

文章编号: 1001-0920(2004)05-0534-04

基于相似度的多传感器数据融合

刘敏华, 萧德云

(清华大学 自动化系, 北京 100084)

摘 要: 利用多传感器状态估计向量(或测量值)的标称化差定义了相似度和相似度矩阵, 用空间信息形成一致性测度, 用时间信息形成可靠性测度, 最终形成了多传感器的组合及加权, 并进行时空融合. 该融合既可在数据层进行, 也可在决策层进行. 仿真计算表明了基于相似度的数据融合的有效性

关键词: 数据融合; 相似度; 一致性测度; 可靠性测度

中图分类号: TP212.9 **文献标识码:** A

Multi-sensor data fusion based on similitude degree

L I U M in-hua, X I A O D e-yun

(Department of Automation, Tsinghua University, Beijing 100084, China Correspondent: L I U M in-hua, E-mail: lmh99@mails.tsinghua.edu.cn)

Abstract: The similitude degree and similitude degree matrix are defined by using the normal difference of status estimate vectors (or measures). The consistency measure is given using spacial information and the reliability measure is given using temporal information. Then the sensor combination and fusion weights are proposed and the space-time fusion is made. The fusion can be made on both the data level and the decision level. The simulation results show the congeniality of data fusion based on similitude degree.

Key words: data fusion; similitude degree; consistency measure; reliability measure

1 引 言

多传感器数据融合是在传感器造价降低、种类和数量不断增多的情况下新兴的一门学科, 其研究重点之一是数据层、特征层和决策层的各种融合方法^[1]. 如果已知多传感器系统各传感器的先验概率分布和关联概率分布, 则可利用最优统计决策法^[2], 通过联合多传感器后验概率分布进行最优融合. 如果系统具有线性动力学模型和高斯分布的系统噪声与测量噪声, 则可利用 Kalman 滤波法^[3]进行融合. 如果缺乏这些条件, 则一般应从多传感器数据自身出发, 考察多传感器数据的置信距离, 进行一致性传感器数据融合^[4]. 后者实际上利用一组多传感器数据形成距离矩阵选择最大传感器组, 然后按一定的

算法或算子^[5]进行融合. 但这种融合仅仅部分地利用了多传感器数据的空间信息, 而没有利用其时间信息. 另一方面, 融合是利用多传感器数据的冗余和互补来增强其结果的一致可靠性, 因此测度一致可靠的程度也是融合的一个尺度, 但这方面研究较少有人涉及.

为充分利用多传感器数据信息和测度一致可靠性, 本文利用多传感器状态估计向量(或测量值)的标称化差定义了相似度和相似度矩阵, 用空间信息形成一致性测度, 用时间信息形成可靠性测度, 最终形成了多传感器的组合及加权, 进行时空融合. 这一融合既可在数据层进行, 也可扩展到决策层. 仿真结果显示了基于相似度的多传感器数据融合方法的有

收稿日期: 2003-04-17; 修回日期: 2003-08-25

基金项目: 国家 863 计划项目(2002AA 412420).

作者简介: 刘敏华(1976—), 男, 江苏海门人, 博士, 从事多传感器数据融合等研究; 萧德云(1945—), 男, 福建莆田人, 教授, 博士生导师, 从事辨识建模、故障诊断及传感器信号融合等研究

效性

2 基于相似度的数据层融合

2.1 相似度矩阵

测量系统的同质传感器集合为 $S = (s_1, s_2, \dots, s_n)$, $z_i(k)$ 表示 k 时刻传感器 s_i 的输出. 设系统可用如下状态方程和输出方程描述:

$$\begin{cases} X(k+1) = \Phi(k)X(k) + G(k)V(k), \\ Z(k) = H(k)X(k) + W(k). \end{cases} \quad (1)$$

其中: $\Phi(k), G(k), H(k)$ 分别表示状态转移矩阵、过程噪声分布矩阵及输出矩阵, 式(1)中省略了各传感器编号的下标; $V(k)$ 和 $W(k)$ 分别表示具有零均值和正定协方差矩阵的高斯噪声向量. 采用 Kalman 滤波算法^[2]进行状态更新.

在 k 时刻, 由于传感器所处的噪声环境和自身性能的不一致, 会形成略有差异的状态估计向量. 为度量这一差异, 定义如下的状态估计向量的标称化差:

$$u_{ij}(k) = C_{ij}^{-1/2}(k|k) [\hat{X}_i(k|k) - \hat{X}_j(k|k)] \quad (2)$$

其中 $C_{ij}(k|k) = P_i(k|k) + P_j(k|k)$, 表示两个测量的估计误差协方差之和. 采用正态型隶属度函数的模糊测度, 定义 k 时刻两状态向量的相似度为

$$d_{ij}(k) = \exp[-bu_{ij}^T(k)u_{ij}(k)], \quad (3)$$

其中: b 是系数, $u_{ij}(k)$ 是列矢量, $d_{ij}(k)$ 是标量. 由式(2)可得

$$d_{ij}(k) = \exp\{-b[\hat{X}_i(k|k) - \hat{X}_j(k|k)]^T \times [P_i(k|k) + P_j(k|k)]^{-1} \times [\hat{X}_i(k|k) - \hat{X}_j(k|k)]\}. \quad (4)$$

因测量系统由独立同质传感器组成, 故状态向量间的相似程度也表征了测量值自身之间的相似程度, 二者是一致的. 当系统无法用状态方程表示时, 可直接用测量值计算相似度.

由相似度可得 k 时刻多传感器测量系统的相似度矩阵为

$$D(k) = \begin{bmatrix} 1 & d_{12}(k) & \dots & d_{1n}(k) \\ d_{21}(k) & 1 & \dots & d_{2n}(k) \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1}(k) & d_{n2}(k) & \dots & 1 \end{bmatrix}_{n \times n}. \quad (5)$$

相似度矩阵包含了 k 时刻传感器系统 $S = (s_1, s_2, \dots, s_n)$ 的测量在空间分布的信息, 是进行多传感器数据空间融合的尺度. 同样, 时间系列 $\{D(k), k = 1, 2, \dots\}$ 包含了到当前时刻为止的测量在时空分布

的信息, 是进行多传感器数据时空融合的尺度.

2.2 数据层融合

设 $c_i(k)$ 是 k 时刻传感器 i 的一个计数器 (初值为零). 考察 $D(k)$ 的第 i 行, 若 $d_{ij}(k) \in \epsilon$ (ϵ 是设定的阈值), 则计数器加 1. 此时第 i 行扫描之后的 $c_i(k)$ 终值表示 k 时刻与传感器 i 测量数据较为相似的测量数据数. $c_i(k)$ 大, 表示 k 时刻传感器 i 的测量值与大多数测量值一致, 这些测量值可能组成一个真值的聚类; $c_i(k)$ 小, 表示 k 时刻传感器 i 的测量值与大多数测量值不一致, 而成为“野值”的可能性较大. 因此 $c_i(k)$ 是测量值一致性的度量. 定义 k 时刻传感器 i 的一致性测度为

$$p_i(k) = c_i(k)/n, \quad (6a)$$

显然有 $0 \leq p_i(k) \leq 1$, 这实际上是一种可能性测度. $p_i(k)$ 也可由相似度直接定义, 即

$$p_i(k) = \frac{1}{n} \sum_{j=1}^n d_{ij}(k). \quad (6b)$$

相对于式(6a)的硬测度而言, 式(6b)给出的一致性测度则是一种软测度. 于是, 由传感器组 $S = (s_1, s_2, \dots, s_n)$ 得到的 k 时刻一致性向量为 $p(k) = [p_1(k), p_2(k), \dots, p_n(k)]$.

对传感器 i 而言, 除了与其他传感器的一致性外, 还存在自身的可靠性问题, 这种可靠性往往通过测量的时间系列表现出来. $[p_i(1), p_i(2), \dots, p_i(k)]^T$ 表示传感器 i 的一致性测度的时间序列, 令

$$\bar{p}_i(k) = \frac{1}{k} \sum_{t=1}^k p_i(t) \quad (7)$$

表示平均或综合的一致性测度. 如果序列波动不大, 则说明传感器 i 的性能比较稳定或环境噪声较小, 即可靠性较高. 故可直观地定义传感器 i 的可靠性测度为一致性测度的方差, 即

$$\sigma_i^2(k) = \frac{1}{k} \sum_{t=1}^k [p_i(t) - \bar{p}_i(k)]^2. \quad (8)$$

用于融合的传感器数据或在融和中权重较大的传感器数据应是一致性较大且可靠性较高者, 即 $\bar{p}_i(k)$ 较大而 $\sigma_i^2(k)$ 较小. 对传感器 i 而言, $\bar{p}_i(k)$ 较大并不意味着 $\sigma_i^2(k)$ 较大或较小, 反之亦然. 故引入映射 $f[\bar{p}_i(k), \sigma_i^2(k)]$ 来综合二者, 使选择的融合数据组合与 $\bar{p}_i(k)$ 正相关, 且与 $\sigma_i^2(k)$ 负相关. 如用双线性定义一致可靠测度

$$q_i(k) = f[\bar{p}_i(k), \sigma_i^2(k)] = [1 - a\sigma_i(k)]\bar{p}_i(k), \quad (9)$$

其中 a 是一个合适的系数, 通常 $0 < a < 1$. 这是因为一致性与可靠性相比显得更为重要一些. 这样, 便可

根据 $q_i(k)$ 值的大小对传感器测量值进行排队, 然后采用不同的门限方法进行融合. 如选择一组大于阈值 ϵ 的传感器组数据加权融合, 或将前 $l(n/3 \sim l \sim n)$ 个传感器组数据进行加权融合. 这两种方法相当于筛除一部分一致性小而可靠性差的数据, 然后对最大传感器组进行融合. 当然, 也可不舍弃任何数据, 将 k 时刻 n 个数据按如下加权公式进行融合:

$$z_f(k) = \frac{\sum_{i=1}^n q_i(k) z_i(k)}{\sum_{i=1}^n q_i(k)} \quad (10)$$

2.3 基于相似度的融合分析

相似度可直接由测量值计算得到, 即 $d_{ij}(k) = \exp\{-b[z_i(k) - z_j(k)]^2\}$, 这相当于度量多传感器数据间的置信距离. 所不同的是状态估计往往是一个具有物理意义、反映系统特性的向量, 而测量值则是一个标量. 因此, 由状态估计计算相似度更加综合而准确.

相似度计算需要计算 n 个传感器两两之间的标称化差, 一致性测度计算需要将 n^2 维的相似度矩阵元素与阈值逐一比较, 计算的复杂度为 $o(n^2)$. 为减少计算量, 在计算综合一致性测度和可靠性测度时, 应使用如下递推算法:

$$\bar{p}_i(k) = \frac{k-1}{k} \bar{p}_i(k-1) + \frac{1}{k} p_i(k), \quad (11a)$$

$$\sigma_i^2(k) = \frac{k-1}{k} \left\{ \sigma_i^2(k-1) + \frac{1}{k} [p_i(k) - \bar{p}_i(k)]^2 \right\} \quad (11b)$$

3 基于相似度的决策层融合

基于相似度的多传感器数据层融合可扩展到决策层融合, 或者说 n 个传感器测量用于模式分类 (类别数为 c), 这时相似度和相似度矩阵有所不同. 设各个模式类的状态向量为 $X_j (1 \leq j \leq c)$, 则标称化差为 $u_{ij}(k) = C_{ij}^{-1/2}(k|k) [X_i(k|k) - X_j]$, 相似度仍按式 (3) 计算, 但此时得到的相似度矩阵为 $n \times c$ 维, 即

$$D(k) = \begin{bmatrix} d_{11}(k) & d_{12}(k) & \dots & d_{1c}(k) \\ d_{21}(k) & d_{22}(k) & \dots & d_{2c}(k) \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1}(k) & d_{n2}(k) & \dots & d_{nc}(k) \end{bmatrix}_{n \times c}$$

根据相似度矩阵, 可采用数据层融合的一致性可靠性测度进行决策层融合. 相应的一致性 (综合) 测度计算为

$$p_j(k) = \frac{1}{n} \sum_{i=1}^n d_{ij}(k), \bar{p}_j(k) = \frac{1}{k} \sum_{t=1}^k p_j(t),$$

而可靠性测度计算为

$$\sigma_j^2(k) = \frac{1}{k} \sum_{t=1}^k [\bar{p}_j(t) - p_j(t)]^2.$$

最终的目标类别号为

$$r = \arg \max_j [\bar{p}_j(k), \sigma_j^2(k)]$$

如果将相似度矩阵视为 k 时刻 n 个传感器对 c 个类别形成决策的可能性分布, 则可先进行时间融合, 再进行空间融合, 形成一个融合可能性分布的决策表. 具体做法如下: 首先将相似度矩阵转化为可能性分布, 即

$$\Pi_i(k) = \frac{d_{ij}(k)}{o_j};$$

其次采用综合映射函数^[6]进行时间融合, 即

$$\Pi_i^k = S[\Pi_i(1), \dots, \Pi_i(k)] = \frac{d_{ij}^k}{o_j};$$

最后采用综合映射函数进行空间融合, 即

$$\Pi^k = S[\Pi_1^k, \dots, \Pi_n^k] = \frac{d_j^k}{o_j}.$$

4 仿真计算

考虑 3 个传感器在 5 个采样时刻的融合情况. 传感器输出如表 1 所示.

表 1 传感器输出

传感器	k				
	1	2	3	4	5
1	0.8	0.4	0.4	0.7	0.2
2	0.8	0.8	0.7	0.8	0.8
3	0.8	0.6	0.5	0.8	0.8

直接利用传感器值计算相似度矩阵, 取式 (3) 中 $b = 5$; 采用式 (6a) 定义的一致性测度, 取 $\epsilon = 0.5$, 取式 (9) 中 $a = 1$. 运算结果如图 1 和表 2 (第 2 行) 所示.

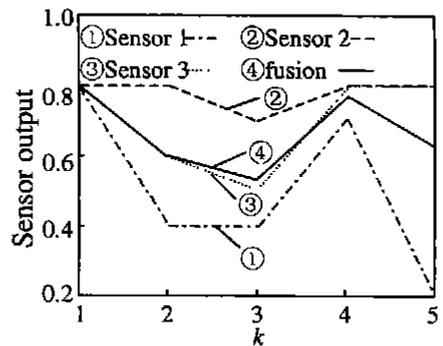


图 1 基于相似度的多传感器数据融合

由图 1 可见, 当 $k = 2$ 时, 传感器输出是围绕中值对称的情形, 这表明相似度融合的结果等于中值, 这是融合方法合理的一个必要条件. 当 $k = 3$ 时, 传感器输出并不对称, 融合结果偏向输出值比较相近

的数值,这体现了相似度融合的聚类作用。当 $k = 4$ 时,传感器 1 和 2 的输出值不同,但具有相同的加权,而传感器 2 和 3 的输出值相同,但具有不同的加权,这说明传感器时间融合已对某一时空的空间融合产生了作用。这一点在 $k = 5$ 时看得更加清楚。传感器 1 的值明显地偏离了传感器 2 和 3 的输出值,因此加权比较小,表现出一定的抗干扰性,但仍有一定的权重,传感器 1 的输出仍有一定的综合一致性和可靠性。从图 1 可以看出,融合曲线与居中的 s_3 输出曲线相近,但走势更加平缓,说明了相似度融合的合意性。

上述结果是按式(6a)计算一致性测度的,如果按式(6b)计算,则最终融合结果略有不同,这与平均值融合结果有一定的差异,因为后者不包含空间和时间信息。对比情况如表 2 所示。

表 2 3 种融合值的比较

融合方法	k				
	1	2	3	4	5
式(6a)	0.8	0.6	0.53	0.77	0.63
式(6b)	0.8	0.6	0.531	0.768	0.625
平均值	0.8	0.6	0.533	0.767	0.6

注意到综合一致性测度既包含了空间信息,也包含了时间信息,而可靠性测度则包含了时间信息,因此可采用一致性测度和可靠性测度融合数据,即令 $q_i(k) = f[p_i(k), \sigma_i^2(k)]$ 与前述方法强调的重点不同,后者更加重视当前多传感器数据的空间分布。如 $k = 5$ 时刻,按式(6b)计算一致性测度,按 $q_i(k) = f[p_i(k), \sigma_i^2(k)]$ 加权,所得结果为 0.67,均大于前述融合值而更接近于 0.8,因为该时刻空间分布更倾向于传感器 2 和 3 的值(但同样没有屏蔽掉传感器 1 的值)。

另外当传感器数据增加时,相似度融合的性能和效果有望得到提高,因为此时“野值”的一致性陡

然降低,而其他聚类的传感器测量值的一致性则没有太大的变化。

5 结 语

本文通过状态估计的标称化差计算得到多传感器数据的相似度矩阵,并以包含了空间和时间信息的相似度矩阵时间序列作为融合的尺度。这一融合方法既可用于数据层融合,也可用于决策层融合,其方法之一是利用所提出的一致性测度和可靠性测度来获取用于融合的传感器组和相应的数据权重。一致性测度和可靠性测度分别包含了多传感器数据的空间和时间信息,因此基于相似度的多传感器数据融合是一种时空融合。分析和计算表明,该方法直观有效,既能控制故障数据的干扰,又能包容多传感器信息。本文选用的一致可靠性测度映射函数是最简单的双线性函数,在实际应用中应选择更加合适的符合条件的测度函数。

参考文献(References):

- [1] Varshney P. K. Multisensor data fusion [J]. *J of Electronics and Communication Engineering*, 1997: 245-253.
- [2] John M. R. Fusion of multisensor data [J]. *The Int J of Robotics Research*, 1988, 7(6): 78-96.
- [3] Rao B. S. Y. A fully decentralized multisensor system for tracking and surveillance [J]. *The Int J of Robotics Research*, 1988, 12(1): 20-45.
- [4] Luo R. C., Lin M., Scherp P. C. Dynamic multisensor data fusion system for intelligent robots [J]. *IEEE J of Robotics and Automation*, 1988, 4(4): 386-396.
- [5] Bloch I. Information combination operators for data fusion: A comparative review with classification [J]. *IEEE Trans on System Man and Cybernetic: Part A*, 1996, 26(1): 52-67.
- [6] 李洪兴, 汪培庄. 模糊数学 [M]. 北京: 国防工业出版社, 1994. 62-64.