多的情况下依然保证较高水平的修复率。

[0076] 尽管本发明的内容已经通过上述优选实施例作了详细介绍,但应当认识到上述的描述不应被认为是对本发明的限制。在本领域技术人员阅读了上述内容后,对于本发明的多种修改和替代都将是显而易见的。因此,本发明的保护范围应由所附的权利要求来限定。

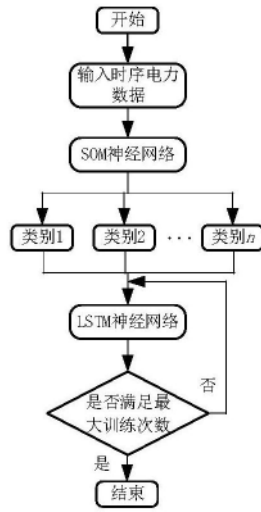


图1

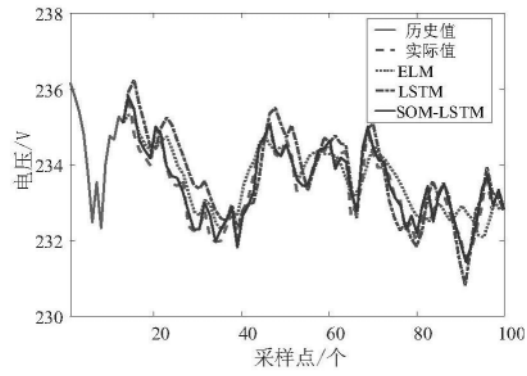


图2

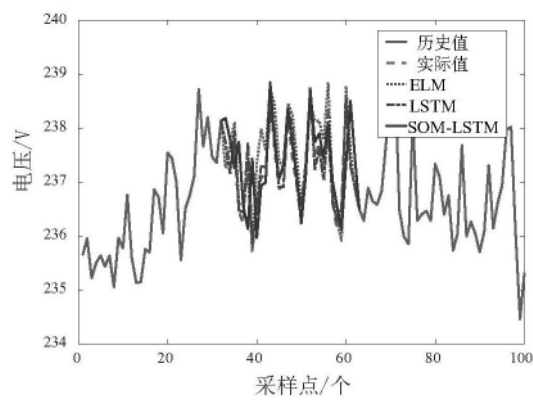


图3