

文章编号: 1001-0920(2013)09-1427-04

一种基于证据理论的数据融合算法

王文庆, 杨远玲, 杨春杰

(西安邮电大学 自动化学院, 西安 710121)

摘要: 针对传感器测量值存在系统误差的情况, 基于证据理论的思想, 提出一种新的数据融合算法. 该算法首先将所有测量值根据其与其真值的偏差进行分组, 并分配不同的基本信任; 然后将其构成的集合视为辨识框架, 进而将各个测量值转换为相应的证据并进行证据组合, 所得合成证据的 Mass 函数即为各个测量值的权值分配函数; 最后对所得分组融合测量值进行加权求和, 即得融合结果. 仿真结果验证了该算法的有效性.

关键词: 数据融合; 证据理论; 基本信任分配; 证据组合

中图分类号: TP212

文献标志码: A

A data fusion algorithm based on evidence theory

WANG Wen-qing, YANG Yuan-ling, YANG Chun-jie

(School of Automation, Xi'an University of Posts & Telecommunications, Xi'an 710121, China. Correspondent: YANG Yuan-ling, E-mail: yangyuanling2006@sina.cn)

Abstract: According to the systematic error of sensors data, a new data fusion algorithm is proposed based on the evidence theory. The proposed algorithm divides sensor data into groups according to the deviation, and the concept of basic probability assignment is applied to generate the frame of discernment. The combination rule is also applied, and the Mass function of the combined evidence is referred as the weight assignment function of all data. Finally, the fusion result is obtained by weighted summation. Simulation results show the effectiveness of the proposed algorithm.

Key words: data fusion; evidence theory; basic probability assignment; evidence combination

0 引言

多传感器数据融合是将来自多个传感器的信息和数据进行综合分析处理, 进而得到比单个传感器更为准确而可靠的结论^[1]. 国内外许多学者针对此问题进行了广泛而深入的研究. 文献[2]提出了基于 Bayes 估计的数据融合方法; 文献[3]利用 Grubbs 判据剔除稀疏数据, 再利用模糊贴近度对余下的有效数据进行融合; 文献[4-6]提出了距离矩阵的概念来度量数据之间的相似度, 建立了新的融合算法; 文献[7]利用置信距离比较各传感器测量值之间的相似度, 从而进行一致性数据融合; 文献[8-9]从多元统计理论角度, 利用最小距离聚类法对多传感器数据进行融合.

以上方法都直接或间接地依赖一个假设: 所有传感器的测量值都服从同一参数的正态分布, 即各传感器的一次测量值都以同一参数的正态分布散落在真值附近. 但实际情况是, 由于器件或测量系统不可避免地存在一定的系统误差, 这种误差因传感器而异;

或者由于环境、人为因素等的影响, 不同传感器的测量值散落在各自的某个近似真值的附近, 在围绕某个近似真值附近的范围内服从一定的分配. 因此, 在进行传感器权重确定和证据组合时, 若忽略这种差异会在一定程度上影响融合结果的客观性和准确性^[10-11].

为提高数据融合结果的准确性, 融合方法应根据测量数据情况进行相应的调整. 为此, 本文基于证据理论的思想, 提出了一种新的数据融合算法. 该算法无需假设传感器测量值服从同一参数的正态分布, 也不需要传感器的先验信息或通过历史数据来获得近似先验信息, 尤其适用于对系统误差不可忽略的多个传感器的实时测量数据进行加权融合, 具有算法简洁实用, 融合精度高, 融合结果客观可信等特点.

1 融合算法描述

1.1 证据理论的基本思想及组合规则

证据理论是由 Dempster 提出, 并由 Shafer 进一

收稿日期: 2012-04-19; 修回日期: 2012-06-25.

基金项目: 国家自然科学基金项目(61100165/F020508); 陕西省自然科学基金项目(2007F18).

作者简介: 王文庆(1964-), 男, 教授, 从事复杂系统分析与控制、智能信息处理等研究; 杨远玲(1987-), 女, 硕士生, 从事智能信息处理的研究.

步完善和发展起来的处理不确定性推理问题的数学方法. 证据理论首先定义一个样本空间 U 为辨识框架, U 是由有限个完备而互斥的所有可能元素构成的集合, 2^U 为 U 的幂集. 对于辨识框架 U 的任意一个子集 A , 定义映射 $m: 2^U \rightarrow [0, 1]$ 为 U 上的基本信任分配函数, 它满足

$$\begin{cases} \sum_{A \subset U} m(A) = 1, \\ m(\emptyset) = 0. \end{cases}$$

若 $m(A) > 0$, 则称 A 为信任函数 m 的焦点, 所有焦点的并称为核^[12].

利用求出的 m 信任函数进行判决, 并利用最小点原则, 希望缩小真值范围或找出真值, 删除可以去掉的元素, 最后利用可靠性的交易融合规则(类似于基于概率的加权平均方法)对证据进行组合, 即

$$m(A) = \sum_i^I R_i m_i(A), \quad (1)$$

其中 $R_i (i = 1, 2, \dots, I)$ 是可靠性因子.

1.2 算法描述

对于多个传感器测得的一组数据, 首先根据其分布情况进行分类, 分类后测量值将被分成不同的组, 对于分类后单独的一组数据将其中所有测量值 $\{S_1, S_2, \dots, S_{N_1}\}$ 视为辨识框架 Θ , 并分别将各个测量值转换为辨识框架 Θ 上的一个证据; 其次, 对生成的这些证据进行组合, 所得合成证据中各个测量值的基本信任分配即为融合加权系数; 再次, 对这些测量值进行加权求和, 得到该组数据融合的结果; 最后, 将每组获得的结果进行重新分配, 权系统进行组合即得融合结果. 算法主要涉及的问题有: 1) 如何将测量数据进行分组是问题的关键, 也是下一步证据生成的前提; 2) 将测量数据进行分组后, 如何确定不同组数据获得的信任度分配是该算法的另一个关键问题; 3) 在获得 N 个证据之后, 采取何种证据组合规则进行证据合成. 目前有修正组合公式方法和预处理方法两大类解决冲突证据组合的方法, 组合规则的选取在很大程度上决定了最终融合结果的精度.

1.2.1 测量值分组

由于环境噪声和人为干扰等各种因素的影响, 各个传感器给出的测量值各不相同, 但总是以一定的值为中心并散布在其周围的某一邻域内. 为了反映不同传感器测量数据之间的偏差大小, 可通过测量数据本身来确定. 定义测量数据 X_p, X_l 之间的距离为

$$d_{pl} = |X_p - X_l|. \quad (2)$$

d_{pl} 的值越小, X_p, X_l 越接近, 否则偏差较大.

假设多个传感器测得一组数据为 $S = \{s_i, i =$

$1, 2, \dots, N\}$, 由于存在系统误差, 这一组数据可分为 k 组, 各组的中心点为 $d_0^{(j)}, j = 1, 2, \dots, k$, 且 $d_0^{(1)} \leq d_0^{(2)} \leq \dots \leq d_0^{(k)}$, 则第 j 组数据表示为 $S^{(j)} = \{s^l, s^l \in (d_0^{(j)} - \delta^{(j)}, d_0^{(j)} + \delta^{(j)}, l = 1, 2, \dots, L_j)\}$, 其中 $\delta^{(j)}$ 为第 j 组邻域的半径, L_j 为第 j 组所含数据个数. 于是 S 数据集可表示为 $S = \bigcup S^{(j)}, j = 1, 2, \dots, k$, 这样可将测量数据分成 k 组, 每组所含数据个数为 $L_j, \sum_{j=1}^k L_j = N$.

1.2.2 基本信任分配

下面在对各测量值进行分组的基础上, 给出将测量值 S_n 转换为证据 e_n 的方法:

对于以 $d_0^{(j)}$ 为中心的一组值 $S^{(j)}$, 定义 $d_n = \sum_{m=1}^{L_j} |S_n - S_m| / L_j$, 当它离真值近时获得的基本信任分配就高; 离真值远时获得的基本信任分配就低; 假设落在以 $d_0^{(j)}$ 为中心, $\delta^{(j)}$ 为半径 ($\Gamma_{n_j} \in S^{(j)} = \{s_l, s_l \in (d_0^{(j)} - \delta^{(j)}, d_0^{(j)} + \delta^{(j)}, l = 1, 2, \dots, L_j\}$) 的范围内共有 L_j 个值, 则这 L_j 个测量值被认为以相同的概率接近真值 S_0 (S_0 为目标真值), 并获得相同的信任分配. 以 $d_0^{(j)}$ 为中心的一组值所得证据 e_{n_j} 的基本信任函数为

$$m_{n_j}(X_{L_j}) = 1/L_j, \forall X_{L_j} \in \Gamma_{n_j}, j = 1, 2, \dots, k, \quad (3)$$

即位于以 $d_0^{(j)}$ 为中心的范围内的 L_j 个测量值均获得 $1/L_j$ 的基本信任分配. 因此, 由第 $j (j = 1, 2, \dots, k)$ 组测量值生成 L_j 个证据, 其中 $\sum_{j=1}^k L_j = N$. 为了数学

描述方便, 下文用下标 i 表示证据编号. 为进一步提高融合精度, 将各分组后测量值的偏差程度考虑在内, 并对各个初始证据 $e_i (i = 1, 2, \dots, L_j)$ 分别进行修正.

1) 对于测量值 $S_i (i = 1, 2, \dots, L_j)$, 若 $d_i \geq \delta \bar{d}$, 则 S_i 被视为野值, 将其剔除, 即

$$m_i(S_i) = 0, d_i \geq \delta \bar{d}. \quad (4)$$

2) 若 $d_k < \delta \bar{d}$, 则 S_k 为有效测量值. 任意两个有效测量值 S_{K1} 和 S_{K2} 获得的基本信任分配之比为

$$m_i(S_{k1})/m_i(S_{k2}) = d_{k2}/d_{k1}, \quad (5)$$

其中常数 $\delta \geq 1$, 经过对多组数据进行融合实验之后, 取 $\delta = 1.25$. 由于 δ 为比例值, 是一个相对量, $\delta = 1.25$ 具有较强的一般性, 本文第 3 部分的实验仿真中均取 $\delta = 1.25$.

在实际操作中, 可由式 (4) 产生一组修正系数 $\{w_n\}, n = 1, 2, \dots, L_j$. 对于有效测量值 $S_k, w_k = 1/d_k$; 对于野值 $S_i, w_i = 0$. 利用修正系数对各个命题进行加权修正并归一化, 有

$$\hat{m}_i(S_n) = w_n g m_i(S_n) / \sum_{p=1}^{L_j} w_p g m_i(S_p). \quad (6)$$

至此,由式(3)~(6)完成了第j组证据的产生及修正过程.

1.3 证据组合

证据理论只适用于证据独立的情况,当合成高度冲突的证据时,组合导出的推理结果往往会出现悖论.由式(6)生成的第j组的 L_j 个证据可能存在较高冲突,采用Dempster组合规则可能会得到不合理的组合结果.文献[13]将支持冲突证据的那部分基本概率赋值全部赋给了 U ,即认为冲突的证据不能提供任何信息;文献[14]对基本概率赋值 m_i 设定一个权重系数 p_i ,再按Dempster组合规则进行融合;传感器的重要程度会对融合结果产生影响,比较重要的传感器对融合结果的影响较大,且各个证据的可信度不一样,可信度较高的证据对融合结果影响较大.因此,本文根据实际提出了如下证据组合方法:将支持证据冲突的概率按比例分配给各个测量值 S_n , $n = 1, 2, \dots, L_j$,组合公式如下所示:

$$m(S_n) = \prod_{i=1}^{L_j} \hat{m}_i(S_n) + c g \bar{m}_i(S_n). \quad (7)$$

这里 m 表示合成证据E的Mass函数,其中冲突因子

$$c = 1 - \sum_{n=1}^{L_j} \prod_{i=1}^{L_j} \hat{m}_i(S_n), \quad (8)$$

$\bar{m}_i(S_n)$ 为 S_n 在所有证据中的平均基本信任分配,即

$$\bar{m}_i(S_n) = \sum_{n=1}^{L_j} \hat{m}_i(S_n) / L_j. \quad (9)$$

合成证据E中 S_n 的基本信任分配 $m(S_n)$ 即为 S_n 获得的权值,则融合结果为

$$S_0^{(j)} = \sum_{n=1}^{L_j} S_n m(S_n). \quad (10)$$

由此得到了第j组数据融合的结果,同理得到了所有 K 组所得融合结果 $S_0^{(j)} = \{s_0^{(j)}, j = 1, 2, \dots, k\}$.由于分组的不同和每组测量数据个数的差异,并且考虑到每组值中心点离真值距离的差异,根据经验可知一般集中在真值附近的测量数据较多且信任度较高,获得的信任分配相对较大.考虑各种因素,对所得的 k 个融合结果进行加权组合,得到最终的融合结果为

$$S_0 = \sum_{j=1}^k L_j S_0^{(j)} / N, N = \sum_{j=1}^k L_j. \quad (11)$$

2 实验仿真

为验证本文算法的优越性,下面给出算法的实验仿真,并与参考文献中的类似融合算法进行性能比较.

以测量恒温箱温度为例,测得某恒温箱的一组数据为

$S = [49.80, 49.92, 50.25, 50.15, 50.55, 50.60, 50.64, 50.75, 50.80, 50.98, 51.05, 51.20, 51.30, 49.70, 49.85, 48.25, 49.10, 47.65, 47.90, 48.00, 48.50, 48.62, 48.90, 52.95, 53.10, 53.20, 53.35, 53.80, 54.05, 53.00]$;

总共有30个数据.假设恒温箱的实际温度为50.50,则分组情况如表1所示.

表1 数据分组情况

参数	分 组		
	第1组	第2组	第3组
Γ_n	S_1, \dots, S_{15}	S_{16}, \dots, S_{23}	S_{24}, \dots, S_{30}
$d_0^{(j)}$	50.05	48.50	53.50
$\delta^{(j)}$	0.85	0.70	0.55
L_j	15	8	7

由式(3)~(9)对3组数据分别进行融合,所得结果以第2组数据为例,如表2所示.

表2 第2组证据的信任函数

各个证据	基本信任分配				
	$m_2(S_{16})$	$m_2(S_{17})$...	$m_2(S_{22})$	$m_2(S_{23})$
e_{16}	0.2000	0	...	0.200	0
e_{17}	0.1250	0	...	0.125	0.125
e_{18}	0.1250	0	...	0.125	0.125
e_{19}	0.2500	0	...	0	0
e_{20}	0.2000	0	...	0	0
e_{21}	0.2000	0	...	0.200	0.200
e_{22}	0.2000	0	...	0.200	0.200
e_{23}	0.1250	0	...	0.125	0.125
合成证据E	0.2657	0	...	0.172	0
S_0^2	48.4625				

同理可得出第1组和第3组的融合结果为

$$S_0^1 = 50.5942, S_0^3 = 53.1298.$$

利用式(11)对3组结果进行加权组合,得出50.50作为标准真值,则本文算法融合结果的偏差为0.06.

结果比较:若采用文献[1]的方法将所有测量值进行证据生成,融合所得的融合结果为50.3618,偏差为0.1382,因此本文融合结果更接近真值,且精度更高.

3 结 论

基于证据理论的思想,提出了一种新的数据融合算法,该算法无需假设传感器测量值服从同一参数的正态分布,也不需要传感器的先验信息或通过历史数据来获得近似先验信息.特别是对于测量数据,由于环境、人为因素存在系统误差,且测量数据相对较多,结果表明本文算法的优越性更强,融合精度较高.

然而,在没有现场经验的情况下,特别是当测量数据量较大时,如何对数据进行准确分组是应用本

文算法的一个难题. 借助于其他统计工具(如聚类算法)对原始测量数据进行分组是下一步研究的内容. 此外, 在证据组合方法上, 还有进一步改进的空间和必要, 为此, 下一步工作将主要集中在这两个方面, 以进一步提高本文算法的融合性能.

参考文献(References)

- [1] 熊彦铭, 杨战平. 受证据理论启发的传感器数据融合算法[J]. 华中科技大学学报, 2011, 39(10): 50-54.
(Xiong Y M, Yang Z P. Sensor data fusion algorithm inspired by the theory of evidence[J]. J of Huazhong University of Science and Technology, 2011, 39(10): 50-54.)
- [2] 吴小俊, 曹奇英, 陈保香. 基于 Bayes 估计的多传感器数据融合方法研究[J]. 系统工程理论与实践, 2000, 20(7): 46-49.
(Wu X J, Cao Q Y, Chen B X. Study on multi-sensor data fusion methods based on Bayes estimation[J]. Systems Engineering-Theory & Practice, 2000(7): 46-49.)
- [3] 项新建. 基于模糊数学与统计理论集成的多传感器数据融合方法[J]. 传感技术学报, 2004, 6(2): 22-24.
(Xiang J X. A method to sensor data fusion based on fuzzy and statistics integration[J]. Chinese J of Sensors and Actuators, 2004, 6(2): 22-24.)
- [4] 刘建书, 李人厚, 常宏. 基于相关性函数和最小二乘的多传感器数据融合[J]. 控制与决策, 2006, 21(6): 714-716.
(Liu J S, Li R H, Chang H. Multi-sensor data fusion based on correlation function and least square[J]. Control and Decision, 2006, 21(6): 714-716.)
- [5] 刘敏华, 萧德云. 基于相似度的多传感器数据融合[J]. 控制与决策, 2004, 19(5): 534-537.
(Liu M H, Xiao D Y. Multi-sensor data fusion based on similitude degree[J]. Control and Decision, 2004, 19(5): 534-537.)
- [6] 孙勇, 景博. 基于支持度的多传感器一致可靠性融合[J]. 传感技术学报, 2005, 18(3): 103-105.
(Sun Y, Jing B. Consistent and reliable fusion of multi-Sensor based on support degree[J]. Chinese J of Sensors and Actuators, 2005, 18(3): 103-105.)
- [7] 刁联旺, 王常武, 商建云. 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): 60-62.
(Diao L W, Wang C W, Shang J Y. Improved and generalized consensus data fusion method[J]. Systems Engineering and Electronics, 2002, 24(9): 60-62.)
- [8] 涂国平, 邓群钊. 多传感器数据的统计融合方法[J]. 传感器技术, 2001, 20(3): 28-30.
(Tu G P, Deng Q Z. A statistical method for the data fusion of multi-sensor[J]. J of Transducer Technology, 2001, 20(3): 28-30.)
- [9] 万树平. 多传感器数据的聚类融合方法[J]. 系统工程理论与实践, 2008, (5): 132-135.
(Wan S P. Method of clustering fusion for multi-sensors data[J]. Systems Engineering-Theory & Practice, 2008, (5): 132-135.)
- [10] 郭惠昕. 基于模糊集的证据组合方法及其应用[J]. 控制与决策, 2008, 23(2): 229-232.
(Guo H X. Approach to evidence combination based on fuzzy theory and its applications[J]. Control and Decision, 2008, 23(2): 229-232.)
- [11] 孙全, 叶秀清, 顾伟康. 一种新的基于证据理论的合成公式[J]. 电子学报, 2000, 28(8): 117-119.
(Sun Q, Ye X Q, Gu W K. A new combination rules of evidence theory[J]. Acta Electronica Sinica, 2000, 28(8): 117-119.)
- [12] 王润生. 信息融合[M]. 北京: 科学出版社, 2007.
(Wang R S. Information fusion[M]. Beijing: Science Press, 2007.)
- [13] Yager R R. On the relationships of methods of aggregation of evidence in experts systems[J]. Cybernetics and Systems, 1985, 16(1): 1-21.
- [14] Wang Ping, Yang Genqing. Improvement method for the combining rule of dempster-shafer evidence theory based on reliability[J]. J of Systems Engineering and Electronics, 2005, 16(2): 471-474.

(上接第1426页)

- [6] Venayagamoorthy G K, Harley R G, Wunsch D C. Dual heuristic programming excitation neurocontrol for generation in a multi-machine power system[J]. IEEE Trans on Industry Applications, 2003, 39(2): 382-394.
- [7] Wei Q L, Zhang H G, Cui L L. Data-based optimal control for discrete-time zero-sum games of 2-D systems using adaptive critic designs[J]. Acta Automatica Sinica, 2009, 35(6): 682-692.
- [8] Vrabie D, Lewis F L. Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear system[J]. Neural Networks, 2009, 22(3): 237-246.
- [9] Chellaboina V, Haddad W M. A unification between partial stability and stability theory for time-varying systems[J]. IEEE Trans on Control Systems, 2002, 22(6): 66-75.
- [10] Khalil H K. Nonlinear system[M]. NJ: Prentice-Hall, 2002.