

基于数据流截获技术的仪器状态监测方案

毛宏宇^{1,2}, 王丰贵¹, 王文良², 王书士²

(1. 吉林大学仪器科学与电气工程学院, 长春 130026; 2. 航空仪器设备计量总站, 北京 100070)

摘要: 针对由计算机控制的仪器工作状态监测问题, 提出一种基于计算机串口的数据流监测方案, 采用纯软件的方式, 规避硬件监测的风险, 通过截获串口与仪器设备之间的数据流进行模式匹配, 实现仪器状态监测。为提高数据流变化的检测效率, 采用加权的核拟合密度估计算法代替传统的 M-核估计方法, 降低系统资源的占用率。

关键词: 数据流; 过滤器驱动; 状态监测; 核拟合

Instruments Status Monitoring Scheme Based on Data Stream Interception

MAO Hong-yu^{1,2}, WANG Feng-gui¹, WANG Wen-liang², WANG Shu-shi²

(1. College of Instrument Science & Electrical Engineering, Jilin University, Changchun 130026;

2. Measuring Central Station of Aeronautical Instrument and Equipment, Beijing 100070)

【Abstract】As for status monitoring for computer-controlled instruments, a monitoring scheme for data stream based on computer serial interface is proposed, which avoids the risk form hardware monitoring by pure software technology and judges instrument status by interception and pattern matching data stream between instruments and serial interface. In order to improve the detection efficiency of data stream and reduce the occupancy rate of resources, weighted core-fitting density approximation approach is chosen instead of traditional M-core.

【Key words】 data stream; filter driver; status monitoring; core fitting

1 概述

随着国家配备的大型综合仪器设备数量和水平的提高, 对大型仪器设备利用效果的评价以及利用率的监测逐渐成为研究热点^[1-2]。文献[3]提出基于 Internet 的设备远程状态监测系统, 实现了仪器设备运行状态的远程实时监测。文献[4]采用能量色散 X 射线荧光光谱分析方法实现了航空发动机大颗粒金属磨削监控。从民用设备到航空工业, 设备状态监测得到了广泛的应用。然而, 对于大型科学仪器有效工作机时的监测和记录, 国内尚处于起步阶段。文献[5]设计了以单片机为核心的科学仪器工作状态监测装置, 通过检测电压信号获取仪器的有效工作机时信息。文献[6]提出基于虚拟仪器技术的设备状态监测系统, 采用数据采集卡实现旋转机械设备的机时监测和诊断。这些研究对于仪器状态监测的发展起到了积极的推动作用, 但这些成果的应用都依赖于能够采集设备的光、电信号, 并通过传感器或数据采集卡传输至中央处理器, 而对于大部分封闭型较好、对环境要求苛刻的精密仪器, 实现难度较大。调研发现, 大部分科学仪器通过 RS-232 总线方式与计算机相连, 其中典型的是德国布鲁克公司生产的 X 射线衍射仪。

本文提出基于计算机串口数据流截获技术的仪器状态监测方案, 采用软件方式对计算机串口数据流的流量状态进行识别, 有效避免了硬件监测的风险, 同时降低了实施难度。

2 监测模型

数据流监测模型如图 1 所示, 按照数据流的处理功能划分为过滤驱动、特征检测与模式识别、数据的网络化应用 3 个部分, 分别实现数据流的动态获取、数据流的分析与处

理及仪器状态信息的远程传送功能。

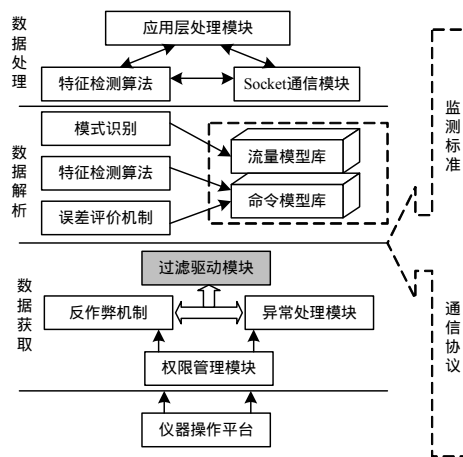


图 1 数据流监测模型

3 串口数据流截获

与科学仪器通信的计算机串口在操作系统中被注册为独占设备, 为了截获数据必须将独占设备更改为共享设备, Spooling 技术实现了应用层的设备共享, 原理是建立虚拟设

基金项目: 国家科技基础条件平台建设基金资助项目(2005DKA 10103)

作者简介: 毛宏宇(1981 -), 男, 博士研究生, 主研方向: 测试计量, 分布式测控; 王丰贵, 博士研究生; 王文良, 教授级高级工程师; 王书士, 工程师

收稿日期: 2009-06-17 E-mail: 24427732@qq.com

备,通过总线的分时复用实现数据共享,该技术对系统资源要求较高,本文采用过滤器驱动技术,借用系统内核的预留资源,降低了资源占用率。

3.1 基于驱动程序的截获原理

Windows 系统驱动程序层次结构为数据流的截获提供了途径。IRP(In/Out Request Packet)是数据从内核层到应用层传输的基本形式,操作系统为了实现用户的请求,必须将 IRP 从顶层传输到底层,通过植入带有特定功能的过滤器驱动程序,即可截获流经功能驱动程序的 IPR,从而实现接口数据流的监测。

3.2 串口过滤器驱动程序

过滤程序分为 2 个模块:驱动程序模块和服务程序模块。驱动程序模块负责数据的截获,服务程序模块负责数据的处理,驱动程序流程如图 2 所示。

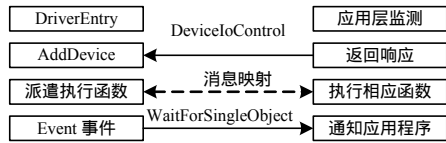


图 2 驱动程序流程

(1)DriverEntry 例程:负责装入并执行驱动程序的某些全局初始化操作。

(2)AddDevice 例程:基本职责是创建一个设备对象并把它连接到以 pdo 为底的设备堆栈中。分为以下 4 个步骤:

- 1)调用 IoCreateDevice 创建设备对象,并建立一个私有的设备扩展对象;
- 2)寄存一个或多个设备接口,以便应用程序能知道设备的存在;
- 3)初始化设备对象的 Flag 成员;
- 4)调用 IoAttachDeviceToDeviceStack 函数把新设备对象放到堆栈上。

(3)等待串口 Event 事件发生,执行相应过滤程序,通知应用层进行处理。

4 串口数据流的状态识别

截获到的原始数据流只是监测程序不能识别的“乱码”,需要对这些数据流进行统计和处理。在数据流的分析处理领域中,密度估计是常用的一项基本技术,对大量数据进行密度估计能够获得数据分布的大致情况,该项技术一直得到研究学者的高度重视。但传统的密度估计方法由于空间复杂度和时间复杂度都较高,对内存空间和处理时间的需求都比较大,不适合数据流动态变化、数据量大的特点,因此,不能直接应用于数据流。

4.1 核密度估计与 M-核

核密度估计的原理与直方图技术有些类似,直方图记录了在每个区间中点的个数或频率,使得图中矩形条的高度随着数值个数的多少而变化,但是直方图很难给出较为精确的密度估计。核密度估计也是计算某一点周围的点的个数,只不过是优先考虑近处的点。具体来说,如果数据为 x_1, x_2, \dots, x_n , 在任意点处 x_n 的一种核密度估计为

$$f_n(x) = \frac{1}{nh} \cdot \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) = \frac{1}{n} \cdot \sum_{i=1}^n K_h(x-X_i) \quad (1)$$

其中, $K_h(t) = h^{-1}K(h^{-1}t)$, $K(\cdot)$ 为核函数, h 为带宽。带宽越大,估计的密度函数就越平滑,选得较小则估计的密度曲线和样本拟合得较好。

在传统的核密度估计方法中,每个数据点都用一个核函数表示。如果从分布中抽取出的随机变量的样本大小为 n , 则按式(1)计算 $k(x)$ 估计时也需要 n 个核函数。因此,要能够有效地处理数据流,就必须找出一种办法减少计算 $k(x)$ 时所需要用的核函数的个数。根据核函数具有可累加性,可以将参数相同的核函数用一个有权重的相同参数的核函数所代替,如果一个或多个数据点有相同的值,那么就可以用一个有权重的核函数去代表这些数据点。在本文中,核函数上的权重的大小就是核函数所代表数据点的个数。

通常,只能代表 1 个数据点的核函数为简单核,能代表多个数据点的核函数为 M-核。式(2)表示了一个中心为 X_i 、带宽为 h 、权重为 ρ 的 M-核核函数。

$$f_i(x) = \frac{\rho}{h} K\left(\frac{x-X_i}{h}\right) \quad (2)$$

选取经验模型中满足正态分布的高斯核函数为核函数:

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \quad (3)$$

构造时间步进为 1 ms、区间为 1 s、每 200 ms 流量数据点阶跃 200 Kb/s 的随机数据流,其统计分布曲线见图 3。

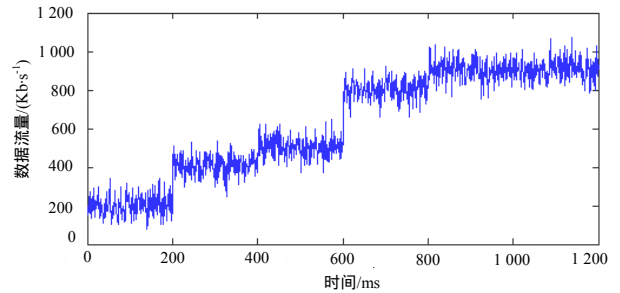


图 3 正态分布模拟数据流

将数据带入式(1)得到如图 4 所示的核密度估计曲线。该方法可以对不同数据流量的变化规律做出判断,在对误差进行估计后,即可应用于串口数据流的流量识别。

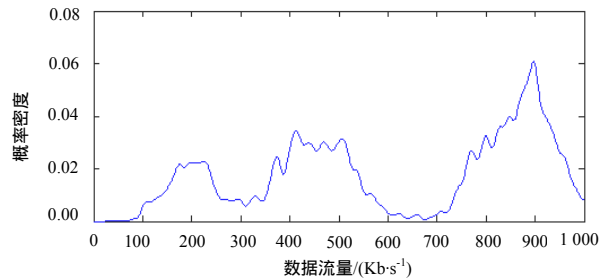


图 4 数据流量的概率密度曲线

4.2 改进的 M-核算法

简单地将中心相同的核函数合并起来并不能有效减少所要保留的核函数的个数,但可以通过拟合中心相近的核函数来进一步减少所要保留的核函数的个数。2 个不同中心的核函数拟合后得到的核函数并不是标准核函数,但是可以用一个非标准核函数去近似它,这里把这种数据流密度估计的方法称为核拟合方法。拟合公式可以定义为

$$\rho_i K_{h_i}(x-X_i) + \rho_j K_{h_j}(x-X_j) \approx (\rho_i + \rho_j) K_{h_m}(x-X_m^*) \quad (4)$$

其中,左边表示 2 个中心分别为 X_i 和 X_j 的并且带有不同权重的核函数 K_i, K_j ; 右边表示权重为 $\rho_i + \rho_j$ 的核函数;中心点为 X_m^* , 用右边的核函数代替左边核函数。选取以下 2 组

数据做拟合实验:(1)曲线1:中心点 $X_i=100$, 权重为 $\rho_i=10$, $h=1000$;(2)曲线2:中心点 $X_j=120$, 权重为 $\rho_j=8$, $h=1200$ 。2组数据的核函数做线性叠加得到曲线3, 2组数据进行参数拟合后得到曲线4, 通过仿真可知, 拟合曲线与线性叠加曲线近似重合, 保证了拟合后的失真度在允许范围内。

结合状态监测特定应用背景, 改进的M-核算法并没有追求计算精度的提高, 而是在计算效率上进行了改进, 见图5, 其中, 实线表示利用标准核函数做密度估计时的计算效率曲线, 虚线为改进后的M-核算法效率曲线。近似的线性统计曲线可以看出, 随着数据流量的加大, 优势会更明显。

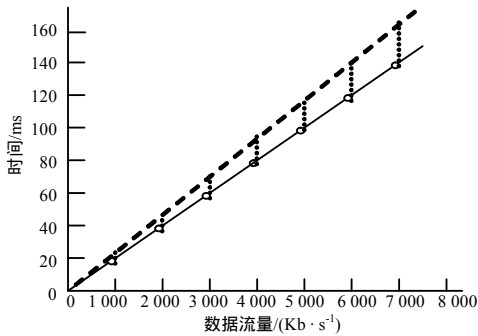


图5 改进前后M-核算法的效率对比曲线

在很多情况下, 不能直接找到一个合适的核函数等于2个不同中心的核函数之和。它们之间的差值就是拟合所带来的误差, 本文用 ε 表示。下文以2个单点数据为例, 计算其误差 ε 。用拟合核函数去代替2个权重不同、中心点相近、带宽不同的核函数之和, 所产生的误差为

$$\varepsilon = \int |(\rho_1 + \rho_2)K_{h_m}(x - X_m^*) - \rho_1 K_{h_1}(x - X_1) + \rho_2 K_{h_2}(x - X_2)| dx$$

采用高斯函数作为核函数, 2个核函数 K_1, K_2 参数分别为 ρ_1, X_1, H_1 和 ρ_2, X_2, H_2 , 且 $|X_1 - X_2| < \sigma_1, |\rho_1 - \rho_2| < \sigma_2$ 。当 σ_1, σ_2 足够小时, 可以用 K_{h_m} 近似代替 K_{h_1} , 用近似代替 X_1 或 X_2 , 拟合后的误差可推导如下:

$$E = \int |\rho_1 K_{h_1}(x - X_1) + \rho_2 K_{h_2}(x - X_2) - (\rho_1 + \rho_2) K_{h_m}(x - X_m^*)| dx$$

$$\int |\rho_1 K_{h_1}(x - X_1) + \rho_2 K_{h_2}(x - X_2) - (\rho_1 + \rho_2) K_{h+\frac{1}{2}\sigma_2}(x - (X_1 + \frac{1}{2}\sigma_1))| dx$$

$$\int |\rho_1 (K_{h_1}(x - X_1) - K_{h+\frac{1}{2}\sigma_2}(x - (X_1 + \frac{1}{2}\sigma_1)))| dx +$$

$$\int |\rho_2 (K_{h_2}(x - X_2) - K_{h+\frac{1}{2}\sigma_2}(x - (X_1 + \frac{1}{2}\sigma_1)))| dx$$

$$\frac{2\rho_1}{\sqrt{2\pi}} \frac{\sigma_1(\sigma_2 + 4)}{8} + \frac{2\rho_2}{\sqrt{2\pi}} \frac{\sigma_1(9\sigma_2 + 12)}{8}$$

结果表明, 拟合后误差所带来的误差是可控的, 该误差在状态监测的允许范围之内, 不影响串口数据流的识别精度。在本文的应用背景下, 把相同核函数的流量数据点的个数定义为该核函数的权值, 当 σ_1, σ_2 足够小时, 核函数和权值都采取近似相等的做法, 因此, 权值的大小不影响误差的大小。

5 测试与应用

5.1 性能及可靠性测试

测试环境为: 操作系统: Windows XP SP2 Pro; 串口资

源: 2个; 处理器: Intel Pentium 4 Processor 630, 3.00 GHz; 内存: 1 GB。

(1) 测试截获串口操作

测试方法: 启动串口调试软件, 并设置好串口号、波特率、数据位、校验位及停止位信息, 点击“打开”和“关闭”串口。此时监测器记录了串口的操作过程, 并显示出了相关的预置信息。测试结果表明: 该监测器能够在不影响串口正常应用的前提下, 截获了串口的操作信息。

(2) 测试监测及通信能力

测试方法: 编写基于串口的大型科学仪器模拟控制平台, 控制该平台对仪器做样过程进行模拟, 使测试数据通过设定的串口进行传输。监测器能够检测到模拟做样过程, 并将有效数据传输到监测服务器。

测试结果表明, 该监测器可以在输入不同格式、不同流量数据的情况下对数据流进行监测和统计, 不同的仪器状态对应不同的流量模型, 依据输入不同的流量模型, 可以判断科学仪器的工作状态; 测试结果也同时显示解析流量的时间随着数据格式的复杂、数据流量的增加而增加, 表明本文提出的核拟合密度估计算法可以应用在科学仪器的流量监测领域。

5.2 应用情况

在大型科学仪器工作状态监测系统项目中, 应用该监测器对德国布鲁克公司生产的X射线单晶衍射仪D8 ISCOVER GADDS进行了状态监测。该监测器能够识别科学仪器开机、关机等工作状态, 并能统计机时数据, 传输到远程监测服务器。

6 结束语

本文提出的基于数据流截获技术的仪器状态监测方案采用软件技术实现了状态信号的采集, 避免了硬件采集带来的风险, 已成功应用于国家大型科学仪器工作状态监测系统。该方案对于工作过程由计算机其他接口控制的仪器设备状态监测也具有重要的应用价值。

参考文献

- [1] Jaesoon C, Park J W, Chung Jinhan, et al. An Intelligent Remote Monitoring System for Artificial Heart[J]. IEEE Transactions on Information Technology in Biomedicine, 2005, 9(4): 564-573.
- [2] Li Jinbo. Design and Realization of USB2.0 Interface in Equipment Condition Monitoring[C]//Proceedings of the 8th Industry Meter and Automatization. Shanghai, China: [s. n.], 2007.
- [3] 韩捷, 董辛. 基于Internet的设备远程状态监测与故障诊断[J]. 机械科学与技术, 2004, 23(12): 1503-1506.
- [4] 朱子新. 航空发动机大颗粒金属磨屑监控技术[J]. 航空维修与工程, 2006, 20(3): 30-32.
- [5] 马骏骁, 李伟. 科学仪器工作状态监测装置的设计与实现[J]. 分析化学, 2007, 35(3): 466-468.
- [6] 王少红, 徐小力. 虚拟仪器技术下的状态监测及趋势预示系统[J]. 北京机械工业学院学报, 2002, 17(1): 1-5.

编辑 顾姣健