

基于动态层次聚类分析的多传感器一致性融合算法

周珍娟¹, 刁联旺²

(1. 江苏开放大学信息与机电工程学院, 江苏 南京 211135)

(2. 中国电子科技集团第二十八研究所信息系统工程重点实验室, 江苏 南京 210007)

[摘要] 为了提高集中式多传感器一致性数据融合的精度, 在数据融合过程中, 必须考虑从各传感器获得的数据可信度. 为了克服现有一致性数据融合算法中定义的距离矩阵和关系矩阵存在的不对称性和主观性等缺点, 本文定义了概率距离矩阵和基于统计置信度的关系矩阵, 然后讨论了多正态分布共同均值的极大似然估计的统计性质和递推特性, 并根据估计量的方差统计性质提出了一种基于动态层次聚类的多传感器一致性数据融合处理的方法, 数据实验的计算结果表明, 该算法优于现有的多传感器一致性数据融合方法.

[关键词] 层次聚类, 数据融合, 置信度, 关系矩阵

[中图分类号] TN911.2 **[文献标志码]** A **[文章编号]** 1001-4616(2018)04-0087-06

Consensus Multi-sensor Data Fusion Algorithm Based on Dynamic Hierarchical Clustering Analysis

Zhou Zhenjuan¹, Diao Lianwang²

(1. College of Electronic and Information Engineering, Jiangsu Open University, Nanjing 211135, China)

(2. S & T on Information System Engineering Lab., The 28th Research Institute of CETC, Nanjing 210007, China)

Abstract: To improve the precision of multi-sensor data fusion method, in the procedure of data fusion, the reliability of sensor data must be considered. The shortcomings of the existing consensus multi-sensor data fusion algorithms, such as asymmetry of distance matrix, and subjectivity of connection matrix are discussed, and a new confidence distance is defined. The common mean of some normal distributions are estimated by maximum likelihood method, and its statistical and recursive characters are derived. A consensus multi-sensor data fusion algorithm based on dynamic hierarchical clustering is proposed. The example shows that this method is better than the existing methods.

Key words: hierarchical clustering, data fusion, degree of confidence, connection matrix

数据融合是把来自多种或多个传感器的信息 and 数据进行综合处理以获得更为准确可靠的结果, 从而减少了信息处理过程中可能造成的失误. 为此需要对各个传感器所得数据的真实性和可信性进行综合判定, 建立不同传感器数据之间的信任关系, 从而决定哪些传感器数据参与融合处理. 参考文献[1]对多传感器一致性数据融合算法的发展现状进行了阐述. 参考文献[2]基于传感器精度分级的思想对一致性融合算法进行了改进. 参考文献[3]首先运用模糊理论中相关性函数对各传感器支持度进行计算排序, 再将支持度较低的数据认定为无效数据予以剔除, 最后应用递推最小二乘法对有效数据进行动态融合. 参考文献[4]讨论了在传感器网络数据融合中利用数据的冗余度处理各传感器节点数据的不确定性和不一致性, 并据此提出一种优化的贝叶斯估计多传感器融合方法. 参考文献[5]基于岭估计给出了测量数据可靠性的定量表示方法, 提出改进的最小二乘无偏估计方法. 参考文献[6]运用截尾均值概念处理多传感器测量数据的不一致性, 提出了加权平均的多传感器融合算法. 参考文献[7]基于传感器测量精度构造了支撑函数度量传感器的可靠性及其权重, 形成一种新型加权融合算法. 参考文献[8]基于多传感器的测量数据的样本均值和方法来度量各传感器测量数据的权重, 提出一种概率加权的多传感器一致性数据融合方

收稿日期: 2018-08-15.

基金项目: 江苏省自然科学基金(BK21060148).

通讯联系人: 刁联旺, 博士, 教授, 研究方向: 信息融合. E-mail: diaolw@sina.com

法. 参考文献[9]首先定义一个模糊型指数信任度函数,对两传感器测量数据间的信任程度进行量化处理,并通过信任度矩阵度量各传感器测量数据的综合信任度,以合理地分配各传感器测量数据的权重,从而实现多传感器数据的一致性融合. 参考文献[10]利用多元正态分布的统计假设检验理论,提出了一种新的置信距离度量不同传感器数据间的距离,从而提升了数据一致性的判定精度. 参考文献[11-12]在传感器支持关系矩阵构造过程中引入椭圆曲线来确定传感器之间的相互支持程度,并用关系矩阵的特征值作为传感器测量数据的权重,从而有效地提高融合结果的准确性. 参考文献[13-14]利用 Pitman 接近度准则来度量传感器测量数据的支持度,并用支持度矩阵的特征值作为传感器测量数据的权重,以提高融合结果的精度. 虽然众多国内外学者对一致性数据融合方法进行了大量研究,但总体上讲,这些研究方法在判定传感器数据间的相互连接关系时都是把高精度与低精度传感器测量数据放在一起建立距离矩阵进行比较处理的,这无疑减少了精度差的传感器测量数据参与融合的机会,显然是不利于对实际情况作出客观判定的. 本文针对这一问题,运用层次聚类分析的思想,动态地将传感器节点进行分群处理,不仅可以发现不同传感器测量数据之间的一致性,而且还可以发现不同类构成的“复合传感器”数据间的一致性,从而更能体现传感器数据一致性的本质,更充分地利用传感器测量数据,具有较高的应用价值.

1 现有一致性数据融合方法分析

本节对现有一致性融合方法的基本原理进行简要介绍,并对其存在的主要问题进行分析.

1.1 测量模型

假设有 n 个传感器对某一物体进行测量,其测量模型可用正态分布 $X \sim N(\mu, \sigma_i^2)$ 描述,即:

$$p_i(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{1}{2\sigma_i^2}(x_i - \mu)^2\right\}, \tag{1}$$

式中, $i=1,2,\cdots,n$. x_i 是第 i 个传感器的测量值, μ 为测量特征的真值,而 σ_i 是该传感器测量误差的标准差.

问题是,如何对此 n 个数据的真实性进行判别,找出不同传感器数据之间的相互关系,从而决定哪些传感器数据参与融合处理.

1.2 现有一致性融合方法的基本原理

现有一致性融合方法的基本原理如图 1 所示.

1.3 现有一致性数据融合方法分析

现有一致性数据融合方法的基本原理是利用支持矩阵 $R=(r_{ij})_{n \times n}$ 计算出各传感器被支持的综合程度,然后求出最大传感器连接组,再利用贝叶斯估计、极大似然估计或其他加权方法进行数据融合^[1-4]. 这种处理方式存在如下不足:

(1)最大传感器连接组的选择只考虑了传感器数量最多,并未考虑传感器的精度作用. 这显然是不正确的,例如有 3 个传感器中,尽管两个很差的传感器数据十分一致,且与最好的传感器数据均不一致,这并不能说明两个差传感器(数量多!)的数据融合结果比一个最好的传感器数据更好.

(2)多传感器数据只作一次处理,没有考虑各个传感器组融合处理结果带来的数据精度影响. 这是现有一致性数据融合算法存在的本质缺陷! 例如,多个传感器进行融合处理(比如采用极大似然法)之后,其结果服从正态分布,且该分布的方差比参与的任何一个传感器测量误差的方差均小,这个融合后的结果很可能会与其他传感器(或组)的数据产生新的一致性,从而通过进一步进行一致性融合处理,提高数据的精度,而这种新一致性在现有的一致性融合方法中完全被忽略了!

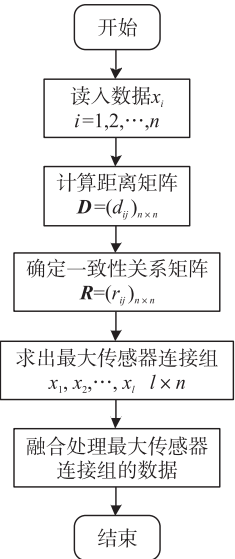


图 1 一致性数据融合方法的基本原理
Fig. 1 The principle of consensus data fusion method

2 基于动态层次聚类的融合算法

基于上节分析,本文利用层次聚类分析的思想以解决不同精度传感器进行一致性判断时存在的信息丢失问题.为此首先定义一种新的度量距离以体现对高精度传感器的偏爱.

2.1 常用统计置信距离

根据数理统计学中的统计假设检验的基本理论,若 $x_i \sim N(\mu, \sigma_i^2)$, $x_j \sim N(\mu, \sigma_j^2)$, 则有:

$$x_i - x_j \sim N(0, \sigma_i^2 + \sigma_j^2). \quad (2)$$

可见,只有在

$$\frac{|x_i - x_j|}{\sqrt{\sigma_i^2 + \sigma_j^2}} \geq k_{1-\alpha/2}, \quad (3)$$

才能拒绝两者之差的均值不为零的假设.式中, $k_{1-\alpha/2}$ 为标准正态分布的随机变量的置信水平为 α 的上门限值^[15].

在多传感器数据融合中,需要对两者获得的数据进行一致性检验,从而避免相差很大(特别是两者的精度相差很大)的数据进行融合.参考文献[10]提出了如下概率距离 d_{ij} 作为第 i 个和第 j 个传感器之间的数据的相互支持程度:

$$d_{ij} = P_r \left(|Z| \leq \frac{|x_i - x_j|}{\sqrt{\sigma_i^2 + \sigma_j^2}} \right), \quad (4)$$

式中, x_i, x_j 是第 i, j 个传感器的测量值; σ_i, σ_j 为相应的方差, Z 表示服从标准正态分布的随机变量.

式(4)定义的概率距离虽然满足对称性的要求,但当两个传感器精度相差很大时,判断两传感器测量数据一致性时数据滤除率不高,对高精度传感器数据的偏向性不够^[11-15],为此本文提出如下改进的概率置信距离.

2.2 一种新的置信距离

本文提出一种新的置信距离定义如下:

$$d_{ij} = P_r \left\{ |Z| \leq \frac{|x_i - x_j|}{\sqrt{2} \min(\sigma_i, \sigma_j)} \right\}. \quad (5)$$

式(5)是对两个传感器 i 和传感器 j ($i \neq j$) 测量数据间距离的一种概率意义上的度量,其含义为

$$P_r(|Z| \leq k_{1-\alpha/2}) = 1 - \alpha, \quad (6)$$

式中, α 为置信水平.它表示一个标准正态随机变量的样本实现(即采样)落入区间 $[-k_{1-\alpha/2}, k_{1-\alpha/2}]$ 的概率为 $1 - \alpha$.因此,置信水平 α 不同,此区间也相应变化,这克服了现有融合方法门限值主观确定的不足^[1,4-5,8].另外,式(5)定义的距离具有如下优点:1)当两个传感器精度相当时,与(4)式相当;2)当两个传感器精度相差很大时,它比式(4)的数据滤除率更高,但比式(2)更宽松一些,能够容纳更多的数据参与融合过程,这实际上体现了对高精度传感器测量数据的偏爱;3)式(5)定义的概率距离矩阵具有对称性.

根据式(5)可以定义概率判定传感器测量数据之间的一致性关系 r_{ij} ,计算方法为^[9,12-14]:

$$r_{ij} = \begin{cases} 1, & d_{ij} \leq 1 - \alpha, \\ 0, & d_{ij} > 1 - \alpha, \end{cases} \quad i, j = 1, 2, \dots, n. \quad (7)$$

$r_{ij} = 1$ 表示两传感器测量数据一致,可以进行融合处理;否则 $r_{ij} = 0$ 表示两传感器测量数据不具有—致性,不宜进行融合处理.

2.3 新算法的基本步骤

本文提出的基于动态层次聚类中心的多传感器一致性数据融合方法的基本步骤如下^[15]:

假定所有 n 个传感器的测量数据 x_1, x_2, \dots, x_n , 各传感器相应的测量精度为 $\sigma_i, i = 1, 2, \dots, n$;

步骤1:将每一个数据对象视为一类,每类只有一个对象,按式(5)计算它们之间的距离 d_{ij} ,此时类与类间的距离就是它们所包含对象之间的距离,从而得到距离矩阵 $\mathbf{D} = (d_{ij})_{n \times n}$;

步骤2:对 \mathbf{D} 求出满足 $\arg \min_{i,j} \{d_{ij}\}$ 的两个传感器 i 和传感器 j ($i \neq j$),并作如下判定:

(1)若 $\arg \min_{i,j}\{d_{ij}\} \leq 1-\alpha$,则将传感器 i 和传感器 j 的测量数据依下式进行融合处理:

$$x_{(i,j)} = \frac{\sigma_i^{-2}}{\sigma_i^{-2} + \sigma_j^{-2}}x_i + \frac{\sigma_j^{-2}}{\sigma_i^{-2} + \sigma_j^{-2}}x_j, \tag{8}$$

$$\sigma_{x_{(i,j)}}^2 = \frac{1}{\sigma_i^{-2} + \sigma_j^{-2}}. \tag{9}$$

将传感器 i 和传感器 j 的两个类合并成一个新的类称之为“复合传感器” (i,j) ,它的测量数据为 $x_{(i,j)}$,其测量精度为 $\sigma_{x_{(i,j)}}$. 并返回步骤 1;

(2)若 $\arg \min_{i,j}\{d_{ij}\} > 1-\alpha$,则转到步骤 3;

步骤 3:此时出现两种情形,

- ①情形一:所有类最后合并成为一个类,此时所获得的最终数据即为多传感器一致性融合结果;
- ②情形二:所有类合并成为 m 个类($n \geq m \geq 2$),此时以“复合传感器”精度指标最小的数据作为多传感器一致性融合结果.

本文提出的一致性数据融合算法最大的优点是不再把不同精度等级的传感器放在一起进行一致性判断,而是根据层次聚类分析的思想^[15],对具有一致性的两个传感器数据先进行融合处理,然后用融合结果再与其他传感器数据进行一致性判定,确定它们的数据之间是否具有 consistency,对数据的一致性理解上升到一个新层次,即不仅度量了各传感器测量结果之间的一致性,而且度量了某些局部具有一致性的传感器数据融合结果与其他传感器测量数据间的一致性,能更加充分地利用传感器测量数据. 另一方面,层次聚类算法每合并成一个新类对象之后,就必须重新计算合并后类间的距离,也就是距离矩阵,无疑增加了计算量. 但好在实际应用多传感器融合中的传感器数量并不大,因此这一点影响并不足虑.

3 计算实例

下面通过一个计算实例说明本文提出的一致性数据融合算法的优点.

设有 10 个传感器组成的探测系统测量某一目标特性参数(其真值为 10),获得测量数据如表 1 所示.

表 1 10 个传感器的精度指标与测量数据
Table 1 The precision and data of ten sensors

序号	1	2	3	4	5	6	7	8	9	10
观测 x_i	9.954 5	9.874 4	10.082 7	10.346 4	10.400 6	9.964 6	9.063 8	10.183 8	9.056 5	9.049 7
精度 σ_i^2	0.05	0.07	0.10	0.20	0.30	0.25	0.10	0.10	0.20	0.30

采用式(5)定义的置信距离可以得到矩阵 D 为

$$D = \begin{bmatrix} 0 & 0.2 & 0.314\ 8 & 0.784\ 7 & 0.841\ 7 & 0.025\ 4 & 0.955\ 2 & 0.429\ 2 & 0.995\ 5 & 0.995\ 8 \\ & 0 & 0.422\ 3 & 0.792\ 8 & 0.840\ 4 & 0.190\ 5 & 0.969\ 7 & 0.511\ 8 & 0.971\ 2 & 0.972\ 5 \\ & & 0 & 0.444\ 5 & 0.522\ 8 & 0.208\ 3 & 0.977\ 2 & 0.090\ 9 & 0.978\ 2 & 0.979\ 1 \\ & & & 0 & 0.068\ 37 & 0.453\ 9 & 0.995\ 8 & 0.365\ 5 & 0.958\ 6 & 0.959\ 7 \\ & & & & 0 & 0.462\ 6 & 0.997\ 2 & 0.449\ 3 & 0.966\ 4 & 0.918\ 8 \\ & & & & & 0 & 0.956\ 0 & 0.294\ 8 & 0.848\ 9 & 0.804\ 2 \\ & & & & & & 0 & 0.983\ 3 & 0.012\ 9 & 0.025\ 1 \\ & & & & & & & 0 & 0.984\ 0 & 0.984\ 6 \\ & & & & & & & & 0 & 0.008\ 6 \\ & & & & & & & & & 0 \end{bmatrix}$$

可见,传感器 9 和 10 之间的置信距离最小且小于 0.95,两者测量数据根据式(8)和(9)进行融合处理得到:

$$x_{(9,10)} = 9.053\ 78, \sigma_{x_{(9,10)}} = 0.346\ 41.$$

限于篇幅,这里不再展示具体计算过程,仅列出计算结果如下:

- (1)可以计算置信距离矩阵(此时其有 9 个传感器),且传感器 7 与复合传感器(9,10)距离最小且小

于 0.95, 因此可以得到融合结果为:

$$x_{(7,9,10)} = 10.1338, \sigma_{x_{(7,9,10)}} = 0.2335.$$

(2) 此时共有 8 个传感器, 计算其置信距离矩阵, 且可以判定出传感器 1 和 6 间的置信距离最小且小于 0.95, 因此可以得到融合结果为:

$$x_{(1,6)} = 9.9562, \sigma_{x_{(1,6)}} = 0.20412.$$

(3) 此时共有 7 个传感器, 通过计算置信距离可以判定出传感器 4 和 5 间置信距离最小且小于 0.95, 因此可以得到融合结果为:

$$x_{(4,5)} = 10.36808, \sigma_{x_{(4,5)}} = 0.34641.$$

(4) 此时共有 6 个传感器, 通过计算置信距离可以判定出传感器 3 和 8 间的距离最小且小于 0.95, 因此可以得到融合结果为:

$$x_{(3,8)} = 10.10825, \sigma_{x_{(3,8)}} = 0.2236.$$

(5) 此时共有 5 个传感器, 通过计算置信距离可以判定出传感器 2 与复合传感器 (1,6) 间置信距离最小且小于 0.95, 因此可以得到融合结果为:

$$x_{(1,2,6)} = 9.92567, \sigma_{x_{(1,2,6)}} = 0.16161.$$

(6) 此时共有 4 个传感器, 通过计算置信距离可以判定出复合传感器 (1,2,6) 与复合传感器 (3,8) 间的距离最小且小于 0.95, 因此可以得到融合结果为:

$$x_{(1,2,6;3,8)} = 9.98832, \sigma_{x_{(1,2,6;3,8)}} = 0.13098.$$

(7) 此时共有 3 个传感器, 通过计算置信距离可以判定出它们之间的距离均大于 0.95, 因此不再进行融合处理.

(8) 通过比较可以得到最终融合结果为:

$$x_{(1,2,6;3,8)} = 9.98832, \sigma_{x_{(1,2,6;3,8)}} = 0.13098.$$

上述计算过程如图 2 所示.

若采用现有的基于统计距离的一致性数据融合方法^[8,12,14], 按阈值 0.95 判定, 可以得出, 支持传感器 {1} 的有 6 个传感器; 支持传感器 {2} 的有 6 个传感器; 支持传感器 {3} 的有 6 个传感器; 支持传感器 {4} 的有 6 个传感器; 支持传感器 {5} 的有 6 个传感器; 支持传感器 {6} 的有 8 个传感器; 支持传感器 {7} 的有 2 个传感器; 支持传感器 {8} 的有 6 个传感器; 支持传感器 {9} 的有 2 个传感器; 支持传感器 {10} 的有 4 个传感器.

如果以获得 6 个传感器测量支持为准则, 则满足条件的原始传感器序号分别为 {1,2,3,4,5,6,8}, 其融合结果为:

$$x_{1,2,3,4,5,6,8} = 10.0358, \sigma_{1,2,3,4,5,6,8} = 0.1225.$$

比较两种算法的融合结果, 显然真值估计不如本文提出的算法.

对表 1 所示的探测系统进行仿真试验, 仿真中国值分别取了 0.75、0.80、0.85、0.90 和 0.95 各进行了 100 次仿真试验, 并应用本文提出的算法与文献[13,2]的一致性融合算法进行计算比较, 结果如表 2 所示.

从表 2 可见, 对上述不同的阈值的比较结果表明, 本文提出的算法给出的融合结果更优.

4 结论

本文提出的数据融合方法将传感器的数据一致性与传感器的测量精度相结合进行综合考虑, 从而避免了在误差较大的数据参与到融合过程, 在不影响融合精度的前提下, 提高了融合结果的准确性, 因此给出的数据融合结果更有效.

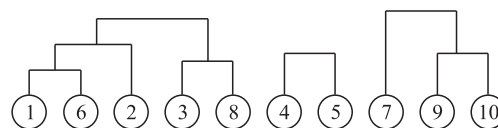


图 2 10 个传感器测量数据的聚类过程

Fig. 2 The hierarchical clustering process of 10 sensors' data

表 2 3 种算法的仿真结果比较

Table 2 Comparison of simulation results of three algorithms

阈值与方法	0.75	0.80	0.85	0.90	0.95
[13]方法优	6	4	2	1	0
[2]方法优	10	8	6	5	2
本文方法优	24	20	18	14	12
三者相同	60	68	74	80	86

[参考文献]

- [1] 熊朝华,刁联旺,张永伟,等. 多传感器探测云下分布式一致信息融合及其发展[J]. 指挥信息系统与技术,2018,9(2): 8-11.
- [2] 刁联旺,王晓璇. 基于精度分级的多传感器一致性数据融合算法[C]//第八届中国信息融合大会论文集. 西安:2017: 638-642.
- [3] 李时辉. 一种多传感器温度数据动态融合方法[J]. 科技通报,2015,31(1):146-170.
- [4] 张品,董为浩,高大冬. 一种优化的贝叶斯估计多传感器数据融合方法[J]. 传感器技术,2014,27(5):643-648.
- [5] 岳元龙,左信,罗雄麟. 提高测量可靠性的多传感器数据融合有偏估计方法[J]. 自动化学报,2014,40(9):1843-1852.
- [6] 薛飞,杨友良,董福涛. 基于多传感器融合算法的CCD测温系统设计[J]. 仪表技术与传感器,2014(8):88-91.
- [7] 张璨,张利,刘征宇. 基于支撑裕的一致性多传感器融合算法[J]. 电子测量与仪器学报,2010,8(24):711-715.
- [8] 万树平. 多传感器数据融合的概率权方法[J]. 传感器与微系统,2009,28(8):12-15.
- [9] 焦竹青,熊伟丽,张林,等. 基于信任度的多传感器数据融合及其应用[J]. 东南大学学报(自然科学版),2008,38(1): 253-257.
- [10] 刁联旺,王常武. 多传感器一致性数据融合方法的改进与推广[J]. 系统工程与电子技术,2002,24(9):60-63.
- [11] 曾黎,蒋沅. 多传感器数据融合的数学方法研究[J]. 云南民族大学学报(自然科学版),2010,19(5):321-325.
- [12] 罗中良,方清城,张前进. 一种多传感器数据融合方法及应用[J]. 传感器技术,2002(2):27-29.
- [13] LOU R C, LIN M, SCHERP P S. Dynamic multi-sensor data fusion system for intelligent robots[J]. Journal of robotics and automation, 1988, 4(4):386-396.
- [14] RAJU G V S, WANG H S. Sensor data fusion using Pitman's closeness technique and complete linkage algorithm[C]//IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems. Las Vegas. NY, 1994:72-75.
- [15] 陈希孺. 高等数理统计学[M]. 合肥:中国科学技术出版社,1999.

[责任编辑:顾晓天]