

文章编号:1001-9081(2015)11-3112-04

doi:10.11772/j.issn.1001-9081.2015.11.3112

基于多维滑窗的异常数据检测方法

花青*, 许国艳, 张叶

(河海大学 计算机与信息学院, 南京 211100)

(*通信作者电子邮箱 517837002@qq.com)

摘要:随着数据流的广泛运用,数据流中异常数据的检测问题也引起了更多的关注。现有的卡尔曼滤波算法需要的历史数据量虽然小,但只适用于单个异常点的检测,对于复杂连续的异常值检测效果较差。针对这个问题,提出一种水文传感器分级标注模型,并在此基础上提出一种基于多维影响因子的卡尔曼滤波算法,加入空间、时间、起源三个维度的影响因子,在天气和汛期等影响因素改变时,对系统模型的控制参数进行适当调整,并且对测量噪声进行更加准确的估计,提高异常检测的准确性。实验结果证明,所提算法在保证运行时间相近的前提下,检测的错误率远低于基于遗忘因子的卡尔曼(AKF)算法和基于小波的卡尔曼(WKF)算法。

关键词:异常数据检测; 数据起源; 分级标注模型; 多维影响因子; 卡尔曼算法

中图分类号: 补充分类号 **文献标志码:**A

Improved Kalman algorithm for abnormal data detection based on multidimensional impact factors

HUA Qing*, XU Guoyan, ZHANG Ye

(College of computer and information, Hohai University, Nanjing Jiangsu 211100, China)

Abstract: With the widespread application of the data flow, the abnormal data detection problem in data flow has caused more attention. Existing Kalman filtering algorithms need small amount of historical data, but they only apply to single abnormal point detection. The effect to complex continuous outlier points is poor. In order to solve the problem, a Kalman filtering algorithm based on multidimensional impact factors was proposed. The algorithm joined the three dimensions of impact factor as space, time, provenance as well. In case of different weather and flood season, the algorithm adjusted the controlling parameters of system model parameters, and got a more accurate estimate of measurement noise. The detection accuracy of the algorithm could be improved significantly. The experimental results show that under the premise of guaranteeing similar running time, the detection error rate of this algorithm is far lower than Amnesic Kalman Filtering (AKF) and Wavelet Kalman Filtering (WKF) algorithms.

Key words: abnormal data detection; data provenance; graded tagging model; multidimensional impact factor; Kalman algorithm

0 引言

随着互联网和传感器技术的发展,数据流这种新型数据不断地渗透到各个领域中。由于数据流信息量庞大,想要将数据流全部存储在数据库中是不现实的,因此,常见的办法是忽略数据流中的冗余项,只抽取一些概要项保存在数据库中。在数据流的处理环节中,对异常数据进行实时检测具有重要意义,例如,在水文传感器网络中,传感器节点可以感知周围的水位或者雨量数值。通过对传感器数据流进行检测,可以判断是否发生了异常事件,从而及时进行相应的处理,避免因为误差造成的错误报警。这里所谓的异常数据是区别于正常数据的,使人怀疑的数据^[1]。这些异常数据并非随机偏差,而是由不同的机制产生,例如测量错误、计算机录入错误、执行错误和人为错误等。异常数据也有可能就是数据的真实性

质的反映,这些数据比一般数据所包含的信息更有价值,例如暴雨引起的水位急剧上涨等,这些数据更需要保留与研究。通过对水文数据中的异常数据进行分析,可发现由环境误差、设备故障引起的错误信息,或者是自然灾害引起的数据跳变,这种分析的方法就是异常数据的检测。

数据流中的异常数据检测方法往往涉及复杂的计算和过滤,需要进行数据转换、修改、融合,或以其他方式处理之后才能用于具体的分析和建模,并且需要使用多个源的传感器数据来支持^[2]。如此复杂的、难以长期存储的数据一经传播就很难保证其准确性和可用性。在这样的网络环境中,起源信息起到了至关重要的作用,它能够让用户理解、验证、再生数据以及确定派生数据产品的质量^[3]。

现有的卡尔曼滤波算法需要的历史数据量虽然小,但只适用于单个异常点的检测,对于复杂连续的异常值检测效果

收稿日期:2015-06-17;修回日期:2015-07-24。基金项目:国家科技支撑计划项目(2013BAB06B04);江苏省自然科学基金资助项目(BK20130852);2013年江苏水利科技项目(2013025);中国华能集团公司总部科技项目(HNKJ13-H17-04)。

作者简介:花青(1991-),男,江苏盐城人,硕士研究生,主要研究方向:大数据、数据起源;许国艳(1971-),女,内蒙古赤峰人,副教授,博士,CCF会员,主要研究方向:大数据、数据起源;张叶(1990-),女,山东泰安人,硕士研究生,主要研究方向:大数据、数据起源。

较差^[4-6]。本文分析了已有的经典检测算法,并进行改进,首先提出一个针对水文传感器的开放起源模型,在此基础上提出了基于空间、时间和起源 3 个维度的影响因子的改进算法,提高了检测算法的准确性。

1 相关工作

1.1 开放起源模型

起源是关于实体、活动以及参与生产数据和事物的人员的信息,这些信息可以用来形成对其质量、可靠性和可信度的评估^[7-8]。数据起源组织 PROV(Provance)的文档定义了一个相应的序列化和其他支持定义的模型,使起源信息在异构环境中互操作,开放起源模型(Open Provenance Model, OPM)把起源描述成一个带有标注的有向无环图,该图是由类型化的节点和有向边组成。图中的节点为:圆代表 Entity、矩形代表 Activity、六边形则代表 Agent。3 个节点的定义如下:

Entity 为实体,物理的、虚拟的、概念上的以及其他种类的事物都可以称为实体;

Activity 为活动,描述实体是如何产生、变化的;

Agent 为代理,是过程的发起者和控制者,代理可以是一个人、一段代码、一个无生命的物体、一个组织或者其他负责的实体。

其边缘定义如下:

wasGeneratedBy 从一个实体到一个活动,描述由活动产生实体的过程;

used 从一个活动到一个实体,是活动开始使用实体的标志;

wasInformedBy 从一个活动到另一个活动,描述一个活动使用了另一个活动产生的实体;

wasDerivedFrom 从一个实体到另一个实体,描述实体到实体的转变;

wasAssociatedWith 从一个活动到一个代理,描述代理在活动中所扮演的角色;

wasAttributedTo 从一个实体到一个代理,描述实体到代理的归因;

actedOnBehalfOf 从一个代理到另一个代理,描述代理赋予自己或另一个代理的权利和职能。

1.2 卡尔曼滤波算法

卡尔曼滤波算法是目前应用比较广泛的状态估计算法,它是一种时域内的递推算法,在计算过程中,算法不需要大量的历史数据,只需要不断地对下一状态进行预测,一旦得到新的测量数据,就对预测算法进行修正,从而提高预测精度。卡尔曼滤波算法需要的数据量小,便于实现实时检测,并且算法实现简单,因此被广泛地运用在异常数据在线监测等应用当中^[6]。

假设系统的离散模型为:

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{F}\mathbf{x}_k + \mathbf{G}\mathbf{u}_k + \mathbf{v}_k \\ \mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{w}_k \end{cases} \quad (1)$$

其中: \mathbf{x}_k 为 n 维状态向量, \mathbf{u}_k 为 m 维控制向量, \mathbf{y}_k 为 r 维输出

向量, \mathbf{v}_k 为 n 维过程干扰向量, \mathbf{w}_k 为 r 维测量噪声向量,假设 \mathbf{v}_k 和 \mathbf{w}_k 均为离散的高斯白噪声序列。

图 1 为卡尔曼算法的核心公式。

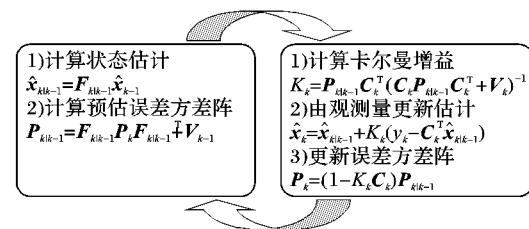


图 1 卡尔曼算法核心公式

1.3 小波变换

小波是一种特殊的信号,其波动能量有限,并且相对集中在局部区域。小波与傅里叶变换中的正弦信号区别在于,傅里叶正弦波的波动范围是一个无穷区间,而小波信号只在一个很小的时间范围内波动^[9]。图 2 为哈尔小波的波形。

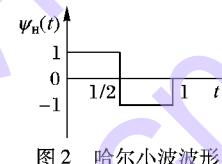


图 2 哈尔小波波形

在信息系统中,经常需要对信号进行去噪、压缩和检测等处理,为了提高信号处理的有效性和效率,通常将信号从时域映射到变换域。基于小波的信号处理一般是先将信号进行离散小波变换,在小波变换域对信号进行处理,再由离散小波逆变换得到处理后的信号。基于小波变换的信号处理有很多优点,这是因为小波变换具有以下 4 个特性:

- 1) 在实际应用中,很多信号的小波展开系数都集中在较少的系数上,这简化了数据处理的复杂度。
- 2) 小波展开具有良好的时频描述,可以更有效地分离出信号中的分量。
- 3) 小波基灵活多变,不同信号可以用不同的小波基展开,从而更好地分离出信号分量。
- 4) 离散小波变换运算简单,不需要复杂的微分和积分,并且重构算法也具有极高的效率。

小波去噪的步骤为:

- 1) 选择合适的小波基,对信号进行分级展开,得到不同尺度的小波展开系数;
- 2) 选择相应的阈值,舍去阈值以外的系数;
- 3) 对处理后的系数进行逆小波变换,重构信号。

2 基于多维影响因子的卡尔曼滤波算法

2.1 水文传感器数据起源模型

本文结合水文传感器数据流实时性、突发性、易失性、連續性、周期性和空间相关性,提出一个基于面向对象的水文数据起源分级标注模型,对某水文传感器网络进行建模,分级类别为属性级、对象级、表级和数据库级。在此模型基础上,本文提取出了影响水文传感器测量值的影响因素,并提出了汛期因子等影响因子,应用到卡尔曼滤波算法中。表 1 展示了水位传感器标注模型,图 3 是其对应的 OPM 流程。

表 1 水位传感器分级标注模型

级别	实体		活动		代理	
	符号	描述	符号	描述	符号	描述
属性级	E1	传感器编码	A1	上传	Ag1	传感器
对象级	E2	测量时间	A2	更新实时表	Ag2	数据库
	E3	测量值	A3	更新基本情况	Ag3	传感器维护人员
对象级	E4	实时数据表	A4	更新使用情况	Ag4	水利工作人员
	E5	传感器基本信息	A5	更新聚合信息	Ag5	第三方服务程序
	E6	传感器使用信息	A6	更新水利表	Ag6	数据库维护人员
	E7	传感器聚合信息	A7	更新天气表		
表级	E8	水利情况表	A8	维护数据库		
	E9	天气情况表				
库级	E10	数据库维护信息				

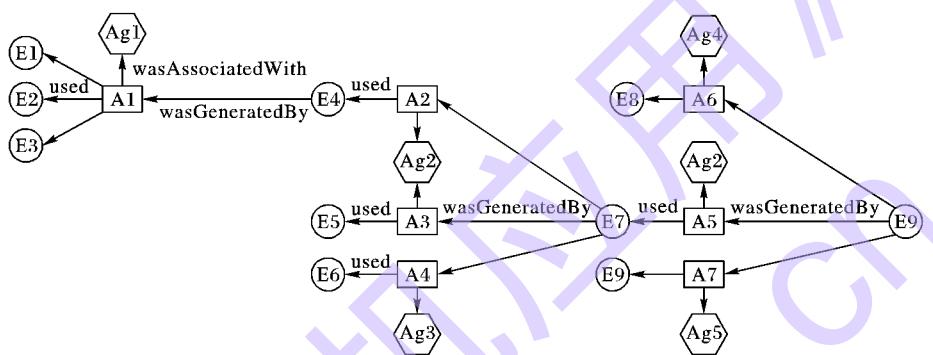


图 3 水位传感器 OPM 流程

2.2 卡尔曼滤波算法的改进

王永利等^[10]提出一种基于遗忘因子的 Kalman 滤波检测算法, 对不同时刻的历史数据给予不同的影响权值, 既限制了滤波的记忆长度, 又提高了存在大量随机成分的数据流预测准确率, 但是并没有解决测量噪声必须已知的缺陷。

高羽等^[11]利用小波变换可以实时分离出信号和噪声的性质, 提出了一种在未知测量噪声情况下的卡尔曼滤波算法。该算法可以实时计算出观测噪声的变化, 即对测量噪声方差进行实时估计, 从而解决了卡尔曼滤波算法在未知测量噪声条件下的失效问题。

本文结合上述两个算法的优点, 提出一种改进的卡尔曼滤波算法, 并且加入空间、时间、起源 3 个维度的影响因子, 对于测量噪声进行更加准确的估计。算法改进如下:

1) 改进小波变换提取测量噪声的算法。

本文算法在进行卡尔曼滤波之前, 先选取检测点之前长度为 L 的观测窗口进行小波变换, 将测量值进行一维 5 层分解, 提取出高频成分作为测量噪声, 从而解决卡尔曼算法在测量噪声未知情况下严重失真的问题。此外, 在小波变换提取测量噪声的基础上, 本文算法还加入了遗忘因子来进一步优化测量噪声的估计。假设在长度为 L 滑窗内, 小波变换提取出的测量噪声值分别为 w_1, w_2, \dots, w_L , 则 $L+1$ 时刻的测量噪声值为:

$$w_{L+1} = \sum_1^L \lambda_i w_i / L \quad (2)$$

其中: $\lambda_i = (1 - \lambda) / (1 - \lambda^i)$ 为 i 时刻测量噪声 w_i 的加权遗忘因子, $\lambda \in (0, 1)$ 。

2) 加入多维影响因子。

常用的传感器数据融合方法主要是时间和空间维度的融

合, 空间维度是指同一时刻, 相邻传感器节点之间数据的融合; 时间维度是指同一节点、不同时刻数据的融合^[12~13]。本文将起源信息融入到时空维度中, 提出了空间、时间和起源这 3 个维度的影响因子。

在空间维度中, 本文算法不仅考虑相邻传感器之间的数据关系, 同时还加入了本文建立的数据起源模型中提出的上下游关系和干流支流关系, 不同关系的节点进行不同的数据融合。空间影响因子定义如下:

$$\lambda_s(t) = \sum_{i=0}^N |y_i(t) - \hat{y}_i(t)| / N \quad (3)$$

其中: $\lambda_s(t)$ 为 t 时刻的空间影响因子, $y_i(t)$ 为 t 时刻传感器 i 的观测值, $\hat{y}_i(t)$ 为 t 时刻传感器 i 的预测值。

在时间维度中, 本文算法在比较同一传感器节点不同时刻测量值的同时, 还会结合该节点的天气因素和传感器使用情况等影响因子, 对系统模型的参数进行动态调整, 使得系统模型在实际应用中具有更好的适应性。时间影响因子定义如下:

$$\lambda_t(t) = \sum_{j=0}^N |y_i(t-j) - \hat{y}_i(t-j)| / N \quad (4)$$

其中: $\lambda_t(t)$ 为 t 时刻的时间影响因子, $y_i(t-j)$ 为 $t-j$ 时刻传感器 i 的观测值, $\hat{y}_i(t-j)$ 为 $t-j$ 时刻传感器 i 的预测值。

起源维度是指结合卡尔曼算法的预测值和影响水位值的起源因素, 进行决策层次的分析, 从而最终判断数据是否为异常值, 并记录判断依据等信息, 方便后续对检测算法进行验证以及重演检测过程的工作。起源维度中的汛期影响因子是针对河流汛期与非汛期时段水位变化差别而提出的, 其定义如下:

$$\lambda_f(t) = \frac{\frac{1}{N} \sum_{i=1}^N P_{N,i}}{\frac{1}{M} \sum_{i=1}^M P_{M,i}} \quad (5)$$

其中: $\lambda_f(t)$ 为 t 时刻的汛期影响因子; N, M 分别为非汛期、汛期的采样天数; $P_{N,t}, P_{M,t}$ 分别为非汛期、汛期的 t 时刻采样水位值。

2.3 算法实现

基于多维影响因子的卡尔曼滤波检测(Multi-Dimensional Factor Kalman Filtering, MDF-KF)算法如下。

算法 基于多维影响因子的卡尔曼滤波检测算法。

Input: $\lambda, y(t), x(0), Qw(0), v(0), Qv(0), L, \xi$

Output: $\hat{y}(t), t$

1) 一步状态预测

$$\hat{x}(t|t-1) = F(t)x(t-1) + v(t-1)$$

2) 一步均方误差预测

$$P(t|t-1) = F(t)P(t-1)F^T(t) + Qv(t)$$

3) 小波提取出测量噪声

$$w(t) = \text{WaveDec}(y(t-1))$$

4) 加权系数(遗忘因子)

$$\lambda_t = (1 - \lambda)/(1 - \lambda^t)$$

5) 测量噪声估计

$$Qw(t) = (1 - \lambda_t)Qw(t-1) + \lambda_t[w(t)w^T(t) - C(t)P(t|t-1)C^T(t)]$$

6) 滤波增益

$$K(t) = \frac{P(t|t-1)C^T(t)}{[C(t)P(t|t-1)C^T(t) + Qw(t)]}$$

7) 更新均方误差差

$$P(t) = [1 - K(t)C(t)]P(t|t-1)$$

8) 状态估计

$$\hat{x}(t) = \hat{x}(t|t-1) + K(t)[y(t) - C(t)x(t|t-1)]$$

9) 数据预测值

$$\hat{y}(t) = C(t)\hat{x}(t) + w(t)$$

10) 更新系统噪声

$$v(t) = \hat{x}(t) - F(t)\hat{x}(t|t-1)$$

11) 多维度判定

if ($[\lambda_s(t) > \lambda_f(t)\xi] \& [\lambda_t(t) > \lambda_f(t)\xi]$) then

12) 输出

Output $\hat{y}(t), t$

13) 修正

$y(t) = \hat{y}(t)$ 转 1)

14) end if

15) 更新 $y(t)$ 并转 1)

3 实验结果

为检测算法的性能,本文对智慧滁河项目中用到的水位传感器网络,选取其中 1 个传感器,对其水位数据进行异常检测。实验机器的配置是 2.20 GHz i7-4702MQ CPU, 8 GB 内存, 1 TB 硬盘, 在 Windows 8.1 系统下, 运用 Matlab 实验箱实现算法。本文对基于遗忘因子的卡尔曼(Amnesic Kalman Filtering, AKF)算法^[10]、基于小波的卡尔曼(Wavelet Kalman Filtering, WKF)算法^[11]和本文算法的准确性和时间进行对比,从而验证本文算法的准确性和效率。

实验 1 对上述选取的传感器的 2015 年 2 月份的测量数据进行采样,得到 1000 条数据,选取其中 5 条不连续的记录分别加上或减去阈值 ϵ 作为单个异常点,选取 5 条连续数据减去阈值 ϵ 作为连续异常点,另外,数据中有一段连续 10 条递减 1 cm 的水位数据,作为水位骤减的异常情况。分别用 AKF 算法、WKF 算法以及本文算法(MDF-KF)对上述数据进行检测,对 3 个方法的检测准确率进行对比分析,检测精度为 3 cm,表 2 为算法检测错误率对比。由实验结果可以看出,

WKF 方法的误检率较高,对连续异常点的检测稍好于 AKF, AKF 对单个异常值和连续不变的异常值检测效果好,但对连续变化的异常值检测效果较差,本文算法结合两者优点,降低了检测的错误率。

表 2 算法检测错误率对比

算法	数据总量	异常点数	连续异常点检出数	漏检数	误检数	错误率/%
WKF	1000	20	6	12	8	2.0
AKF	1000	20	5	10	0	1.0
MDF-KF	1000	20	12	3	4	0.7

实验 2 同样利用上述数据,对本文算法的空间、时间和起源维度影响因子分别进行测试,分析它们对检测结果的影响情况。表 3 为 3 个维度的影响因子的检测错误率情况,实验表明,空间维度影响因子可以减少误检数,时间影响因子可以明显增加连续异常点的检出数,而起源维度影响因子可以减少漏检数。

表 3 3 个维度的影响因子的检测错误率

维度	数据总量	异常点数	连续异常点检出数	漏检数	误检数	错误率/%
空间	1000	20	6	8	4	1.2
时间	1000	20	10	5	6	1.1
起源	1000	20	8	3	8	1.1
三者结合	1000	20	12	3	4	0.7

实验 3 对上述选取的传感器的 2015 年 2 月份的测量数据进行采样,分别取 100 条、200 条、500 条、1000 条和 2000 条记录,对 3 种算法的运行时间进行对比分析,如图 4 所示。由实验可知,AKF 算法的运行时间与数据量呈线性增加关系,因为 AKF 算法不需要对测量噪声进行实时调整。而 WKF 算法中需要使用小波变换对噪声进行实时修正,所以运行时间与数据量呈指数增加关系。本文算法的运行时间和 WKF 相近,但检测准确率却大大高于另外 2 种算法。

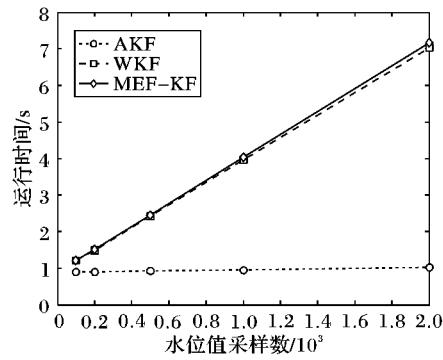


图 4 3 种算法运行时间对比

4 结语

本文研究了已有的异常检测算法,并提出一种水文传感器的分级标注模型,在此基础上提出了一种基于三维影响因子的检测算法。本文方法与前人的算法相比,主要有两处改进:第 1 点是将前人的两个算法结合,将遗忘因子加入到测量噪声的小波变换中;第 2 点是将起源信息中的影响因子加入到算法中,提高算法的准确性。实验结果表明,本文提出的基于三维影响因子的异常检测算法能有效提高检测的准确性。

(下转第 3156 页)

- 2012: 5763 – 5768.
- [2] LIN W H. A Gaussian maximum likelihood formulation for short-term forecasting of traffic flow [C]// Proceedings of the 2001 Intelligent Transportation Systems. Piscataway: IEEE, 2001: 150 – 155.
- [3] SMITH B L, WILLIAMS B M, KEITH OSWALD R. Comparison of parametric and nonparametric models for traffic flow forecasting [J]. Transportation Research Part C: Emerging Technologies, 2002, 10 (4): 303 – 321.
- [4] CLARK S D, DOUGHERTY M S, KIRBY H R. The use of neural networks and time series models for short term traffic forecasting: a comparative study [C]// Proceedings of Seminar D Held at the PTRC European Transport, Highways and Planning 21st Summer Annual Meeting. Berkshire: Transport Research Laboratory, 1993: 151.
- [5] CASTRO-NETO M, JEONG Y S, JEONG M K, et al. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions [J]. Expert Systems with Applications, 2009, 36 (3): 6164 – 6173.
- [6] XIE Y, ZHANG Y, YE Z. Short-term traffic volume forecasting using Kalman filter with discrete wavelet decomposition [J]. Computer-Aided Civil and Infrastructure Engineering, 2007, 22(5): 326 – 334.
- [7] YIN J, GAO J, WANG B, et al. Research on methods of urban road traffic condition identification [J]. Journal of Qingdao University: Engineering and Technology Edition, 2012, 27(3): 84 – 87. (尹婧, 高军伟, 王彬, 等. 城市交通状态判别方法的研究[J]. 青岛大学学报: 工程技术版, 2012, 27(3): 84 – 87.)
- [8] ZHANG X, LU H. Short-term traffic flow characteristics and instance analysis [J]. Journal of Highway and Transportation Research and Development, 2009, 26 (S1): 62 – 68, 73. (张晓利, 陆化普.)
- [9] JIN S, WANG D, XU C, et al. Short-term traffic safety forecasting using Gaussian mixture model and Kalman filter [J]. Journal of Zhejiang University Science A, 2013, 14(4): 231 – 243.
- [10] SMITH B L, WILLIAMS B M, OSWALD R K. Comparison of parametric and nonparametric models for traffic flow forecasting [J]. Transportation Research Part C: Emerging Technologies, 2002, 10 (4): 303 – 321.
- [11] GONG W, CAI Z, JIANG L. Enhancing the performance of differential evolution using orthogonal design method [J]. Applied Mathematics and Computation, 2008, 206(1): 56 – 69.
- [12] BAR-SHALOM Y, LI X, KIRUBARAJAN T. Estimation with application to tracking and navigation: theory, algorithm and software [M]. New York: Wiley, 2001: 321 – 340.
- [13] BREST J, GREINER S, BOKOVI B, et al. Self-adapting control parameters in differential evolution: a comparative study on numerical benchmark problems [J]. IEEE Transactions on Evolutionary Computation, 2006, 10(6): 646 – 657.
- [14] VESTERSTROEM J, THOMSEN R. A comparative study of differential evolution, particle swarm optimization and evolutionary algorithms on numerical benchmark problems [C]// Proceedings of the 2004 IEEE Congress Evolutionary Computation. Piscataway: IEEE, 2004: 1980 – 1987.
- [15] HUANG F Y. Integration of an improved particle swarm algorithm and fuzzy neural network for Shanghai stock market prediction [C]// Proceedings of the 2008 Workshop on Power Electronics and Intelligent Transportation System. Piscataway: IEEE, 2008: 242 – 247.

(上接第 3115 页)

参考文献:

- [1] LIU H, TONG D, CHEN S. The research of algorithm of attribute-based detection of outlier data [J]. Computer Science, 2005, 32 (5): 164 – 166. (刘洪涛, 童德利, 陈世福. 一种基于属性的异常点检测算法[J]. 计算机科学, 2005, 32(5): 164 – 166.)
- [2] HANDCOCK R N, SWAIN D L, BISHOP-HURLEY G J, et al. Monitoring animal behaviour and environmental interactions using wireless sensor networks, GPS collars and satellite remote sensing [J]. Sensors, 2009, 9(5): 3586 – 3603.
- [3] DAI C, WANG T, ZHANG P. Survey of data provenance technique [J]. Application Research of Computers, 2010, 9(4) : 3215 – 3221. (戴超凡, 王涛, 张鹏程. 数据起源技术发展研究与综述 [J]. 计算机应用研究, 2010, 9(4): 3215 – 3221.)
- [4] WANG Y, LIU J, YU Z, et al. Multi-granularity annotations and multi-resolution compression algorithm for sensor data provenance [J]. Journal of Nantong University: Natural Science Edition, 2010, 9(3): 92 – 101. (王永利, 刘景春, 于占龙, 等. 传感数据起源信息的多粒度标注与多分辨率压缩[J]. 南通大学学报: 自然科学版, 2010, 9(3): 92 – 101.)
- [5] LIU Z, WANG J, WU Y, et al. Kalman filter outlier detection methods based on M-estimation [C]// Proceedings of the 2013 32nd Chinese Control Conference. Piscataway: IEEE, 2013: 4652 – 4655.
- [6] GUO X, GE F, GUO L. Improved adaptive Kalman filtering and its application in acoustic maneuvering target tracking [J]. Acta Acustica, 2011, 36(6): 611 – 618. (郭鑫, 葛凤翔, 郭良浩. 改进的自适应 Kalman 滤波及其在声学机动目标跟踪中的应用[J]. 声学学报, 2011, 36(6): 611 – 618.)
- [7] LI Y. Data provenance's annotation schema and description model [J]. New Technology of Library and Information Service, 2007(7) : 10 – 13. (李亚子. 数据起源标注模式与描述模型[J]. 现代图书情报技术, 2007(7): 10 – 13.)
- [8] MOREAU L, CLIFFORD B, FREIRE J. The open provenance model core specification (v1.1) [J]. Future Generation Computer Systems, 2011, 27(6): 743 – 756.
- [9] ZHANG D, YANG Y, QIN Q. Nonlinear adaptive wavelet transform for lossless image compression [J]. Wuhan University Journal of Natural Sciences, 2007, 12(2): 267 – 270.
- [10] WANG Y, XU H, DONG Y, et al. Detection and repairing method for outliers over data streams [J]. Journal of Applied Science, 2006, 24(3): 256 – 261. (王永利, 徐宏炳, 董逸生, 等. 数据流上异常数据的在线检测与修正[J]. 应用科学学报, 2006, 24 (3): 256 – 261.)
- [11] GAO Y, ZHANG J. Kalman filter with wavelet-based unknown measurement noise estimation and its application for information fusion [J]. Acta Electronica Sinica, 2007, 35(1): 108 – 111. (高羽, 张建秋. 小波变换域估计观测噪声方差的 Kalman 滤波算法及其在数据融合中的应用[J]. 电子学报, 2007, 35(1): 108 – 111.)
- [12] YANG X. EKF-based anomaly data detection algorithm of wireless sensor network [D]. Shenyang: Northeastern University, 2013: 42 – 45. (杨晓龙. 基于 EKF 的无线传感器网络异常数据检测算法研究[D]. 沈阳: 东北大学, 2013: 42 – 45.)
- [13] HILL D J, MINSKER B S. Anomaly detection in streaming environmental sensor data: a data-driven modeling approach [J]. Environmental Modelling and Software, 2009, 25(9): 1014 – 1022.