基于物联网的焊机监测系统中突发检测算法研究*

徐 明[1,2], 刘广钟[1]

(1. 上海海事大学 信息工程学院,上海 201306; 2. 复旦大学 上海市智能信息处理重点实验室,上海 200433)

摘 要: 针对基于物联网的焊机监测系统中异常情况检测和预警的需要,提出一种基于多种群萤火虫的突发检测算法;通过不同种群萤火虫的协同工作实现滑动窗口大小的优化选择与配置,提高突发检测模型的处理速度与检测性能。仿真实验结果表明,基于多种群萤火虫的突发检测算法在突发概率或者最大滑动窗口大小相同的情况下,处理时间少于传统的突发检测算法,并具有更高的准确率和召回率。

关键词: 物联网; 焊机监测系统; 突发检测; 滑动窗口

中图分类号: TP393 **文献标志码**: A **文章编号**: 1001-3695(2013)01-0142-04

doi:10.3969/j.issn.1001-3695.2013.01.035

Research on burst detection algorithm in Internet of things-based welding machine monitoring system

XU Ming[1,2], LIU Guang-zhong[1]

(1. *College of Information Engineering, Shanghai Maritime University, Shanghai* 201306, *China*; 2. *Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai* 200433, *China*)

Abstract: For the need of detecting and early warning of abnormal conditions in an Internet of things-based welding machine monitoring system, this paper presented a multi-population firefly based burst detection algorithm for reducing processing time and improving detection performance through optimizing the sizes of slide windows. The simulations experimental results demonstrate that the burst detection algorithm consumes less processing time in the same conditions of burst probability or maximum slide window size, and also achieves higher precision and recall rate than traditional algorithms.

Key words: Internet of things; welding machine monitoring system; burst detection; slide window

物联网被认为是继计算机和互联网之后世界信息产业的第三次浪潮。物联网通过射频识别、传感器、全球定位系统、激光扫描器等信息传感设备,按照约定的协议,把任何物体与互联网连接起来,进行信息交换和通信,以实现智能化识别、定位、跟踪、监控和管理的一种网络,通俗地说就是可实现感知世界的网络[1,2]。为了更好地实现焊接现场的精细化管理和精益制造的水平,有必要在焊接全过程中应用信息化技术来提高焊接效率和降低能耗。通过基于物联网的焊机监测系统可以实时监控焊机的工作状态及相关信息,实现焊接工时、物量、动力能源消耗等基础数据的计算与精细化管理,对操作者的焊接规范进行监督,及时对操作者超规范焊接情况进行管理;并且能够通过焊机异常情况报警历史信息分析焊机出现故障的原因,而焊机监测系统中的异常数据可以通过对物联网数据流进行突发检测来正确识别。

1 相关工作

基于物联网的焊机监测系统接收到的数据是一组持续不断到达的、带有时间标签的元素序列,因此需要用数据流挖掘的方法进行处理。数据流的本质特征在于其变化性及不可预测性。作为数据流的一种基础且重要的数据分析方法,数据流突发检测能够及时检测到特定元素数量的异常变化,因此在许多领域有着广泛的应用,包括监测宇宙中高能光子伽玛射线的爆发、检测金融领域异常波动以及监控互联网中分布式拒绝服务攻击[3~5]。针对数据流中的突发检测问题,研究人员已经提出了一些方法。采用偏移小波树(shifted wavelet tree, SWT)[6]可以表示数据流滑动窗口中所有的元素,能解决弹性窗口问题,但是搜索效率有待提高。偏移二叉树(shifted binary tree, SBT)[7]实现了对数级复杂度内多滑动窗口数据流序列的突发检测。偏移聚合树(shifted aggregation tree, SAT)[8]可以根据应用场景的不同灵活选择聚合塔(aggregation pyramid)的层数以及每层中子节点的个数,减少详细搜索的次数以降低时间复杂度。基于平均似然比的突发检测方法利用平均似然比对突发信号作贝叶斯检测,解决了以往提出的突发检测方法要求最佳定时采样的问题[9]。基于小波的突发检测算法利用聚合向量对原始数据流的窗口和观测数据进行了压缩处理[10]。基于迁移小波树的突发检测算法在突发定义的基础上,设定适当的阈值,并利用迁移小波树的结构有效地定位突发[11]。然而,传统的数据流突发检测技术都是针对单一数据流设计单一突发

收稿日期: 2012-06-11; **修回日期**: 2012-07-30 **基金项目**: 国家自然科学基金资助项目(61202370);上海市科学技术委员会浦江人才计划项目(11PJ1404300);上海市教委科研创新重点资助项目(12ZZ151);上海市智能信息处理重点实验室开放课题(IIPL-2011-008);上海海事大学科研基金资助项目(20110049)

作者简介: 徐明(1977-),男,安徽含山人,博士,主要研究方向为 P2P 网络、传感器网络、计算智能理论与方法(mingxu@shmtu.edu.cn);刘广钟(1962-),男,江苏徐州人,教授,博士,主要研究方向为计算机网络、水声通信、网格计算、CIMS 技术、分布式数据库等.

检测算法来进行的,因此无法适用于物联网环境下多源数据流的突发检测:a)由于无法预测突发时间的长短,因此在基于滑动窗口模型的突发检测中很难确定滑动窗口的大小以及同时对多个滑动窗口大小进行突发检测;b)针对物联网中不同的应用场景需要采用相应的数据模型表示多源数据流以确保能够高效进行突发检测;c)采用何种度量方式评估是否达到突发以及阈值的设定问题有待解决。

群智能优化算法是近几十年发展起来的仿生模拟进化算法,具有操作简单、宜于并行处理、鲁棒性强等特点。由于群智能优化算法的性能特征及其自组织能力与无线传感器网络自组织性质的相似性,本文将研究以萤火虫算法[12]为代表的群体智能优化算法,提出一种基于多种群萤火虫(multi-population firefly, MPF)的突发检测算法;每个种群的萤火虫监测一种属性的物理量,并通过不同种群萤火虫的协同工作实现滑动窗口大小的优化选择与配置,提高突发检测模型的处理速度与检测性能。

2　体系结构与功能模块

基于物联网的焊机监测系统采用基于感知层、网络层和应用层的三层架构,主要由无线数据采集模块、服务器/客户端模块、车间焊接过程远程监控软件和可视化显示四个模块组成。其中:感知层负责采集焊机的工作电压、工作电流等物理量,并将其转换为数字信息;网络层主要实现信息的传送和通信以及相关部分的处理,采用基于 ZigBee 的无线通信;应用层实现数据处理、管理和人机交互等功能。

1)无线数据采集模块主要实现两个功能,即采集焊机的电流、电压信号,并以无线的方式把数据发送到无线网关,应用软件使用 API 通过无线网关实现对所有数据采集模块数据的轮询。无线数据采集模块内置 ZigBee 无线通信模块,综合运用信号转换和光电隔离技术,具有高可靠性、高精度、结构紧凑、安装使用方便等特点。根据设备状况和实际使用情况,数据采集模块安装于设备内部,天线置于设备外表面,提高系统的可靠性和美观度,能适应较大的信号输入范围,可以接收 4 路模拟信号(2 路 DC 0 ~ 100 V 和 2 路 DC 0 ~ 100 MV),并对输入信号进行了转换和处理,确保信息不发生畸变、失真、延迟现象,具有较高的可靠性和良好的线性度。此外,系统还可以采用数据定时主动上报模式,定时时间的设置可以自由设定。主动上报模式下,数据采集设备可以进入低功耗休眠模式。由于数据采集设备有开关信号输入和高低电平输入接口。当检测到电平变化,立即结束休眠,上报数据。服务器端则负责数据的接收,不需要发送查询命令。

2)服务器/客户端模块中,服务器运行数据采集模块软件,负责实时采集焊机焊接数据,并将数据保存到数据库中;客户端运行车间焊接过程远程监控软件客户端,支持多个用户同时实时访问焊接检测数据库。

3)车间焊接过程远程监控软件提供数据采集、处理、焊机监测、统计分析的功能,并具有开放的数据库,提供给其他管理软件使用。软件实时采集各台焊机的工作电压和电流,通过电压和电流的变化得到焊机的工作状态,包括工作、报警、待机、关机四种不同状态,并按区域对焊机进行单机和总体监测。软件将采集到的数据存入数据库,用于后续的分析处理,提供单机/多机焊机状态、负荷率、开机/待机时间、班组工作分析、区域分析等统计分析功能。

a)可视化显示模块可以实现点、线、面三个层次的显示:a)基于节点标志的显示,将所有节点按网络内的标志符大小进行实时数据和状态显示,以便监测者抽取和导出监测数据,同时观察节点异常状态,对整个网络系统进行及时有效的维护;b)基于时间曲线的显示,将一个或多个传感器节点的历史或实时数据显示在以时间为横轴的曲线图上,以便于监测者分析一段时间内焊机焊接数据的变化规律,结合环境因素分析原因,完善预警机制;c)基于地理信息的显示,结合地理信息系统,将所有焊机的位置及其实时数据显示在地图上,以便于监测者把握全局监测结果。

此外,焊机运行状况监测数据和处理结果还可以通过 IP 骨干网、焊机监测现场的本地中继网传送到现场工作人员随身携带的便携式监视终端,以便及时通报故障。

3　突发检测模型和算法

3.1　突发检测模型和问题定义

在突发检测的研究中,最关键的问题是突发检测模型的设计,它决定了后续突发检测算法性能及其评价指标。基于点检测的自动机模型不用预先指定阈值等参数,适合数据相对均匀正态分布的情况。但是,在数据海量以及复杂多变的情况下,自动机模型有较强的领域局限性。因此,目前大部分突发检测模型和算法都采用基于聚合检测的滑动窗口模型,并且采用偏移小波树方法或者聚合塔结构来解决弹性窗口问题。然而,同一突发检测模型在不同的应用领域或者应用场景下效果不尽相同。例如,有的场景下元素聚合值的绝对值超过阈值就可认为是突发,而有的场景下只有元素聚合值与前一时间段聚合值的比值超过阈值才认为是突发。此外,突发检测是典型的监测任务,但不是一个孤立的工作,并且通常与其他监测工作同时存在。特别是在焊机监测环境下,为了更好地覆盖监测区域,通常需要多个传感器节点之间协同工作以便处理多源数据流。

图 1 描绘了基于物联网的焊机监测系统中多源数据流的突发检测模型。当数据流到达网络节点后,首先通过分类器进行特征提取与分类;然后由调度器放入相应的突发检测队列,经过多线程操作处理器处理后,将处理结果送到路由选择器并决定数据的走向,即传给存储管理器进行归档处理或者输出。负载均衡器不断对输出性能进行检测,如果性能过低,则采取措施减轻系统负载直到达到标准。由于数据流到达网络节点的速率是动态变化的,因此即使某一节点的平均负载不是很高,但在峰值期间,一个节点将有可能遇到暂时的负载峰值,其数据处理的等待时间会受到影响。所以,为了使数据处理时间最小化,需要使用负载均衡策略来避免暂时过载。

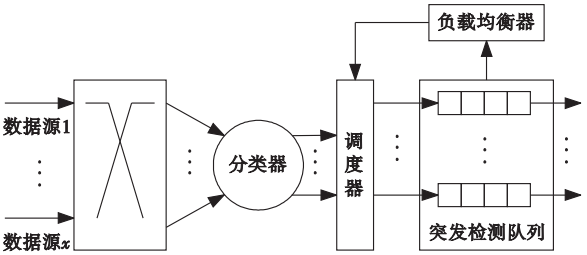

图1　多源异质数据流的突发检测模型

定义 1　给定 T 个单元时间序列 $t=0,1,\cdots,T$ 上的 n 个事件集合 $E=\{e_1,e_2,\cdots,e_n\}$,数据流来自 x 个传感器节点,节点 j

上的事件序列表示为 $E_j = \{e_{1j}, e_{2j}, \cdots, e_{nj}\}$，则事件集合 E 为

$$E = \sum_{j=1}^{x} E_j \tag{1}$$

假设事件到达的速率为 λ，并且各个事件之间独立同分布，则任意时间的事件数量服从泊松分布：

$$P(n|\lambda) = e^{-\lambda}\frac{\lambda^n}{n!} \quad n = 0,1,2,\cdots \tag{2}$$

假设不同时刻到达的事件独立，则事件流中事件数量 $\{n_1, n_2, \cdots, n_T\}$ 的概率分布为

$$P(n_1, n_2, \cdots, n_T|\lambda_1, \lambda_2, \cdots, \lambda_T) = \prod_{t=1}^{T} P(n_t|\lambda_t) \tag{3}$$

定义 2 当时间序列 T 上的事件集合 E 内的事件数量 B 大于或者小于平均事件数一个量级时称为突发，此时 $B > \mu E + c\sigma E$ 或者 $B < \mu E - c\sigma E$。其中 μ 为均值，c 为常量，σ 为标准差。

d 维空间多元高斯密度函数分布定义为

$$p(x, M_j, C_j) = \frac{1}{\sqrt{(2\pi)^d |C|}} e^{-(x-M_j)^T C_j (x-M_j)} \tag{4}$$

其中：M_j 代表 d 维平均向量，C_j 是一个代表协方差向量的 $d \times d$ 对角线矩阵。

单信息源情况下使用效用函数 $U(x,a)$（其中 $a \in S_e$，S_e 为所有可能执行的一个集合）最大化，效用函数表示当真实状态是一个特定值 x 时，执行 a 所带来的效用的一种度量。期望的效用则用 β 表示：

$$\beta(p(x|Z^k), a) = \int \{U(x,a)\} p(x|Z^k) \mathrm{d}x \tag{5}$$

则最优执行 $\bar{a}$ 是使期望效用最大的策略，即

$$\bar{a} = \max \beta(p(x|Z^k), a) \tag{6}$$

式(6)等价于式(7)的最大化：

$$\int U(x,a) p(Z^k|x) p(x) \mathrm{d}x \tag{7}$$

当前方子系统定期报告时间到达、检测到异常情况或接收到数据查询请求时，符合条件的传感器节点将传感单元获得的信息经过 A/D 转换及初步处理，形成监测数据包；监测数据包经过一跳或多跳无线传输到达汇聚节点，再经过一跳或多跳无线传输接入 IP 骨干网，最后经 IP 骨干网传送到后方子系统；后方子系统的数据处理模型单元对收到的监测数据按照突发检测算法进行处理，判断焊机的焊接状态是否发生异常，以便及时作出预警处理。

3.2 突发检测算法

本文提出的基于多种群萤火虫的突发检测算法中每个种群的萤火虫负责监测一种属性的物理量，并且也包含基本萤火虫算法的两个要素，即荧光素亮度和吸引度。荧光素亮度与萤火虫所在位置的目标函数适应度值有关，而萤火虫在某一时刻的适应度值是该萤火虫所在节点进行突发检测时所采用的滑动窗口大小的函数。吸引度决定了亮度较弱的萤火虫往亮度较强的萤火虫方向移动的距离，即亮度较弱的萤火虫所在节点滑动窗口大小变化的幅值。通过各萤火虫荧光素亮度和吸引度的不断更新，实现突发检测滑动窗口大小的优化，从而提高突发检测模型的处理速度与检测性能。从数学角度对萤火虫算法的优化机理进行如下描述：

定义 3 萤火虫的荧光素亮度定义为

$$I(r) = I_0 e^{-\gamma \cdot r^2} \tag{8}$$

其中：I_0 为萤火虫的自身荧光素亮度（$r = 0$ 处），与目标函数值 φ 相关，φ 值越大则自身荧光素亮度越高；γ 为荧光素吸收系数，用来表示荧光素亮度随着距离的增加而逐渐衰减的速率，可设置为常数；r 表示萤火虫之间的距离。

定义 4 萤火虫的吸引度定义为

$$\beta(r) = \beta_0 e^{-\gamma \cdot r^2} \tag{9}$$

其中：β_0 为 $r = 0$ 处的吸引度，γ 和 r 的意义同上。当 $r = 1/\sqrt{\gamma}$ 时称为特征距离，此时吸引度会从 β_0 显著变化到 $\beta_0 e^{-1}$。

假设任意两只萤火虫 i 和 j 所在节点的滑动窗口大小分别为 x_i 和 x_j，则这两只萤火虫之间的距离表示为 $r_{ij} = \| x_i - x_j \|$，则萤火虫 i 被吸引到荧光素亮度更高的萤火虫 j 方向的位移定义为

$$\Delta x_i = \beta_0 e^{-\gamma r^2}(x_j^t - x_i^t) + \alpha \cdot \varepsilon_i, x_i^{t+1} = x_i^t + \Delta x_i \tag{10}$$

其中：α 为步长因子，是一个随机参数；ε_i 是一个服从高斯分布的随机数向量。

算法 1 基于多种群萤火虫的突发检测算法

输入：滑动窗口数量 K，滑动窗口大小初始值 $x = [x_1, x_2, \cdots, x_K]$，对应的阈值大小 $f = [f_1, f_2, \cdots, f_K]$，最大迭代次数 L。

输出：滑动窗口大小的最优值 x^{opt}；

```
每只萤火虫初始荧光素亮度 I_i(0) = I_0;
for i = 1 to K do
  x_i←x. InitialValues( );
endfor
iter = 0;
while iter < L do;
  for i = 1 to K do
    for j = 1 to K do
      if φ(x_i) < φ(x_j) then
        根据式(10): x_i←x_i - Δx_i;
      else 根据式(10): x_j←x_j - Δx_j;
        根据式(9)更新 β_i 和 β_j 的值;
        m = log x_i;
        y = sum(x[c, c - 1 + m]);
        if y > f_i
        search x[c, c - 1 + j];
      endif
    endif
    endfor
    return x_i^opt
    endfor
iter + +;
endwhile
return x^opt;
```

算法 1 的第 1 ~ 4 行对各滑动窗口大小进行初始化；第 5、6 行根据输入的迭代次数进行突发检测算法的迭代操作；第 7 ~ 12 行比较各个节点滑动窗口大小的适应度值，并以此决定萤火虫的移动方向和移动距离，然后更新相关萤火虫的吸引度值；第 13 ~ 16 行采用 sum 作为聚集函数，当超过阈值时进行详细搜索，可以排除很多明显没有达到阈值的监测数据项。

4 仿真结果与分析

本文的仿真环境采用 Crossbow 公司生产的 MICA2[13] 节点为标准，节点的通信距离为 50 m。在仿真过程中采用随机数据来模拟无线传感器网络节点的感知数据，包括焊接电流和电极电压。具体方法如下：首先将每个节点在每维上的数值 v 随机初始化为正常范围内的某个值。在随后的每一轮实验中，数值 v 较上轮的最大变化幅度为 $f(f > 0)$。假设第 r 轮中第 i 维

上的数值为 $v(r)[i]$,在第 $r+1$ 轮中数值变化为 $v(r)[i]+\text{rand}(-f,f)$。其他实验参数与默认值如表 1 所示。

表 1　实验参数与默认值

实验参数	默认值	实验参数	默认值
传感器节点数量	80	焊接电流范围/kA	3.3 ~ 6.0
网络节点平均度数	5	电极电压范围/kN	2.1 ~ 5.1
汇聚节点数量	4	焊接数据集大小	86 281
数据包大小/Byte	40		

表 1 给出了仿真实验的参数与默认值。本文将比较 SBT、SAT 和基于多种群萤火 MPF 的在处理时间、准确率(precision)和召回率(recall)等指标上的性能。准确率是指检测到的真实突发数目与所有真实突发数目的比值,召回率是指检测到的真实突发数目与所有检测到的突发数目的比值。

第 1 组实验比较了不同突发检测算法在不同突发概率下的处理时间。在该组实验中,最大滑动窗口大小为 64,仿真结果如图 2 所示。随着突发概率的增大,三种突发检测算法的处理时间都会增多。在突发概率相同的情况下,MPF 的处理时间平均要比 SAT 少 16.8%,比 SBT 平均少 42.7%。

第 2 组实验比较了不同突发检测算法的处理时间与最大滑动窗口大小之间的关系,仿真结果如图 3 所示。三种突发检测算法的处理时间都与最大滑动窗口大小成正比。在相同滑动窗口大小的情况下,MPF 的处理时间要少于 SAT 和 SBT 的处理时间,并且随着滑动窗口尺寸的增大,MPF 的性能优势更为明显。当最大滑动窗口大小为 8 时,MPF 的处理时间相当于 SAT 的 61.1%,SBT 的 50.6%;当最大滑动窗口大小为 512 时,MPF 的处理时间仅相当于 SAT 的 54.3% 以及 SBT 的 39.5%。

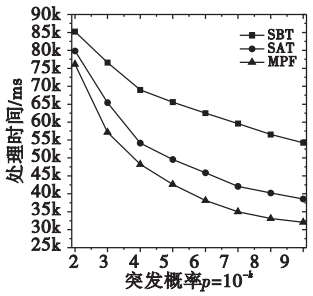

图2　不同突发概率下突发检测的处理时间

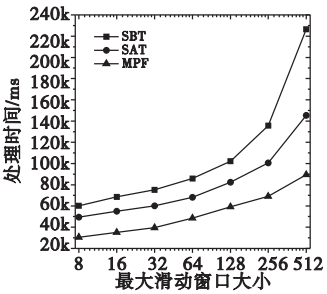

图3　最大滑动窗口大小对处理时间的影响

第 3 组实验主要比较最大滑动窗口大小对准确率的影响,仿真结果如图 4 所示。随着最大滑动窗口大小的增大,三种突发检测算法的准确率都有所上升。但是,当最大滑动窗口大小超过 128 时,三种突发检测算法准确率上升的幅度明显减缓。在最大滑动窗口大小相同的情况下,MPF 的准确率要高于 SAT 和 SBT 的准确率。

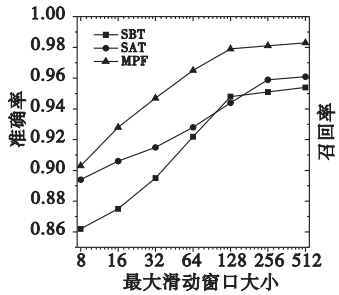

图4　最大滑动窗口大小对准确率的影响

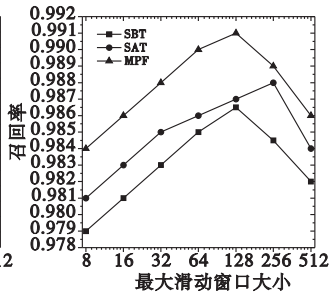

图5　最大滑动窗口大小对召回率的影响

第 4 组实验主要比较最大滑动窗口大小对召回率的影响,仿真结果如图 5 所示。随着最大滑动窗口大小的增大,三种突发检测算法的召回率呈现出先增大后减小的趋势。当最大滑动窗口大小达到 128 时,SBT 的召回率达到了它的最高值 0.985,而 MPF 的召回率达到了它的最高值 0.991;当最大滑动窗口大小达到 256 时,SAT 的召回率达到了它的最高值 0.988。但是,在同等情况下,MPF 的召回率要高于 SAT 和 SBT 的召回率。

5　结束语

作为数据流突发检测算法的一个普遍特性,速度与准确率作为数据流算法的两个最关键参数始终是一对矛盾体,设计算法时需在两者之间取得良好的折中。一般来说,突发检测算法的处理时间与最大滑动窗口大小成正比关系。随着最大滑动窗口大小的增大,处理时间会变长,突发检测的速度变慢,但同时准确率也会有所上升。与准确率的变化规律不同,召回率在最大滑动窗口大小增大的过程中呈现出先增大后减小的趋势,这与准确率到了后期上升幅度变缓有关。进一步的研究工作将考虑突发检测算法的假阳性和假阴性问题,通过对现有算法进行优化处理,使假阳/阴性率综合起来最低。

参考文献:

[1] 吴振强, 周彦伟, 马建峰. 物联网安全传输模型[J]. 计算机学报,2011,34(8): 1351-1364.

[2] 沈苏彬, 毛燕琴, 范曲立,等. 物联网概念模型与体系结构[J]. 南京邮电大学学报:自然科学版,2010,30(4): 1-8.

[3] HE Dan, PARKER D S. Topic dynamics: an alternative model of bursts in streams of topics[C]//Proc of the 16th ACM SIGKDD Internatio-nal Conference on Knowledge Discovery and Data Mining. New York:ACM Press, 2010: 443-452.

[4] SUN A, ZENG Da-jun, CHEN Hsin-chun, *et al*. Burst detection from multiple data streams: a network-based approach[J]. IEEE Trans on Systems, Man, and Cybernetics,2010,40(3): 258-267.

[5] 秦首科, 钱卫宁, 周傲英. 基于分形技术的数据流突变检测算法[J]. 电子学报,2006,17(9):1969-1979.

[6] ZHU Yun-yue, SHASHA D. Efficient elastic burst detection in data streams[C]//Proc of the 9th ACM International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2003: 336-345.

[7] LERNER A, SHASHA D, WANG Zi-hua, *et al*. Fast algorithms for time series with applications to finance, physics, music, biology, and other suspects[C]//Proc of ACM International Conference on Management of Data. New York:ACM Press, 2004: 965-968.

[8] ZHANG Xin, SHASHA D. Better burst detection[C]//Proc of the 22nd International Conference on Data Engineering. Washington DC: IEEE Computer Society, 2006: 146-150.

[9] 黄渊凌,路友荣,袁强. 基于平均似然比的鲁棒性突发检测[J]. 电子与信息学报,2010,32(2): 345-349.

[10] SINGH L, SAYAL M. Privately detecting bursts in streaming, distributed time series data[J]. Data and Knowledge Engineering, 2009,68(6):509-530.

[11] 朱维军. 基于迁移小波树的突发检测分析[J]. 中国科技信息, 2010,10(6): 36-38.

[12] YANG Xin-she. Firefly algorithms for multimodal optimization[C]//Proc of the 5th International Conference on Stochastic algorithms: Foundations and Applications. Berlin: Springer-Verlag, 2009: 169-178.

[13] FARSHCHI S, NUYUJUKIAN P, PESTEREV A. A TinyOS-enabled MICA2-based wireless neural interface[J]. IEEE Trans on Biomedical Engineering,2006,53(7):1416-1424.