

基于云理论和多传感器加权平均的 关键性能指标状态描述

陈 梵, 薛 敏, 葛 蓬, 李俊达

(国电南京自动化有限公司, 江苏 南京 211100)

摘 要: 利用了云理论以及多传感器加权平均的方法来对关键性能指标的状态进行定性描述. 第一步用分布图法对状态正常情况下由数个传感器得到的数据进行一致性检验, 第二步采用多传感器加权平均方法并且融合出正常的状态下均方误差为最小的数据融合值, 第三步将任意时刻的关键性能指标数据和均方误差为最小的数据融合值相比较得出偏差值, 最后一步基于云理论构造出 5 个评语的评语集, 和结合偏差大小对当前的关键的性能指标进行了定性描述. 实验结果表明该方法的有效性.

关键词: 一致性检验; 多传感器加权平均; 云理论; 关键性能指标

中图分类号: TP391

文献标识码: A

文章编号: 1000-7180(2012)01-0159-05

Key Index State Description Based on Cloud Theory and Multi-sensor Weighted Mean

CHEN Fan, XUE Min, GE Peng, LI Jun-da

(Guodian Nanjing Automation Co. Ltd., Nanjing 211100, China)

Abstract: Key index analysis state was described qualitatively by using cloud theory and multi-sensor weighted mean method in this paper. Firstly, the data got from the sensors in natural state were tested conformably by using distributing graph method. Secondly, we can use the multi-sensor weighted mean method and get the smallest data confluent value of the average error in natural state. And then compared the key index data in any time with foregoing confluent value and the warp value was gained. At last we can construct five remark sets based on cloud theory, and combining the warp value we can describe the key index analysis state qualitatively. The experiment result indicates the validity of our method.

Key words: consistency testing; multi-sensor weighted mean; cloud theory; key capability index

1 引言

通信和电子计算机不断发展, 系统的安全和关键服务方面的要求不断提高. 海量事实证明, 很难实现绝对安全的系统, 并且在非常多的工作环境下, 系统的关键服务等能否正常完成更被受到关注. 也就是说, 对系统的综合性能指标关心以和对某些关键性能指标的状态都受到了高度得重视^[1]. 本文所述关键性能指标是指对于具体的某个系统, 为了保证其关键服务的正常工作前提下, 还要保证状态良好

的一个或者几个性能指标. 所以研究关键性能指标具有现实意义, 可以良好的补充以往系统安全状态判断.

为了克服单一传感器采集数据的不足, 本论文利用多传感器采集数据来实现单一传感器发生错误和故障, 从而继续保证数据的完整性和可靠性. 在考虑到了关键性能指标数据的随机性后, 第一步对正常状态下由多个传感器得到的数据用分布图法进行一致性检验, 第二步采用文献[2]提出得使用多个传感器加权平均方法来融合出正常状态下的均方误差

最小的数据融合值 $\hat{X}$, 第三步将这个融合值 $\hat{X}$ 和任意时刻 t 得到的关键性能指标数据 x_t 来比较, 得出偏差值 $b = |x_t - \hat{X}| / \max(x_t, \hat{X})$. 基于云理论构造 5 个评语得评语集 $V = (v_1, v_2, \dots, v_5)$, 用来表示关键性能指标的当前所处状态, 这些状态包括非常好、良好、中等、差、很差 5 个评语. 最后一步基于云理论与偏差的大小来对当前得关键性能指标进行定性语言得描述. 与过去相比, 更准确、更科学的定量判断. 本文通过使用 Matlab 仿真结果可以看到, 本文所提出的方法确实具有较好得可行性.

2 分布图法利用于对关键性能指标数据进行一致性得检验

现有系统之中, 不存在任何随机性产生得理想状态. 总是存在某些数据在各种各样的因素影响之下而混杂在正常的数据之中, 所以要非常好的反映系统的统计规律^[3]. 本文就要把这些数据在融合前删除. 本文采用了分布图法来对数据的一致性检验进行实现. 本文在一个分布图之中反映出数据分布结构的参数包括一下几个, 上分位数 x_U 、中位数 x_M 、分位数离散度 d_F 和下分位数 x_L .

设得到一组数据. 将这组数据按照从小到大的顺序排列:

(1) 当数据得数目是偶数时: $x_1, x_2, \dots, x_n, x_{n+1}, \dots, x_{2n}$, 则将 x_1 称为测量列下限, x_{2n} 称之为上限.

定义了中位数 x_M 为:

$$x_M = \frac{x_n + x_{n+1}}{2}$$

上分位的区间 $[x_M, x_{2n}]$ 中位数为 x_U , 下分位的区间 $[x_1, x_M]$ 中位数为 x_L .

① 如果 n 为偶数, 那么

$$x_U = x_{3n/2}; x_L = x_{(n+2)/2}$$

分数位的离散度为 d_F 为:

$$d_F = x_U - x_L = x_{3n/2} - x_{(n+2)/2}$$

② 如果 n 为奇数, 那么

$$x_U = (x_{(3n-1)/2} + x_{(3n+1)/2})/2$$

$$x_L = (x_{(n+1)/2} + x_{(n+3)/2})/2$$

分数位离散度为 d_F 为:

$$d_F = x_U - x_L = (x_{(3n-1)/2} + x_{(3n+1)/2})/2 - (x_{(n+1)/2} + x_{(n+3)/2})/2$$

(2) 如果数据的数目是奇数时: $x_1, x_2, \dots, x_n, x_{n+1}, \dots, x_{2n}, x_{2n+1}$, 那么 x_1 称为测量列的下限,

x_{2n+1} 称为上限.

若中位数 x_M 为:

$$x_M = x_{n+1}$$

上分位的区间 $[x_M, x_{2n+1}]$ 中位数为 x_U , 下分位的区间 $[x_1, x_M]$ 中位数为 x_L .

① 如果 n 为偶数, 那么

$$x_U = x_{(3n+2)/2}; x_L = x_{(n+2)/2}$$

分数位离散度是 d_F 为:

$$d_F = x_U - x_L = x_{(3n+2)/2} - x_{(n+2)/2}$$

② 如果 n 为奇数, 那么

$$x_U = (x_{(3n+1)/2} + x_{(3n+3)/2})/2$$

$$x_L = (x_{(n+1)/2} + x_{(n+3)/2})/2$$

分数位离散度是 d_F 为:

$$d_F = x_U - x_L = (x_{(3n+1)/2} + x_{(3n+3)/2})/2 - (x_{(n+1)/2} + x_{(n+3)/2})/2$$

认定与中位数的距离大于 αd_F 数据为不正常数据, 就是说无效数据的判断区间是:

$$|x_i - x_M| > \alpha d_F$$

式中, α 为常数. 它的大小根据系统得测量要求定, 一般取 0.6, 1.1, 2.1 等值.

测量列去除了误差数据之后, 其余数据被认为是有效得一致性测量数据. 随后对一致性数据进行数据融合的处理.

3 多传感器的加权平均数据融合算法

3.1 在某一时刻求估计值

假设要估计得真值是 X , n 个传感器得测量值为 $X_1, X_2, \dots, X_n$, 彼此相互独立, 而且是 X 的无偏估计. 传感器方差为 $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$. 各传感器得加权

因子为 $W_1, W_2, \dots, W_n$, 那么融合后的 $\hat{X}$ 值:

$$\hat{X} = \sum_{i=1}^n w_i x_i; \text{其中 } \sum_{i=1}^n w_i = 1$$

总均方误差为:

$$\sigma^2 = E[(x - \bar{x})^2] = E\left[\sum_{i=1}^n w_i^2 (x - x_i)^2 + 2 \sum_{i=1, j=1, i \neq j}^n w_i w_j (x - x_i)(x - x_j)\right]$$

因为 $X_1, X_2, \dots, X_n$ 彼此是相互独立, 而且为 X 的无偏估计, 所以 $E[(X - X_p)(X - X_q)] = 0 (p \neq q; p = 1, 2, \dots, n; q = 1, 2, \dots, n)$, 故 σ^2 可写成:

$$\sigma^2 = E\left[\sum_{i=1}^n w_i^2 (x - x_i)^2\right] = \sum_{i=1}^n w_i^2 \sigma_i^2$$

利用柯西不等式: $\left(\sum_{i=1}^n w_i^2 \sigma_i^2\right) \left(\sum_{i=1}^n 1/\sigma_i^2\right) \geq$

$(\sum_{i=1}^n w_i)^2$, 当 $w_1^2 \sigma_1^2 = w_2^2 \sigma_2^2 = \dots = w_n^2 \sigma_n^2$ 时候等号成立, 总得均方差取最小值。

这个最小值得求取是当加权因子 $w_i, i=1, 2, \dots, n$ 能满足式 $\sum_{i=1}^n w_i = 1$ 约束条件时, 利用总的均方差极值求取。

根据多元函数求极值理论, 可以求出总均方误差最小时所对应的加权因子为:

$$W'_i = 1 / \left(\sigma_i^2 \sum_{i=1}^n (1/\sigma_i^2) \right)$$

这时所对应最小均方误差为:

$$\sigma_{\min}^2 = 1 / \sum_{i=1}^n (1/\sigma_i^2)$$

3.2 在某一段时间 t 内求估计值

当估计真值 X 是常量时, 那么根据各个传感器在一段时间 t 内数据均值来进行估计。设在 t 时间内, 当第 i 个传感器共测得 k 个数据 $x_j, j=1, 2, \dots, k$; 则第 i 个传感器在 t 时间内数据的均值为:

$$\bar{x}_i = (1/k) \sum_{j=1}^k x_j$$

这时的估计值为:

$$\hat{X} = \sum_{i=1}^n w_i \bar{x}_i$$

总均方误差为:

$$\begin{aligned} \bar{\sigma}^2 &= E[(x - \bar{x})^2] = E \left[\sum_{i=1}^n w_i^2 (x - x_i)^2 \right. \\ &\quad \left. + 2 \sum_{i=1, j=1, i \neq j}^n w_i w_j (x - x_i)(x - x_j) \right] \end{aligned}$$

同理, 因为 $X_1, X_2, \dots, X_n$ 为 X 的无偏估计, 所以 $x_1, x_2, \dots, x_n$ 也必定是 X 的无偏估计, 故 $\bar{\sigma}^2$ 为:

$$\bar{\sigma}^2 = E \left[\sum_{i=1}^n w_i^2 (x - x_i)^2 \right] = (1/k) \sum_{i=1}^n w_i^2 \sigma_i^2$$

显然, 当 $\bar{\sigma}^2$ 最小时对应得最优加权因子仍然满足 $W'_i = 1 / \left(\sigma_i^2 \sum_{i=1}^n (1/\sigma_i^2) \right)$, 这时所对应的最小均方

误差为: $\bar{\sigma}_{\min}^2 = 1 / \left(k \sum_{i=1}^n (1/\sigma_i^2) \right) = \sigma_{\min}^2 / k$

则 $\bar{\sigma}_{\min}^2$ 必定小于 $\sigma_{\min}^2$, 而且 $\bar{\sigma}_{\min}^2$ 将随 k 的增加而进一步减小。

3.3 多传感器加权平均融合算法

(1) 分别计算每个传感器数据平均值:

$$\bar{x}_i = (1/k) \sum_{j=1}^k x_j$$

(2) 分别计算出每组的方差: σ_i^2

(3) 利用前面求出的值来计算最优加权因子和总得均方差:

$$W'_i = 1 / \left(\sigma_i^2 \sum_{i=1}^n (1/\sigma_i^2) \right);$$

$$\sigma_{\min}^2 = 1 / \sum_{i=1}^n (1/\sigma_i^2)$$

(4) 求出加权数据融合值: $\hat{X} = \sum_{i=1}^n w_i x_i$

(5) 计算对应的最小方差:

$$\bar{\sigma}_{\min}^2 = 1 / \left(k \sum_{i=1}^n (1/\sigma_i^2) \right) = \sigma_{\min}^2 / k$$

4 运用云模型实现关键性能指标的定性描述

云模型的定义是用语言值描述得某个定性概念和其数值表示之间得不确定性转换模型。是用来统一量化数值间和语言值的模糊性和随机性的^[4-5]。相当多的云滴组成云, 而云的整体反映了一个定性概念。对应的数值域作为函数概念中的自变量。而云的数字特征使用期望值 Ex (Expected Value)、超熵 He (Hyper Entropy)、熵 En (Entropy) 这三个数值来表征^[6]。超熵 He , 它反映了云滴的离散程度, 超熵的大小可以间接地反映出云滴厚度, 超熵 Ex 如果越大, 那么隶属度的随机性越大, 云滴离散度越大, 云的厚度也就越大。对于期望值 Ex 是概念, 是指在论域中的中心值, 它是最能代表这个定性概念的值。也就是说, 它 100% 得隶属于这个定性概念, 完整的反映了相应得模糊概念的信息中心值。而定性概念模糊度的度量用熵 En 表示, 它反映了在论域中能够被这个概念所接受的数值范围, 集中体现了定性概念亦彼亦此性的裕度。熵 Ex 越大, 则概念所接受的数值范围越大, 概念也就越模糊。云模型的这三个数字特征值把随机性(隶属度的随机性)和模糊性(定性概念的亦彼亦此性)集成到一起, 从而构成定量和定性相互间的映射, 反映出了定性知识的定量特性。

在描述系统关键性能指标上, 云模型比传统的定量描述具有更高地科学性和直观性^[6]。按照求得正常状态下均方误差最小的数据融合值 $\hat{X}$, 和任意时刻 t 得到的关键性能指标数据 x_t , 来比较得到偏差值 $a = |x_t - \hat{X}| / \max(x_t, \hat{X})$ 。定义出 5 个评语的评语集 $V = (v_1, v_2, \dots, v_5)$, 并且将定义得 5 个评语放置于连续的语言值标尺上, 每个评语都用云模型来实现, 从而构成一个定性评测云发生器, 如图 1 所示。

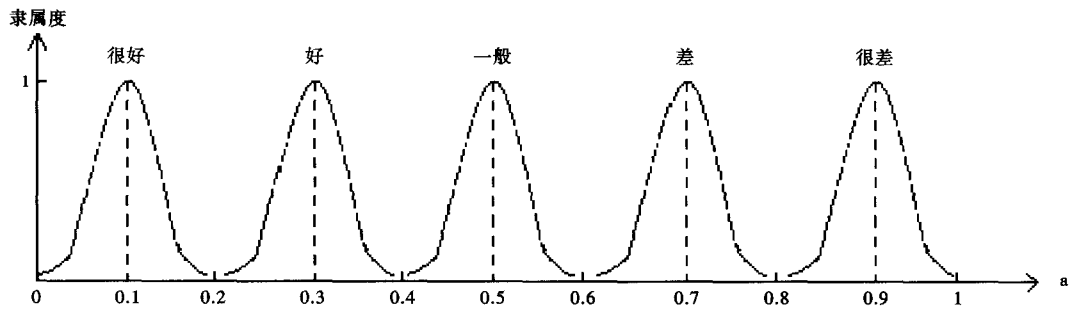


图 1 定性评测云发生器标尺表示

5 基于分布图法、云模型、和多传感器的加权平均方法的 Matlab 仿真试验实例

采集一组数据,第一步用分布图法对数据进行一致性的检验,并且将异常的数据设置为 0,第二步通过多传感器加权平均方法求出加权得数据融合值和其对应的最小方差,最后一步通过基于云理论创建的 5 评语定性评测云发生器来给出定性描述.采集到的数据如表 1、表 2 所示.

表 1 采集到的数据(一)

0.1474	0.0179	0.0797	0.1498	0.1579
0.1483	0.2525	0.2435	0.1523	0.1748
0.2045	0.3120	0.2484	0.1817	0.2057
0.2938	0.3050	0.2566	0.2823	0.2171
0.2991	0.3622	0.2637	0.3328	0.2612
0.3581	0.3847	0.2629	0.3347	0.3162
0.3662	0.4036	0.3138	0.3975	0.3535
0.3857	0.4076	0.3452	0.4179	0.3675
0.3932	0.4393	0.3883	0.4431	0.3863
0.5539	0.4603	0.5175	0.4653	0.4192

表 2 采集到的数据(二)

0.0503	0.1752	0.0607	0.1195	0.0266
0.1281	0.2119	0.0832	0.2405	0.0594
0.2353	0.2721	0.0994	0.2715	0.1425
0.2942	0.2797	0.1965	0.2978	0.1637
0.3184	0.2856	0.3446	0.4023	0.1772
0.3265	0.3085	0.3857	0.4022	0.2961
0.3436	0.3173	0.4073	0.4065	0.3389
0.3497	0.4084	0.4288	0.4225	0.3396
0.4789	0.4608	0.4884	0.4782	0.3703
0.4781	0.6270	0.5432	0.4933	0.5113

删除质量不好的数据后如表 3、表 4 所示.

表 3 处理后的数据(一)

0	0.1721	0.0607	0	0
0	0.2128	0.0837	0	0.0595
0.2353	0.2722	0.0995	0.2766	0.1416
0.2942	0.2794	0.1963	0.2974	0.1613
0.3187	0.2852	0.3443	0.4004	0.1793
0.3260	0.3087	0.3856	0.4036	0.2912
0.3430	0.3172	0.4072	0.4067	0.3320
0.3492	0.4087	0.4284	0.4225	0.3383
0	0	0.4882	0.4785	0.3794
0	0	0.5436	0.4937	0

表 4 处理后的数据(二)

0.1483	0	0	0.1478	0.1585
0.1485	0	0.2431	0.1522	0.1763
0.2036	0.3001	0.2472	0.1888	0.2043
0.2927	0.3061	0.2553	0.2802	0.2165
0.2972	0.3645	0.2615	0.3328	0.2606
0.3573	0.3865	0.2645	0.3347	0.3133
0.3764	0.4012	0.3176	0.3937	0.3584
0.3886	0.4095	0.3471	0.4189	0.3666
0.3921	0.4344	0	0.4421	0.3857
0	0.4642	0	0.4613	0.4178

进行数据一致性检测的融合数据为:

$\hat{X}=0.3116773414134638463230375853746$

那么其方差值是:

$0.12169577852720253465535298050930e-3$

取两个时刻的数值随机进行一下比较.

如果某一时刻 t_1 测得的数据 $x_1=0.665$, 时刻 t_2 测得的数据 $x_2=0.224$, 然后代入公式 $a=|x_i-\hat{X}|/\max(x_i,\hat{X})$ 求得 $a_1=0.529, a_2=0.324$ 定性评测云发生器评测后, 可得到时刻 t_1 的关键性能指标定性描述为普通, 时刻 t_2 的关键性能指标定性描述是好。

从以上实验结果可以得出, 本文设计的基于云模型的和多传感器的加权平均方法关键性能指标状态描述方法确实是高效的和可行的。

6 结束语

分析关键性能指标可以帮助我们更好更全面的了解系统当前的状态, 而且有利于关键服务正常工作的保证。采用了分布图法的数据一致性检验, 使得数据统计的准确性有所提高, 融和数据的误差有所减少。单个传感器不稳定、易出错等不足可以通过多传感器加权平均方法来克服。相比过去的定量判断, 基于云理论的定性语言描述更加科学、准确和直观。本论文还存在着很多不足, 有待于进一步的研究定性评测云发生器得区间划分。

参考文献:

[1] 付华, 杜晓坤. 基于 Bayes 估计理论的数据融合方法

[J]. 自动化技术与应用, 2005, 24(4): 10-12.

[2] 翟翌立, 戴逸松. 多传感器数据自适应加权融合估计算法的研究[J]. 计量学报, 1998, 19(1): 69-75.

[3] 李德毅, 刘常昱. 论正态云模型的普适性[J]. 中国工程科学, 2004, (08): 29-34.

[4] Li D Y, Cheng D W, Shi X M, et al. Uncertainty reasoning based on cloud models in controllers [J]. Computers and Mathematics with Applications, 1998, 35(3): 99-123.

[5] 邓自立. 两种最优观测融合方法的功能等价性[J]. 控制理论与应用, 2006, 23(2): 319-323.

[6] 胡振涛, 刘先省. 基于动态加权的分布式多传感器航迹融合算法[J]. 计算机应用研究, 2006, (06): 59-61.

作者简介:



陈 梵 男, (1980—), 硕士, 工程师。研究方向为计算机及其应用。

薛 敏 女, (1966—), 学士, 高级工程师。研究方向电力系统自动化。

(上接第 158 页)

作者简介:

陈泰伟 男, (1986—), 硕士研究生。研究方向为 RFID 无线射频、信息系统。

项 颖 男, (1968—), 博士, 教授, 高级工程师, 硕士研究生导师。研究方向为基于液晶的平板显示器和光存储器件、基于超薄膜垂直磁化的高密度磁存储器件、电信企业数据仓库

的规划和设计、计算机网络的规划和网络安全。

吴黎明 男, (1964—), 教授, 硕士研究生导师。研究方向为计算机应用及数字图像处理技术。

吴学文 男, (1985—), 硕士研究生。研究方向为嵌入式系统开发。