



(12) 发明专利

(10) 授权公告号 CN 110347666 B

(45) 授权公告日 2023. 03. 31

(21) 申请号 201910571954.0

(22) 申请日 2019.06.28

(65) 同一申请的已公布的文献号
申请公布号 CN 110347666 A

(43) 申请公布日 2019.10.18

(73) 专利权人 佛山科学技术学院
地址 528000 广东省佛山市南海区狮山镇
广云路33号

(72) 发明人 张彩霞 王向东 曾平

(74) 专利代理机构 广州嘉权专利商标事务所有
限公司 44205
专利代理师 蔡伟杰

(51) Int. Cl.
G06F 16/215 (2019.01)
G06F 16/2458 (2019.01)

(56) 对比文件

- CN 109469896 A, 2019.03.15
- CN 106895946 A, 2017.06.27
- CN 102141403 A, 2011.08.03
- CN 108167653 A, 2018.06.15
- CN 104574848 A, 2015.04.29
- US 2014379904 A1, 2014.12.25

审查员 张锐峰

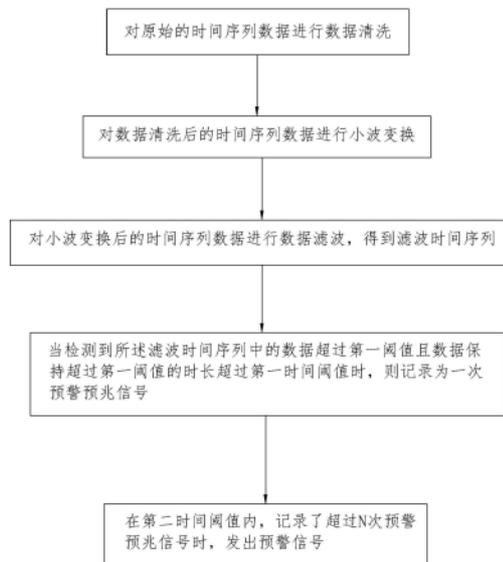
权利要求书1页 说明书4页 附图1页

(54) 发明名称

一种改善时序数据质量和预警的方法及装置

(57) 摘要

本发明公开了一种改善时序数据质量和预警的方法及装置,包括:对原始的时间序列数据进行数据清洗;对数据清洗后的时间序列数据进行小波变换;对小波变换后的时间序列数据进行数据滤波,得到滤波时间序列;当检测所述滤波时间序列中的数据超过第一阈值且数据保持超过第一阈值的时长超过第一时间阈值时,则记录为一次预警预兆信号;在第二时间阈值内,记录了超过N次预警预兆信号时,发出预警信号。本发明对时间序列数据进行去噪,同时降低去噪过程中失真情况,提高时间序列数据的质量,通过设置阈值判断预警预兆信号,根据检测情况发出预警信号。



1. 一种改善时序数据质量和预警的方法,其特征在于,包括:
 对原始的时间序列数据进行数据清洗;
 对数据清洗后的时间序列数据进行小波变换;
 对小波变换后的时间序列数据进行数据滤波,得到滤波时间序列;
 当检测到所述滤波时间序列中的数据超过第一阈值且数据保持超过第一阈值的时长超过第一时间阈值时,则记录为一次预警预兆信号;
 在第二时间阈值内,记录了超过N次预警预兆信号时,发出预警信号;
 使用Mallat算法进行小波变换,当尺度参数较大时,频率分辨率较高,适合分析低频信号,当尺度参数较小时,时间分辨率较高,适合分析高频信号;
 将数据清洗后的时间序列数据定义为 $f(x)$,小波变换的函数表达式为:

$$W_f(a, b) = \int_R f(x) \bar{\psi}_{(a,b)}(x) dx = \frac{1}{\sqrt{a}} \int_R f(x) \bar{\psi}\left(\frac{x-b}{a}\right) dx;$$

其中, a 为尺度参数, b 为时间中心参数, ψ 为小波母函数。

2. 根据权利要求1所述的一种改善时序数据质量和预警的方法,其特征在于,所述对原始的时间序列数据进行数据清洗的过程包括:

对原始的时间序列数据中的脏数据进行去除,所述脏数据包括缺失值和异常值。

3. 根据权利要求2所述的一种改善时序数据质量和预警的方法,其特征在于,所述对原始的时间序列数据中的脏数据进行去除的过程包括:对所述缺失值进行均值填补;

利用单变量的散点图对所述异常值进行初步判断,再利用统计学 3σ 法原则进行述异常值的排查,先采用零值替换法剔除异常值,然后利用均值填补。

4. 根据权利要求1所述的一种改善时序数据质量和预警的方法,其特征在于,对小波变换后的时间序列数据进行数据滤波的过程包括:

对小波变换后的时间序列数据进行卡尔曼滤波。

5. 一种改善时序数据质量和预警的装置,其特征在于,包括:

数据采集模块,用于监测和采集时间序列数据,并将采集得到的时间序列数据发送到处理模块;

处理模块,用于对数据采集模块采集得到的时间序列数据进行处理,检测预警预兆信号,控制预警模块发出预警信号;

预警模块,当数据处理模块检测在第二时间阈值内记录了超过N次预警预兆信号时,控制预警模块发出预警信号;

所述处理模块分别与数据采集模块和预警模块相连接;

使用Mallat算法进行小波变换,当尺度参数较大时,频率分辨率较高,适合分析低频信号,当尺度参数较小时,时间分辨率较高,适合分析高频信号;

将数据清洗后的时间序列数据定义为 $f(x)$,小波变换的函数表达式为:

$$W_f(a, b) = \int_R f(x) \bar{\psi}_{(a,b)}(x) dx = \frac{1}{\sqrt{a}} \int_R f(x) \bar{\psi}\left(\frac{x-b}{a}\right) dx;$$

其中, a 为尺度参数, b 为时间中心参数, ψ 为小波母函数。

一种改善时序数据质量和预警的方法及装置

技术领域

[0001] 本发明涉及数据处理技术领域,更具体地说涉及一种改善时序数据质量和预警的方法及装置。

背景技术

[0002] 近几年随着科技的发展,出现了越来越多的以时间和空间为维度的数据。例如在生物医学方面,基因数据就是以空间为维度的数据,针对基因数据的异常检测可以发现疾病;在金融方面,持卡人的消费记录就是以时间为维度的数据,通过检测持卡人的消费记录所对应的时间序列数据,可以找到异常持卡人;在工业方面,温度传感器、压力传感器等的的数据都是以时间和空间为维度的数据。

[0003] 这些时间序列数据中蕴含了大量的信息,但时间序列数据中存在许多噪声,对时间序列数据进行数据挖掘之前需要对时间序列数据进行预处理,提高时间序列数据的质量,不然时间序列数据中的噪音严重影响了进一步地分析和处理,同时对数据进行判断是不准确的。但现有技术中时间序列数据预处理方法通常是不彻底的,这会影响后续的数据挖掘效果。

发明内容

[0004] 本发明提供一种改善时序数据质量和预警的方法及装置,提高时间序列数据的质量和发出预警。

[0005] 本发明解决其技术问题的解决方案是:

[0006] 一种改善时序数据质量和预警的方法,其特征在于,包括:

[0007] 对原始的时间序列数据进行数据清洗;

[0008] 对数据清洗后的时间序列数据进行小波变换;

[0009] 对小波变换后的时间序列数据进行数据滤波,得到滤波时间序列;

[0010] 当检测到所述滤波时间序列中的数据超过第一阈值且数据保持超过第一阈值的时长超过第一时间阈值时,则记录为一次预警预兆信号;

[0011] 在第二时间阈值内,记录了超过N次预警预兆信号时,发出预警信号。

[0012] 作为上述技术方案的进一步改进,所述对原始的时间序列数据进行数据清洗的过程包括:

[0013] 对原始的时间序列数据中的脏数据进行去除,所述脏数据包括缺失值和异常值。

[0014] 作为上述技术方案的进一步改进,所述对原始的时间序列数据中的脏数据进行去除的过程包括:

[0015] 对所述缺失值进行均值填补;

[0016] 利用单变量的散点图对所述异常值进行初步判断,再利用统计学 3σ 法原则进行述异常值的排查,先采用零值替换法剔除异常值,然后利用均值填补。

[0017] 作为上述技术方案的进一步改进,对小波变换后的时间序列数据进行数据滤波的

过程包括：

- [0018] 对小波变换后的时间序列数据进行卡尔曼滤波。
- [0019] 一种改善时序数据质量和预警的装置,包括:
- [0020] 数据采集模块,用于监测和采集时间序列数据,并将采集得到的时间序列数据发送到处理模块;
- [0021] 处理模块,用于对数据采集模块采集得到的时间序列数据进行处理,检测预警预兆信号,控制预警模块发出预警信号;
- [0022] 预警模块,当数据处理模块检测在第二时间阈值内记录了超过N次预警预兆信号时,控制预警模块发出预警信号;
- [0023] 所述处理模块分别与数据采集模块和预警模块相连接。
- [0024] 本发明的有益效果是:本发明对时间序列数据进行去噪,同时降低去噪过程中失真情况,提高时间序列数据的质量,通过设置阈值判断预警预兆信号,根据检测情况发出预警信号。

附图说明

[0025] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单说明。显然,所描述的附图只是本发明的一部分实施例,而不是全部实施例,本领域的技术人员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他设计方案和附图。

[0026] 图1是本发明方法的流程示意图。

具体实施方式

[0027] 以下将结合实施例和附图对本发明的构思、具体结构及产生的技术效果进行清楚、完整的描述,以充分地理解本发明的目的、特征和效果。显然,所描述的实施例只是本发明的一部分实施例,而不是全部实施例,基于本发明的实施例,本领域的技术人员在不付出创造性劳动的前提下所获得的其他实施例,均属于本发明保护的范围。另外,文中所提到的所有连接关系,并非单指构件直接相接,而是指可根据具体实施情况,通过添加或减少连接辅件,来组成更优的连接结构。本发明创造中的各个技术特征,在不互相矛盾冲突的前提下可以交互组合。

- [0028] 实施例1,参照图1,一种改善时序数据质量和预警的方法,包括:
- [0029] 对原始的时间序列数据进行数据清洗;
- [0030] 对数据清洗后的时间序列数据进行小波变换;
- [0031] 对小波变换后的时间序列数据进行数据滤波,得到滤波时间序列;
- [0032] 当检测到所述滤波时间序列中的数据超过第一阈值且数据保持超过第一阈值的时长超过第一时间阈值时,则记录为一次预警预兆信号;
- [0033] 在第二时间阈值内,记录了超过N次预警预兆信号时,发出预警信号。
- [0034] 本实施例使用Mallat算法进行小波变换,当尺度参数较大时,频率分辨率较高,适合分析低频信号,当尺度参数较小时,时间分辨率较高,适合分析高频信号。
- [0035] 将数据清洗后的时间序列数据定义为 $f(x)$,小波变换的函数表达式为:

$$W_f(a, b) = \int_R f(x) \bar{\psi}_{(a,b)}(x) dx = \frac{1}{\sqrt{a}} \int_R f(x) \bar{\psi}\left(\frac{x-b}{a}\right) dx;$$

[0036] 其中, a为尺度参数, b为时间中心参数, ψ 为小波母函数。

[0037] 现有技术中通过傅里叶变换对数据进行处理, 去掉数据中的高频信息, 但是通过利用傅立叶变换对数据进行处理, 得到的结果却不尽人意, 因为存在比较严重的失真情况, 而本实施例利用小波变换中的多分辨特点来对数据进行分解, 通过多分辨信号算法来实现对数据的逐层分解, 最后得到数据清洗后的时间序列数据的低频信息, 起到了消除噪音的作用, 不会出现严重的失真情况, 保持数据的准确性。

[0038] 进一步作为优选的实施方式, 所述对原始的时间序列数据进行数据清洗的过程包括:

[0039] 对原始的时间序列数据中的脏数据进行去除, 所述脏数据包括缺失值和异常值。

[0040] 进一步作为优选的实施方式, 所述对原始的时间序列数据中的脏数据进行去除的过程包括:

[0041] 对所述缺失值进行均值填补;

[0042] 利用单变量的散点图对所述异常值进行初步判断, 再利用统计学 3σ 法原则进行所述异常值的排查, 得到异常值;

[0043] 对于所述异常值, 先采用零值替换法剔除异常值, 然后利用均值填补。

[0044] 当时间序列数据中出现特殊字符的数据时, 先采用零值替换法剔除异常值, 然后利用均值填补。

[0045] 所述均值为所述原始的时间序列数据的均值。

[0046] 进一步作为优选的实施方式, 对小波变换后的时间序列数据进行数据滤波的过程包括:

[0047] 对小波变换后的时间序列数据进行卡尔曼滤波。

[0048] 所述一种改善时序数据质量的方法可应用在所述一种改善时序数据质量和预警的装置, 所述装置包括:

[0049] 数据采集模块, 用于监测和采集时间序列数据, 并将采集得到的时间序列数据发送到处理模块;

[0050] 处理模块, 用于对数据采集模块采集得到的时间序列数据进行处理, 检测预警预兆信号, 控制预警模块发出预警信号;

[0051] 预警模块, 当数据处理模块检测在第二时间阈值内记录了超过N次预警预兆信号时, 控制预警模块发出预警信号;

[0052] 所述处理模块分别与数据采集模块和预警模块相连接。

[0053] 本实施例的工作过程:

[0054] 在工业方面, 温度传感器、压力传感器等的的数据都是以时间和空间为维度的数据。本实施例以检测工业上检测温度为例, 所述数据采集模块为温度传感器。

[0055] 所述数据采集模块实时采集锅炉的温度值, 并将采集得到的时间序列数据发送到所述处理模块, 所述处理模块对所述时间序列数据进行数据清洗, 对数据清洗后的时间序列数据进行小波变换, 得到低频数据, 消除噪音。对小波变换后的时间序列数据进行卡尔曼滤波, 得到滤波时间序列。

[0056] 当处理模块检测到到所述滤波时间序列中数据超过第一阈值,且保持超过第一阈值的时长超过第一时间阈值时,则记录为一次预警预兆信号。

[0057] 本实施例的数据采集模块用于检测低压锅炉的蒸汽温度,低压锅炉的蒸汽温度需要在400℃以下,所述第一阈值为380℃,所述第一时间阈值为15s,所述第二时间阈值为12h。低压锅炉的蒸汽温度超过380℃,且超过380℃的时长超过15s,则记录为一次预警预兆信号。

[0058] 本实施例中N为5,在12小时内,处理模块记录了超过5次预警预兆信号时,则处理模块控制所述预警模块发出预警信号。

[0059] 低压锅炉的蒸汽温度如果长期超过400℃,引起锅炉金属部件的蠕变速度加快,许用应力大大降低,更严重的会导致管道过热爆破。低压锅炉瞬时的蒸汽温度上升是很正常的,当蒸汽温度上升维持一段时间,且在固定时间内蒸汽温度上升过于频繁,则说明低压锅炉内部有出现故障的可能。但现有技术中无法提前发现锅炉的故障,本实例检测低压锅炉的蒸汽温度超过380℃以及每次超过380℃的时长,当12小时内,记录了超过5次预警预兆信号时,则低压锅炉的蒸汽温度上升的频率较大,可能低压锅炉内部出现了故障,提醒工作人员对各个部件进行检查,及时发现问题。

[0060] 本发明对时间序列数据进行去噪,同时降低去噪过程中失真情况,提高时间序列数据的质量,通过设置第一阈值、第一时间阈值和第二时间阈值判断预警预兆信号,根据检测情况发出预警信号。这样有利于通过处理时间序列数据及时发现潜在的问题,预警和定位故障。

[0061] 以上对本发明的较佳实施方式进行了具体说明,但本发明创造并不限于所述实施例,熟悉本领域的技术人员在不违背本发明精神的前提下还可作出种种的等同变型或替换,这些等同的变型或替换均包含在本申请权利要求所限定的范围内。

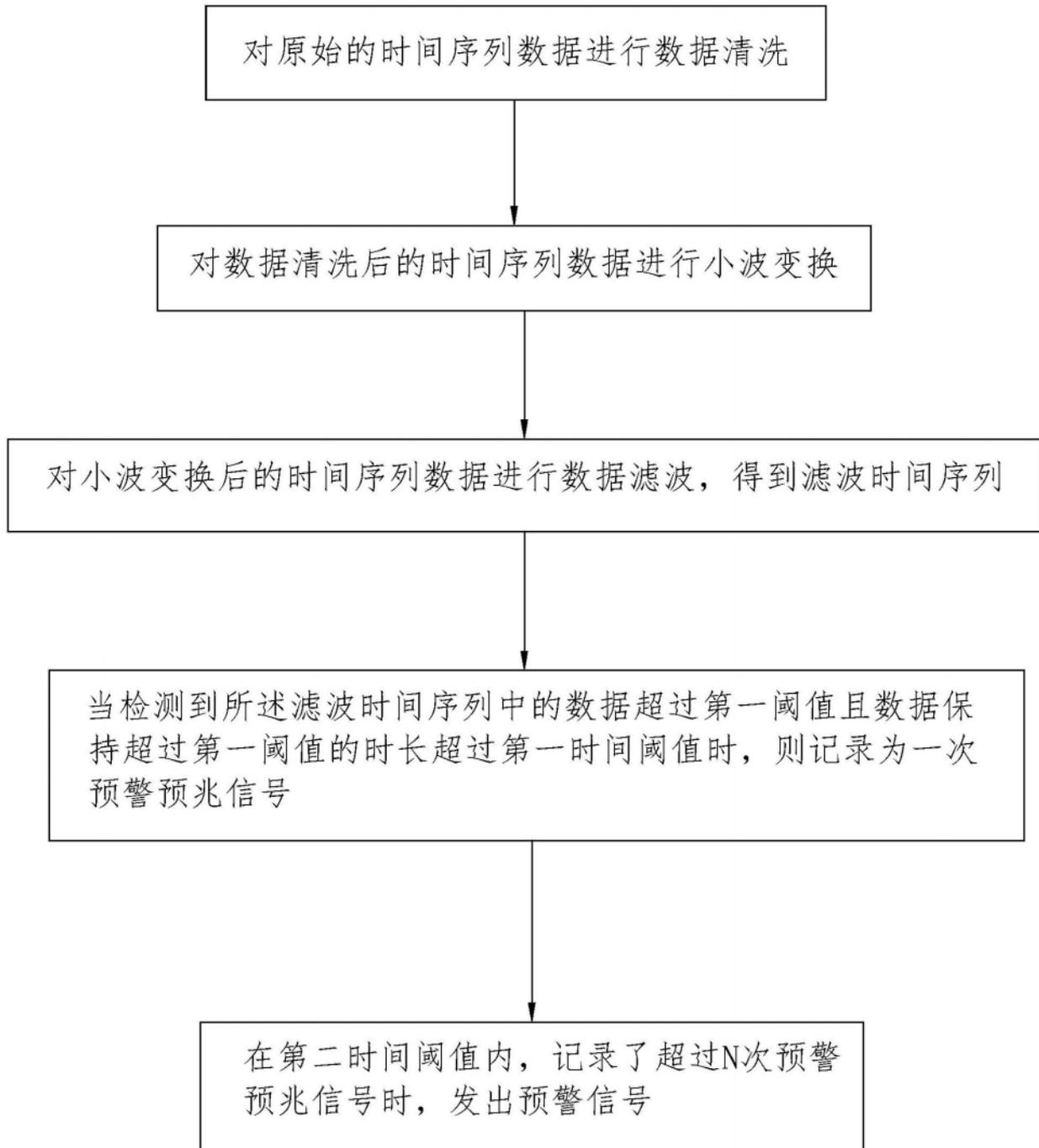


图1