

改进拉伊达准则在 LabVIEW 温度采集系统中的应用

李蕊¹, 王慧忠², 冯建文¹

(1. 杭州电子科技大学 计算机学院, 杭州 310018; 2. 浙江省计量科学研究院, 杭州 310016)

摘要: 对温度进行采集是温控系统的前提之一, 好的采集系统可以真实有效地反映温度的变化, 可靠的采集数据可以为后续的系统性能分析提供有力的保障; 在采集的数据当中, 粗大误差总是存在的, 有效地剔除这些粗大误差可以帮助实验者更好地分析采集系统的性能; 对于温度采集系统来说, 粗大误差也总是存在的, 为了剔除在温度采集系统中出现的粗大误差, 文章首先介绍了 3 种常见的粗大误差剔除方法, 包括格拉布斯准则、迪克逊准则和拉伊达准则; 比较了拉伊达准则和改进拉伊达准则的优劣, 选取了改进拉伊达准则, 将其应用于基于 Labview 设计的温度采集系统当中; 整个系统不仅实现了对温度的自动化采集和处理, 而且也实现了对非正态分布的数据进行粗大误差剔除, 并且剔除的效果满足预期。

关键词: LabVIEW; 改进拉伊达准则; 粗大误差

Application of Pauta Criterion in Temperature Acquisition System Based on LabVIEW

Li Rui¹, Wang Huizhong², Feng Jianwen¹

(1. Computer Department, Hangzhou Dianzi University, Hangzhou 310018, China;
2. Zhejiang Academy of Metrology, Hangzhou 310016, China)

Abstract: The collection of temperature is one of the prerequisites for temperature control systems, a suitable acquisition system will reflect the real changes of the temperature, and reliable data can provide a strong guarantee for subsequent data analysis. Gross error is modal value in acquisition system, effective removal of these gross error can help the experimenter to better analyze the performance of the acquisition system. In this paper, we first introduces three common gross error elimination methods, including Grabs criteria, Dixon guidelines and Pauta Criterion. Comparing the merits of Pauta Criterion and improved Pauta Criterion, and apply the improved Pauta criterion to LabVIEW temperature acquisition system, and this system not only realized the automatic collection and compute, but also got rid of the coarse error of the non-normal distribution data successfully, and reached expected goal.

Keywords: LabVIEW; improved pauta criterion; gross error

0 引言

在常见的数据采集系统中, 采集到的数据通常包含较大的误差, 误差虽然是被允许的, 但是粗大误差(也称为野值)的出现会大大影响测量结果的准确性, 所以在测量得到一系列数据之后需要对粗大误差进行剔除, 来保证测量结果的准确性。粗大误差的出现通常是因为检测环境受到外部的干扰或者是实验者的粗心大意引起的。常见的粗大误差剔除方法有格拉布斯准则、迪克逊准则、拉伊达准则等^[1]。格拉布斯准则和迪克逊准则适用于测量样本数较小时使用、计算公式复杂, 在使用时需要查表; 拉伊达准则由变化的贝塞尔公式推导而来, 使用时简单方便, 适用于测量样本数较多的场景。在本系统中, 温度是一种迟滞性的缓变量, 随着不断加热, 温度会不断线性缓慢上升直到目标温度, 为了实时准确地监控温度的变化, 要频繁的对整个过程进行数据采集, 所以拉伊达准则在本系统中比较

适用。

LabVIEW 是 Laboratory Virtual Instrument Engineering Workbench 的缩写, 是一款使用图形符号来编写程序的编程环境, 具有开发周期短、界面美观大方、灵活方便、跨平台等特性^[2]。LabVIEW 编程开发环境与常见的 C、Java 等开发环境有一个重要的区别是: LabVIEW 是使用图形的编程语言, 而不是使用文本进行编辑。一个 VI 界面可以分为前面板和程序框图两部分, 前面板是 VI 的交互式图形用户界面, 模拟了物理仪器的面板, 可以有按钮、温度计、波形等显示控件和输入控件, 且这些控件还能与 Photoshop 等图片优化软件结合使用, 使界面更加美观大方^[3]。程序框图是 VI 的源代码, 类似于面向对象语言的类, 由 LabVIEW 的图形化编程即 G 语言构成, 是实际可执行的程序。程序框图由子 VI、库函数、控制结构和常量、对象等构成, 用连线将对象连接起来定义它们之间的数据流。对象对应着前面板的终端, 这样就可以实现用户和后台程序的交互^[4]。由于 LabVIEW 的以上特性, 所以 LabVIEW 具有开发周期短, 适合于实验研究场景等优点^[5]。

1 常见粗大误差剔除方法

剔除粗大误差的常见方法有格拉布斯准则, 迪克逊准则、拉伊达准则等^[6], 这几种准则的应用都是以检测样本服从正太

收稿日期: 2017-10-19; 修回日期: 2017-11-14。

作者简介: 李蕊(1994-), 女, 四川南充人, 硕士研究生, 主要从事嵌入式应用技术方向的研究。

冯建文(1971-), 女, 河南汝阳人, 副教授, 主要从事智能控制方向的研究。

分布为前提的，以下内容将进行简单介绍。

1.1 格拉布斯准则

使用格拉布斯准则对测量样本 T 进行粗大误差剔除的步骤如下：

- 1) 计算升序数列的平均值 μ 和标准差 σ ；
- 2) 以平均值 μ 为基准计算计算可疑数据的偏离值 V_i ，其中偏离值最大的为 $|V_i|_{\max}$ ；
- 3) 根据查表所得的置信概率 α 和测量次数 n ，确定临界统计系数 $g(\alpha, n)$ ；
- 4) 若 $|V_i|_{\max} > g(\alpha, n)$ ，则认为 x_i 为粗大误差数据，应当剔除，否则认为序列 T 中不含粗大误差数据，剔除完毕；
- 5) 剔除某一粗大误差后继续执行 1) ~ 5)，直到没有粗大误差为止。

以上步骤以流程如图 1 所示。

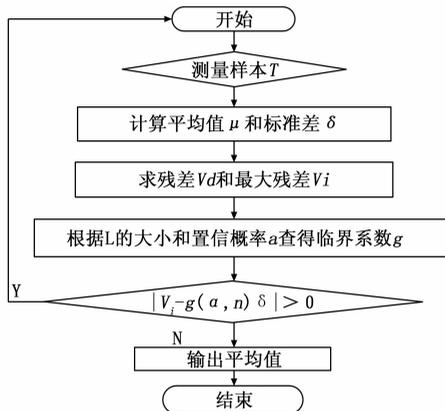


图 1 格拉布斯准则流程图

1.2 迪克逊准则

迪克逊准则是一种利用极差比双侧检验来判别粗大误差的方法，可以不用计算采集数据的平均值、标准差和残差，简化了计算过程，但是需要查表^[7]。这种方法对于序列中只有一个粗大误差值的测量数据效果较好，但是对于在同侧存在多个粗大误差值时检测效果不佳。其检测步骤如下：

- 1) 将采集数据规模为 n 的数据从小到大进行排序^[2]形成待检测序列 $T: t_1 < t_2 < t_3 < \dots < t_i < \dots < t_n$ ；
- 2) 对待检测序列观察待检测值是最大值还是最小值，如果是最大值，则构造统计量：

$$y_{22} = \frac{t_n - t_{n-2}}{t_n - t_2} \quad (1)$$

如待检测值是最小值，则构造统计量：

$$y'_{22} = \frac{t_n - t_2}{t_{n-2} - t_2} \quad (2)$$

- 3) 查看迪克逊准则给出的这些统计量的分布函数以及临界值 $D(\alpha, n)$ ^[6]，则：

$$P(y_{22} > D(\alpha, n)) = \alpha \quad (3)$$

y_{22} 和 y'_{22} 临界值可以查看文献 [8]。

- 4) 根据迪克逊的判别公式，若统计量 $y_{22} > D(\alpha, n)$ 或者 $y'_{22} > D(\alpha, n)$ ，则认为该值为粗大误差值，应该剔除，否则保留该值。

1.3 标准拉伊达准则

拉伊达准则是一种依靠测量数据三倍标准差为极限标准取

舍测量值的方法，其给定的置信率是 99.73%，是一种适用于测量次数较多 ($n > 15$) 的场景或者是经过大量重复测量已经统计出标准差 σ 的情况。同时，在文献 [9] 中，作者提出了一种基于拉伊达准则进行目标值预测的方法，发现将拉伊达准则用于测量数据较多的目标预测场景时也能得到较好的结果。使用拉伊达准则对规模为 n 的测量数据 T 进行误差剔除的一般步骤如下：

- 1) 计算测量数据的均值 \bar{T} ：

$$\bar{L} = \sum_{i=0}^n t_i \quad (4)$$

- 2) 根据均值求出残差：

$$V_i = l_i - \bar{L} \quad (5)$$

- 3) 根据残差求出标准偏差：

$$\sigma = \sqrt{\frac{1}{n-1} V_i^2} \quad (6)$$

- 4) 如 $|l_i - \bar{L}| > 3\sigma$ ，则 l_i 为粗大误差值，应剔除；否则，应保留。

- 5) 计算 $n-1$ 个测量数据的均值、残差、标准偏差继续执行 (1) ~ (4)，直至没有粗大误差时停止。

1.4 改进的拉伊达准则

从上面标准拉伊达准则的讨论中可以看到在使用该准则进行剔除粗大误差时，每剔除一个误差后就要重新计算一次 $n-1$ 个数的平均值、残差和标准差。这大大增加了在应用时的计算量，会导致 LabVIEW 程序框图在设计时变得非常繁复。为了解决这一问题，提出一种直接用测量数据 t_i 来计算标准偏差 σ ，用某个固定值 t_0 (接近于均值) 稍加修正来计算平均值 \bar{T} 的方法^[10]。在该方法中，

$$\delta_i = t_i - t_0 \quad (7)$$

平均值：

$$\bar{T} = \frac{1}{n} \sum_{i=1}^{i=n} t_i = t_0 + \frac{1}{n} \sum_{i=1}^{i=n} \delta_i \quad (8)$$

残差：

$$|V_i| = |t_i - \bar{T}| \quad (9)$$

推导变化的贝塞尔公式可以得出的标准偏差如下：

$$\sigma = \sqrt{\frac{1}{n-1} \left[\sum_{i=1}^n \delta_i^2 - \frac{1}{n} \left(\sum_{i=1}^n \delta_i \right)^2 \right]} \quad (10)$$

可以发现，通过以上四式去剔除粗大误差时，可以一次计算出 δ_i 数组和 δ_i^2 数组后，不必每次循环时重复计算原始测量数据的累加值，只需要用 $\sum_{i=1}^n \delta_i$ 和 $\sum_{i=1}^n \delta_i^2$ 分别减去第 i 个 δ_i 和 δ_i^2 即可，从而达到节省运行时间的目的。

2 系统设计与实现

2.1 系统基本组成

一般温度采集系统由采集对象、测温热电阻作为传感器、温度采集回路和温度采集软件组成^[11]。

在本实验系统中，采集对象是 RTS-35A 恒温水箱，RTS-35A 恒温水箱是一款可按照恒定功率进行加热或者降温的设备，其升温 and 降温过程一般是按照某一斜率线性上升或者线性下降。测温热电阻是温度系统中重要的传感器，将被测对象的温度以电阻的形式反应出来，然后交由温度采集回路进行处理。温度采集回路是将测温热电阻的值传送到温度采集软件进行分析和处理，在本系统中，温度采集回路使用的是 Agilent34970A 进行采集，Agilent34970A 的精度为六位半，有精

确的测量能力, 可以应用于各种场合。Agilent34970A 使用两种标准通讯接口: GPIB (IEEE-488)、RS232, GPIB 接口具有通讯质量高性能稳定等优点, 但是 PC 机没有 GPIB 接口, 而且 GPIB 接口通信距离短等缺点使得 GPIB 接口不适用本系统, 本系统采用 RS232 接口协议进行通信。温度采集软件使用虚拟仪器语言 Labview 来实现^[12]。本系统结构如图 2 所示。

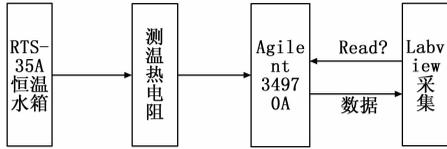


图 2 温度采集系统结构图

2.2 系统采集详细设计

在本系统中, 利用 Agilent34970A 六位半采集仪来对采集到的电阻值进行转换, 再上传到上位机 Labview 程序中, Agilent34970A 与上位机的通讯是采用 SCPI (可编程仪器控制命令) 进行, 上位机发送 “read?” 指令给采集仪, 采集仪就会返回所有通道的数据。Labview 上位机采集数据使用 VISA 串口通信方式, VISA 是应用于仪器编程标准 I/O 应用程序接口 (API), 是可以调用底层驱动器的 API, Labview 在使用的时候也要安装相应的驱动程序^[13]。

采集数据可以分为五步, 第一步上位机写入 “read?” 指令到端口; 第二步是等待延时时间保证能收到缓冲区里面正确的数据; 第三步是从端口中读出数据; 第四步是对读出的数据进行处理和显示; 第五步采集结束后将采集到数据导入到 excel 文档中。其程序如图 3 所示。

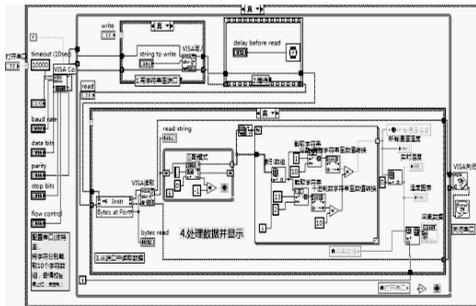


图 3 采集程序程序框图

2.3 利用改进拉伊达准则进行数据处理

图 4 是利用改进拉伊达准则进行野值剔除的程序框图, 程序在实现的过程中先去 Excel 表格中读取测量样本, 存到原始数据数组中。取到测量样本后, 对原始数据数组利用改进拉伊达准则进行处理。

处理步骤可以分为 5 个部分:

- 1) 对原始数据进行累加得到累加值, 初始化 $t_i - t_0$ 和 $(t_i - t_0)^2$ 数组;
- 2) 利用累加值求出算术平均值, 在本实验中, 经过多次取值计算发现初始值 t_0 选择为平均值时能得到最好的结果;
- 3) 求得 $t_i - t_0$ 和 $(t_i - t_0)^2$ 数组;
- 4) 利用 $t_i - t_0$ 和 $(t_i - t_0)^2$ 数组分别求出 $t_i - t_0$ 和 $(t_i - t_0)^2$ 数组的和放在临时变量 1 和临时变量 2 中; 图 4 为利用改进拉伊达准则剔除粗大误差的前四步程序框图。

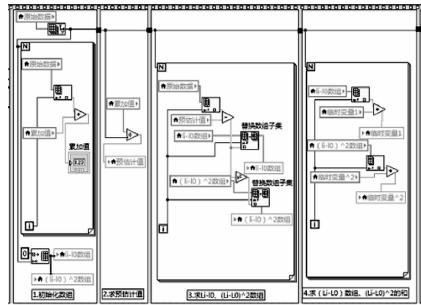


图 4 改进拉伊达准则剔除粗大误差程序框图

5) 设置一个 while loop 循环, 在循环体里面利用前 4 步求得的临时变量对测量样本进行粗大误差剔除, 首先根据临时变量 1、临时变量 2、估计均值求出平均值和标准偏差的 3 倍, 并把剔除完成标记置为 0, 接下来利用一个 for loop 循环对数据进行实际操作, 利用原始数据减去平均值和标准差的 3 倍进行比较看是否是粗大误差值, 如果该值为粗大误差值, 粗大误差数+1, 临时变量 1 和临时变量 2 分别减去 $t_i - t_0$ 数组、 $(t_i - t_0)^2$ 数组中位置为 i 的元素, 并把原始数组位置为 i 的元素替换为平均值, 并把剔除完成标记置为 1, 表示剔除没有完成, 还要进行下一轮的剔除, 直到剔除完成为止。图 5 是利用改进拉伊达准则剔除粗大误差第五步程序框图。

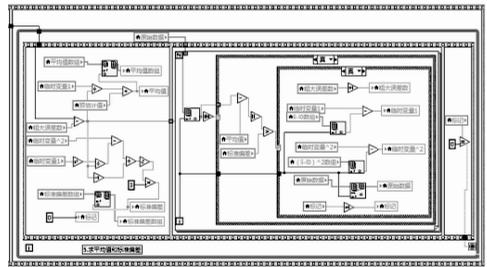


图 5 改进拉伊达准则剔除粗大误差程序框图

对采集对象恒温水箱进行设置, 采集了从 30 °C 上升到 80 °C 的温度, 每隔 500 ms 采集一次, 得到了数据每个时间间隔温度上升为 0.061 °C 左右, 对其中的数据按每 100 个提取一个的方法进行提取得到表 1 的数据作为原始数据。

表 1 采集原始数据 °C

31.749	32.349	32.987	33.57	34.176	34.785
35.389	36.178	36.781	37.385	37.988	38.583
39.046	39.173	39.768	40.363	40.951	41.54
42.34	42.934	43.522	44.144	44.704	45.293
45.881	46.465	47.055	47.637	48.219	48.804
49.386	49.994	50.575	51.156	51.737	52.31
52.89	53.465	54.077	54.84	55.24	55.815
56.387	56.962	57.526	58.093	58.665	59.233
59.799	60.362	60.974	61.54	62.337	62.935
63.532	64.005	64.58	65.156	65.741	66.353
66.918	67.682	68.264	68.814	69.246	69.801
70.369	70.972	71.559	72.105	72.896	73.434
74.147	74.762	75.36	75.946	76.479	77.012
77.594	78.094	78.696	79.245	79.808	80.366
80.934	81.458	81.985			

在实际应用中，利用拉伊达准则进行数据处理时，要求被测数据稳定在某个值附近，满足正态分布；对于波动特别大的数据，就判定其为粗大误差。在本实验系统中，温度是一种带有迟滞性的研究对象，其单位时间温度上升是一个平滑的过程，对表 1 的数据进行分析，发现数据处于一个基本稳定上升的过程中，原因是恒温加热水箱通过某一恒定功率进行加热，其特点是在单位时间内温度的增量波动不大。可以通过表 1 得到其单位时间内温度增量表表 2，可以看出是符合前面提到的温度单位时间内上升 0.061 ℃ 的要求的，为了进一步的验证改进拉伊达算法在本实验中可行性，在原始数据中加入毛刺数据，共加入 7 个，得到表 3 的数据。对表 3 的数据进行减法分析处理得到单位时间的增量表表 4。

表 2 原始数据增量表 ℃

0.6	0.638	0.583	0.606	0.609	0.604
0.798	0.603	0.604	0.603	0.595	0.59
0.56	0.595	0.595	0.588	0.589	0.8
0.594	0.588	0.592	0.59	0.589	0.588
0.584	0.59	0.582	0.582	0.585	0.582
0.608	0.581	0.581	0.581	0.573	0.58
0.575	0.612	0.595	0.588	0.605	0.572
0.575	0.584	0.567	0.572	0.588	0.586
0.563	0.612	0.596	0.797	0.598	0.588
0.282	0.575	0.616	0.585	0.612	0.565
0.856	0.592	0.55	0.391	0.586	0.568
0.603	0.587	0.546	0.903	0.588	0.602
0.588	0.598	0.586	0.543	0.597	0.6
0.595	0.602	0.569	0.583	0.618	0.598
0.584	0.588				

表 3 毛刺数据表 ℃

31.749	32.349	32.987	33.57	34.176	34.785
30.181	36.178	36.781	37.385	37.988	38.583
39.046	39.173	36.573	40.363	40.951	41.54
44.34	42.934	43.522	44.144	44.704	45.293
45.881	46.465	47.055	45.475	48.219	48.804
49.386	49.994	50.575	51.156	51.737	52.31
52.89	53.465	54.077	54.84	55.24	55.815
56.387	56.962	57.526	55.956	58.665	59.233
57.643	60.362	60.974	61.54	62.337	62.935
63.532	64.005	64.58	65.156	65.741	66.353
66.918	67.682	68.264	68.814	69.246	69.801
70.369	70.972	71.559	72.105	75.718	73.434
74.147	74.762	75.36	75.946	76.479	77.012
77.594	78.094	78.696	79.245	79.808	80.366
80.934	81.458	81.985			

对表 4 的数据通过图 4、5 的 Labview 程序进行处理，共循环两次，发现加入的 7 个毛刺数据为粗大误差值。第一次处理时得到平均值为 0.482221，3 倍的标准偏差值为 2.616408，剔除了第 6、13、68 位毛刺数据；第二次循环得到平均值为 0.541301，3 倍标准偏差值为 2.054999，剔除了第 18、27、45、48 位毛刺数据。至此，毛刺数据剔除完毕，第三次循环时平均值为 0.538886，3 倍标准偏差值为 2.106758，发现无粗大误差，剔除过程结束。

表 4 毛刺数据增量表 ℃

0.6	0.638	0.583	0.606	0.609	-4.604
0.798	0.603	0.604	0.603	0.595	0.59
0.56	-2.595	0.595	0.588	0.589	2.8
0.594	0.588	0.592	0.59	0.589	0.588
0.584	0.59	-1.582	0.582	0.585	0.582
0.608	0.581	0.581	0.581	0.573	0.58
0.575	0.903	0.595	0.588	0.605	0.572
0.575	0.584	-1.567	0.572	0.588	-1.586
0.563	0.612	0.596	0.797	0.598	0.588
0.282	0.575	0.616	0.585	0.612	0.565
0.856	0.592	0.55	0.391	0.586	0.568
0.603	0.587	0.546	3.612	0.588	0.602
0.588	0.598	0.586	0.543	0.597	0.6
0.595	0.602	0.569	0.583	0.618	0.598
0.584	0.588				

3 结论

通过分别比较剔除前和剔除后增量数据表与原始数据表，见图 5、6，可以发现原用于对稳定数据进行粗大误差剔除的拉伊达准则，对于增量数据的粗大误差剔除，也能取得良好的效果，在保证漏剔率和误剔率的同时，还可以保证得到较为准确的实验结果，对于后续的实验分析具有很大的帮助。

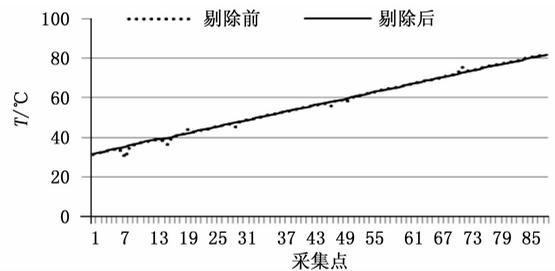


图 5 增量数据剔除前后对比

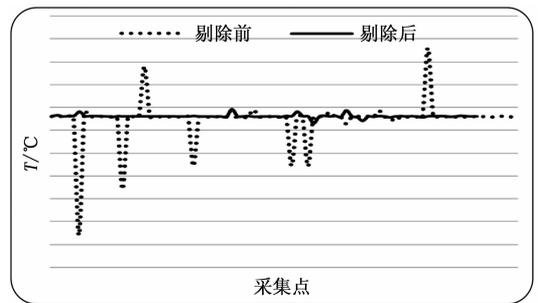


图 6 原始数据剔除前后对比

在工程的实际设计和测量过程中，灵活的运用 LabVIEW 进行开发可以大大的缩短开发周期和节约人力成本，也可以方便地构建直观的虚拟仪器界面。在基于 LabVIEW 设计的采集测量平台中，进一步改进了拉伊达准则，将其应用在稳定持续变化的温度测量场合，扩大了其应用范围。实验表明，改进使用后的拉伊达准则可以准确的对试验产生的粗大误差进行剔除，大大提高了试验数据的准确性，确保测试结果可信性，在后期涉及到多数据的采集中具有良好的实用意义。