



(12)发明专利

(10)授权公告号 CN 107346301 B

(45)授权公告日 2020.09.04

(21)申请号 201710523755.3

(22)申请日 2017.06.30

(65)同一申请的已公布的文献号
申请公布号 CN 107346301 A

(43)申请公布日 2017.11.14

(66)本国优先权数据
201611092983.1 2016.12.02 CN

(73)专利权人 西交利物浦大学
地址 215123 江苏省苏州市工业园区独墅湖科教创新区仁爱路111号

(72)发明人 岳勇 朱晓辉 张晋 张一新
王威

(74)专利代理机构 南京钟山专利代理有限公司
32252

代理人 戴朝荣

(51)Int.Cl.

G06F 17/18(2006.01)

(56)对比文件

CN 102945320 A,2013.02.27

CN 106302487 A,2017.01.04

余宇峰 等.基于滑动窗口预测的水文时间序列异常检测.《计算机应用》.2014,第2217-2220、2226页.

何明亮 等.基于多窗口机制的聚类异常检测算法”.《信息安全》.2016,(第11期),第33-39页.

裴丽娟.一种基于滑动窗口的时间序列异常检测算法.《巢湖学院学报》.2011,第13卷(第3期),第28-31页.

审查员 郭瑞

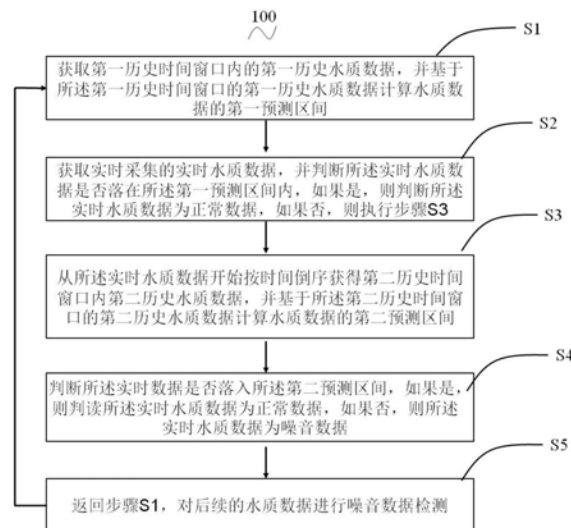
权利要求书2页 说明书5页 附图1页

(54)发明名称

基于双时间窗口验证的水质监测噪音数据实时检测方法

(57)摘要

本发明提供了一种基于双时间窗口验证的水质监测噪音数据实时检测方法,并包括如下步骤:一、获取第一历史时间窗口内的第一历史水质数据,并基于第一历史时间窗口的第一历史水质数据计算水质数据的第一预测区间;二、获取实时水质数据,并判断实时水质数据是否落在第一预测区间内,如果是,则判断实时水质数据为正常数据,如果不是,则执行步骤三;三、从实时水质数据开始按时间倒序获得第二历史时间窗口内第二历史水质数据,并基于第二历史时间窗口的第二历史水质数据计算水质数据的第二预测区间;四、判断实时数据是否落入所述第二预测区间,如果是,则判断实时水质数据为正常数据,如果不是,则判断实时水质数据为噪音数据。



1. 一种基于双时间窗口验证的水质监测噪音数据实时检测方法, 其特征在于: 包括如下步骤:

一、获取第一历史时间窗口内的第一历史水质数据, 并基于所述第一历史时间窗口的第一历史水质数据计算水质数据的第一预测区间;

二、获取实时采集的实时水质数据, 并判断所述实时水质数据是否落在所述第一预测区间内, 如果是, 则判断所述实时水质数据为正常数据, 如果否, 则执行步骤三;

三、从所述实时水质数据开始按时间倒序获得第二历史时间窗口内第二历史水质数据, 并基于所述第二历史时间窗口的第二历史水质数据计算水质数据的第二预测区间;

四、判断所述实时水质数据是否落入所述第二预测区间, 如果是, 则判读所述实时水质数据为正常数据, 如果否, 则所述实时水质数据为噪音数据;

所述步骤一包括如下步骤:

选择一个连续时间段内的历史水质数据作为模型训练数据, 并均分为Y组, 其中, Y为正整数;

计算每一组所述模型训练数据的残差, 并计算Y组所述历史水质数据的残差平均值;

通过交叉交叉法计算所有所述模型训练数据的残差标准差;

利用delta学习算法确定所述第一历史时间窗口内每个历史水质数据的初始权重系数, 并根据所述模型训练数据得到的残差标准差计算出待采集的下一个水质数据的第一预测值;

基于学生分布函数和所述第一预测值计算出待采集的下一个水质数据的第一预测区间。

2. 根据权利要求1所述的基于双时间窗口验证的水质监测噪音数据实时检测方法, 其特征在于: 所述步骤三包括如下步骤:

利用delta学习算法确定所述第二历史时间窗口内每个数据的初始权重系数, 并根据步骤一得到的残值标准差计算出待采集的下一个水质数据的第二预测值;

基于学生分布函数和所述第二预测值计算出所述待采集的下一个水质数据的第二预测区间。

3. 根据权利要求1所述的基于双时间窗口验证的水质监测噪音数据实时检测方法, 其特征在于: 在步骤二和步骤四中, 如果所述实时水质数据为正常数据, 则按时间序列删除时间窗口中的第一个水质数据, 并补入所述实时水质数据, 构造新的时间窗口。

4. 根据权利要求1所述的基于双时间窗口验证的水质监测噪音数据实时检测方法, 其特征在于: 在步骤四中, 如果所述实时水质数据为噪音数据, 则按时间序列删除时间窗口中的第一个水质数据, 并加入基于所述时间窗口计算得到的水质数据的预测值, 构造新的时间窗口。

5. 根据权利要求3或4所述的基于双时间窗口验证的水质监测噪音数据实时检测方法, 其特征在于: 所述时间窗口内的数据按先进先出的队列模式并根据时间序列进行动态调整。

6. 根据权利要求1-4任一所述的基于双时间窗口验证的水质监测噪音数据实时检测方法, 其特征在于: 所述第一历史时间窗口和所述第二历史时间窗口内的数据初始权重系数随时间窗口的变化而进行动态计算。

7. 根据权利要求1所述的基于双时间窗口验证的水质监测噪音数据实时检测方法, 其特征在于: 利用第二个历史时间窗口对可能是噪音数据进行二次验证。

8. 根据权利要求1-4任一所述的基于双时间窗口验证的水质监测噪音数据实时检测方法, 其特征在于: 在步骤四之后, 还包括如下步骤: 返回步骤一, 对后续的水质数据进行噪音数据检测。

基于双时间窗口验证的水质监测噪音数据实时检测方法

技术领域

[0001] 本发明属于水质监测技术领域,具体地涉及一种基于双时间窗口验证的水质监测噪音数据实时检测方法。

背景技术

[0002] 随和人类经济发展,一方面我们需要越来越多的清洁淡水资源,另一方面我们的工业生产和人类生活对淡水河流造成越来越多的污染,严重影响了河流生态系统和人类可持续发展。通过传感器对河流水质进行实时监测是对水资源进行保护的一个有效手段,得到了越来越多的应用。受野外恶劣自然环境、传感器可靠性、数据传输稳定性等多方面影响,传感器实时采集的水质监测数据中带有许多噪音数据。由于这类数据不能反映水质变化的真实情况,其对水质监测数据的可靠性造成很大影响,不但会引起监测系统的误报警,也会对后期的水质数据处理和分析带来误差。因此需要研究新的算法来实时检测并剔除这类噪音数据,提升水质实时监测系统的稳定性和可靠性。

[0003] 目前异常数据检测方法较多,诸如:回归分析、灰色系统、人工神经网络、概率模型等。上述方法都基于水质数据完全没有背景噪音数据干扰的假设下,对水质数据进行分析 and 预测,这与实际情况不符。现有的时间窗口法综合考虑时间窗口内所有水质数据的权重系数来预测下一个水质参数的上下限区间,当实际值超过预测区间时,即认为该值为噪音数据。这种方式实现简单,但只考虑了当前数据与时间窗口内历史数据关联性,却没有考虑该当前数据与后续数据的关联性,因此误检率较高。

发明内容

[0004] 本发明的目的在于提供一种通过第二个时间窗口对可能的噪音数据进行二次验证,从而有效降低噪音数据误检率的基于双时间窗口验证的水质监测噪音数据实时检测方法。

[0005] 本发明的技术方案如下:一种基于双时间窗口验证的水质监测噪音数据实时检测方法包括如下步骤:

[0006] 一、获取第一历史时间窗口内的第一历史水质数据,并基于所述第一历史时间窗口的第一历史水质数据计算水质数据的第一预测区间;

[0007] 二、获取实时采集的实时水质数据,并判断所述实时水质数据是否落在所述第一预测区间内,如果是,则判断所述实时水质数据为正常数据,如果不是,则执行步骤三;

[0008] 三、从所述实时水质数据开始按时间倒序获得第二历史时间窗口内第二历史水质数据,并基于所述第二历史时间窗口的第二历史水质数据计算水质数据的第二预测区间;

[0009] 四、判断所述实时数据是否落入所述第二预测区间,如果是,则判读所述实时水质数据为正常数据,如果不是,则所述实时水质数据为噪音数据。

[0010] 优选地,所述步骤一包括如下步骤:选择一个连续时间段内的历史水质数据作为模型训练数据,并均分为Y组,其中,Y为正整数;计算每一组所述模型训练数据的残差,并计

算Y组所述模型训练数据的残差平均值;通过十字交叉法计算所有所述模型训练数据的残差标准差;利用delta学习算法确定所述第一历史时间窗口内每个历史水质数据的初始权重系数,并根据模型训练数据得到的残差标准差计算出待采集的下一个水质数据的第一预测值;基于学生分布函数和所述第一预测值计算出待采集的下一个水质数据的第一预测区间。

[0011] 优选地,所述步骤三包括如下步骤:利用delta学习算法确定所述第二历史时间窗口内每个数据的初始权重系数,并根据步骤一得到的残值标准差计算出待采集的下一个水质数据的第二预测值;基于学生分布函数和所述第二预测值计算出所述待采集的下一个水质数据的第二预测区间。

[0012] 优选地,在步骤二和步骤四中,如果所述实时数据为正常数据,则按时间序列删除时间窗口中的第一个水质数据,并补入所述实时水质数据,构造新的时间窗口。

[0013] 优选地,在步骤四中,如果所述实时数据为噪音数据,则按时间序列删除时间窗口中的第一个水质数据,并加入基于所述时间窗口计算得到的水质数据的预测值,构造新的时间窗口。

[0014] 优选地,所述时间窗口内的数据按先进先出的队列模式并根据时间序列进行动态调整。

[0015] 优选地,所述第一历史时间窗口和所述第二历史时间窗口内的数据初始权重系数随时间窗口的变化而进行动态计算。

[0016] 优选地,利用第一个历史时间窗口来判断实际采集到的实时水质数据是否有可能是噪音数据。

[0017] 优选地,利用第二个历史时间窗口对可能的噪音数据进行二次验证。

[0018] 优选地,在步骤四之后,还包括如下步骤:返回步骤一,对后续的水质数据进行噪音数据检测。

[0019] 本发明提供的技术方案具有如下有益效果:

[0020] 所述基于双时间窗口验证的水质监测噪音数据实时检测方法采用双时间窗口对可能的噪音数据进行两次验证,从而有效降低了噪音数据误检率,进一步减少了系统因噪音数据引起的误报警,提升水质监测系统可靠性和稳定性。。

附图说明

[0021] 图1是本发明基于双时间窗口验证的水质监测噪音数据实时检测算法检测流程图。

具体实施方式

[0022] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。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。

[0023] 除非上下文另有特定清楚的描述,本发明中的元件和组件,数量既可以单个的形式存在,也可以多个的形式存在,本发明并不对此进行限定。本发明中的步骤虽然用标号进行了排列,但并不用于限定步骤的先后次序,除非明确说明了步骤的次序或者某步骤的执

行需要其他步骤作为基础,否则步骤的相对次序是可以调整的。可以理解,本文中所使用的术语“和/或”涉及且涵盖相关联的所列项目中的一者或一者以上的任何和所有可能的组合。

[0024] 请参阅图1,是本发明实施例提供的基于双时间窗口验证的水质监测噪音数据实时检测方法100包括如下步骤:

[0025] S1、获取第一历史时间窗口内的第一历史水质数据,并基于所述第一历史时间窗口的第一历史水质数据计算下一个水质数据的第一预测区间。

[0026] 在水质数据采集过程中,水质监测节点把采集的水质数据按一定频率发送到远程数据中心。

[0027] 具体地,在步骤S1中,所述第一历史时间窗口的第一历史水质数据是指水质监测节点采集历史中某个时间段内的历史水质数据。

[0028] 所述步骤S1具体包括如下步骤:

[0029] 选择一个连续时间段内的历史水质数据作为模型训练数据,并均分为Y组,其中,Y为正整数;

[0030] 计算每一组所述模型训练数据的残差,并计算Y组所述模型训练数据的残差平均值;

[0031] 通过交叉法计算所有所述模型训练数据的残差标准差;

[0032] 利用delta学习算法确定所述第一历史时间窗口内每个历史水质数据的初始权重系数,并根据模型训练数据的残值标准差计算出待采集的下一个水质数据的第一预测值;

[0033] 基于学生分布函数和所述第一预测值计算出待采集的下一个水质数据的第一预测区间。

[0034] 例如,将N个历史水质数据划分为10组,则10组所述水质数据与的残差平均值计算公式如下:

$$[0035] \quad \mu_r = \frac{1}{N} \sum_{i=1}^{10} \left(\sum_{j=1}^{\frac{N}{10}} r_j^i \right),$$

[0036] 其中, r_j^i 表示第i个分组中第j个水质数据的残差, μ_r 表示所有水质数据残差平均值,且 $1 \leq i \leq 10$,i为正整数, $1 \leq j \leq N$,j为正整数。

[0037] 10组所述水质数据的残差标准差计算公式如下:

$$[0038] \quad d = \left[\frac{1}{N-1} \sum_{i=1}^{10} \left(\sum_{j=1}^{\frac{N}{10}} (r_j^i - \mu_r^i)^2 \right) \right]^{\frac{1}{2}},$$

[0039] 其中, μ_r^i 表示第i个分组中水质数据残差平均值。通过delta学习算法确定所述第一历史时间窗口内每个数据的初始权重系数,所述第一历史时间窗口内的数据初始权重系数随时间窗口的变化而进行动态计算,计算公式如下:

$$[0040] \quad W_{ij}(t+1) = W_{ij}(t) + \alpha (d_i - y_i) x_j(t),$$

[0041] 其中, α 是学习因子,这里设为0.0001; $x_j(t)$ 是激活函数,这里设为常量1; d_i 是期望值; y_i 为实测值。

[0042] 通过所述初始权重系数和所述残差标准差计算所述水质数据的第一预测值,计算公式如下:

$$[0043] \quad X'_{t+1} = d + \sum_{i=0}^{q-1} w_i X_i ,$$

[0044] 其中, X_i 表示 t 时间窗口内第 i 个水质数据的值, w_i 表示元素权重, d 为残差标准差。

[0045] 基于学生分布函数和所述第一预测值计算所述水质数据的第一预测值的第一预测区间,计算公式如下:

$$[0046] \quad X'_{t+1} \pm T_{\frac{\alpha}{2}, n-1} \times d \times \sqrt{1 + \frac{1}{N}} ,$$

[0047] 其中 t 为学生分布, d 是所述残差标准差。

[0048] S2、获取实时采集的实时水质数据,并判断所述实时水质数据是否落在所述第一预测区间内,如果是,则判断所述实时水质数据为正常数据,如果否,则执行步骤S3。

[0049] 具体地,在步骤S3中,获取水质监测节点采集的实时水质数据,并比较实时采集的所述实时水质数据是否在所述第一预测区间的范围内。

[0050] 如果所述实时水质数据落在所述第一预测区间内上下限范围内,则表明所述实时水质数据为正常数据,则按时间序列删除时间窗口中的第一个水质数据,并补入所述实时水质数据,构造新的时间窗口。

[0051] 如果所述实时水质数据超出所述第一预测区间的上下限,则表示该数据有较高的概率是噪音数据,需要通过后续的第二历史时间窗口进行二次验证。

[0052] 即,在本发明实施例中,利用第一个历史时间窗口来判断实际采集到的实时水质数据是否有可能是噪音数据,而且,利用第二个时间窗口对可能的噪音数据进行二次验证。

[0053] S3、从所述实时水质数据开始按时间倒序获得第二历史时间窗口内第二历史水质数据,并基于所述第二历史时间窗口的第二历史水质数据计算水质数据的第二预测区间。

[0054] 具体地,在步骤S3中,从所述实时水质数据位置开始取第二个历史时间窗口内所有水质数据,并按时间倒序计算第二个历史时间窗口内下一个水质数据的预测范围。

[0055] 所述步骤S3包括如下步骤:

[0056] 利用delta学习算法确定所述第二历史时间窗口内每个数据的初始权重系数,并利用步骤S1中得到的残差标准差计算出水质数据的第二预测值;

[0057] 基于学生分布函数和所述第二预测值计算出所述水质数据的第二预测区间。

[0058] 需要说明的是,所述第二历史水质数据内的数据数量与所述第一历史水质数据内的数据数量相同。

[0059] S4、判断所述实时数据是否落入所述第二预测区间,如果是,则判读所述实时水质数据为正常数据,如果否,则所述实时水质数据为噪音数据。

[0060] 具体地,在步骤S4中,如果所述实时数据为正常数据,则按时间序列删除时间窗口中的第一个水质数据,并补入所述实时水质数据,构造新的时间窗口;

[0061] 如果所述实时数据为噪音数据,则按时间序列删除时间窗口中的第一个水质数据,并加入基于所述时间窗口计算得到的水质数据的预测值,构造新的时间窗口。

[0062] S5、返回步骤S1,对后续的水质数据进行噪音数据检测。

[0063] 需要说明的是,在本实施例中,时间窗口内的数据按先进先出的队列模式并根据

时间序列进行动态调整。而且，

[0064] 相较于现有技术，本发明提供的基于双时间窗口验证的水质监测噪音数据实时检测方法采用双时间窗口对可能的噪音数据进行两次验证，从而有效降低了噪音数据误检率，进一步减少了系统因噪音数据引起的误报警，提升水质监测系统可靠性和稳定性。

[0065] 对于本领域技术人员而言，显然本发明不限于上述示范性实施例的细节，而且在不背离本发明的精神或基本特征的情况下，能够以其他的具体形式实现本发明。因此，无论从哪一点来看，均应将实施例看作是示范性的，而且是非限制性的，本发明的范围由所附权利要求而不是上述说明限定，因此旨在将落在权利要求的等同要件的含义和范围内的所有变化囊括在本发明内。不应将权利要求中的任何附图标记视为限制所涉及的权利要求。

[0066] 此外，应当理解，虽然本说明书按照实施方式加以描述，但并非每个实施方式仅包含一个独立的技术方案，说明书的这种叙述方式仅仅是为清楚起见，本领域技术人员应当将说明书作为一个整体，各实施例中的技术方案也可以经适当组合，形成本领域技术人员可以理解的其他实施方式。

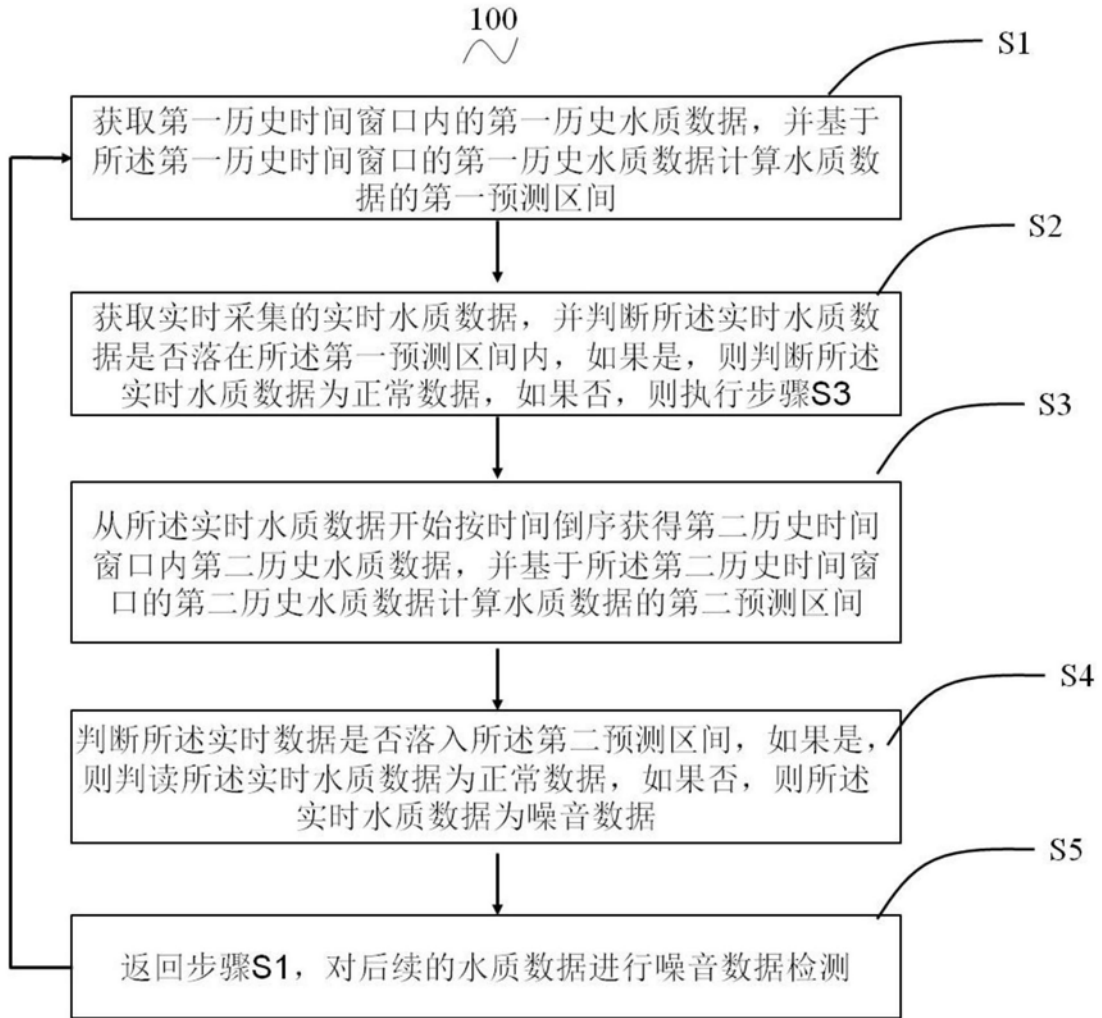


图1