

# 多传感器数据融合的主成分方法研究

董九英

DONG Jiu-ying

江西财经大学 信息管理学院, 南昌 330013

College of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330013, China

E-mail: jiuyingdong@126.com

**DONG Jiu-ying. Study on principle component method for multi-sensor data fusion. Computer Engineering and Applications, 2009, 45(33): 111-113.**

**Abstract:** Due to data fusion of multi-sensor experiment on some characteristic index, a new fusion method is proposed based on the principle component analysis. The method views the measured data of every sensor as a variant. After defining each principle components for the collectivity, the synthesis support degrees of all sensors are given according to the compound relationship between the measured value and the principle component. The formula of data fusion is obtained. The applied example proves that the method is both effective and accurate.

**Key words:** multi-sensor; data fusion; principle component analysis; synthesis support degree

**摘要:** 针对多个传感器对某一特性指标进行测量实验的数据融合问题, 提出了一种基于主成分分析的融合方法。该方法把各传感器的测量数据作为一变量, 定义总体的各主成分, 利用测量值与主成分的复相关关系, 给出了各传感器的综合支持程度和数据融合公式。应用实例验证了方法的有效性和精确性。

**关键词:** 多传感器; 数据融合; 主成分分析; 综合支持度

**DOI:** 10.3778/j.issn.1002-8331.2009.33.036 **文章编号:** 1002-8331(2009)33-0111-03 **文献标识码:** A **中图分类号:** TP212

## 1 引言

为了降低单个传感器由于干扰或不好的环境所造成的不确定性, 传感器信息融合一般采用三种方式: 单传感器多测量周期、多传感器单测量周期和多个传感器多测量周期进行信息融合<sup>[1-7]</sup>。

在数据采集与信号处理系统中, 经常要利用多个传感器对同一目标进行检测, 并运用合理的算法融合多个传感器的观测信息, 这是一种典型的多传感器多测量周期融合方式。在融合算法方面, 文献[5-7]分别给出了基于支持度、相似度、信噪比的融合方法。然而, 文献[5-6]中的指数衰减系数、最终加权系数中的可调参数的选择由人为给定, 受主观因素的影响较大。选择不同的参数值将导致不同的融合结果。

为此, 运用多元统计分析中的主成分分析思想, 给出了一种新的融合方法。

## 2 数据融合

### 2.1 主成分定义和性质

主成分分析是将研究对象的多个相关变量(指标)化为少数几个不相关的变量的一种多元统计方法。

**定义 1** 设总体  $X=(x_1, x_2, \dots, x_n)$  的协方差阵为  $\Sigma$ , 其特征

根为  $\lambda_1 \geq \dots \geq \lambda_n$ , 对应的单位化正交特征向量为  $e_1, e_2, \dots, e_n$ , 其中  $e_i=(e_{i1}, e_{i2}, \dots, e_{in})$ , 则第  $i$  个主成分为<sup>[8]</sup>:

$$y_i = e_i X' = e_{i1}x_1 + e_{i2}x_2 + \dots + e_{in}x_n \quad (1)$$

从代数的观点来看, 主成分是  $n$  个变量  $x_1, x_2, \dots, x_n$  的一种特殊的线性组合, 它仅依赖于总体的协方差阵, 并不要求总体是正态分布的。定义 1 中, 当总体的协方差阵未知时, 可以用样本的协方差矩阵或相关矩阵代替。

**性质 1** 主成分满足如下性质:

- (1)  $y_i$  与  $y_j (i \neq j)$  不相关。
- (2)  $y_1$  是  $X$  的一切线性组合中方差达到最大的;  $y_2$  是与  $y_1$  不相关的一切  $X$  的线性组合中方差达到最大的,  $\dots$ ,  $y_i$  是与  $y_1, y_2, \dots, y_{i-1}$  不相关的一切  $X$  的线性组合中方差达到最大的。
- (3) 第  $i$  个主成分对应于第  $i$  大特征根  $\lambda_i$  的单位化正交特征向量, 其方差为  $\lambda_i$ 。

### 2.2 主成分融合方法

基于主成分分析的思想提出一种新的融合方法, 先叙述有关的定理。

**定义 2** 在主成分分析中, 通过定义主成分  $y_i$  的方差贡献率

$$a_i = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j}, \quad i=1, 2, \dots, n \quad (2)$$

**基金项目:** 国家自然科学基金(the National Natural Science Foundation of China under Grant No.10626029); 江西省自然科学基金(the Natural Science Foundation of Jiangxi Province of China under Grant No.0611082); 江西省教育厅科技项目(No.GJJ08350)。

**作者简介:** 董九英(1974-), 女, 讲师, 研究方向: 信息融合, 数据处理。

**收稿日期:** 2008-12-04 **修回日期:** 2009-02-13

来说明各主成分  $y_i$  概括原变量  $X$  信息的大小。称  $\sum_{j=1}^k a_j$  为前  $k$  个主成分的累计方差贡献率。

主成分个数  $k$  的确定一般情况下取  $k$  使得累计方差贡献率  $\sum_{j=1}^k a_j$  达到 80% 以上。

**定理 1** 总体  $X$  的原变量  $x_j$  与主成分  $y_j$  的相关系数为<sup>[8]</sup>:

$$\rho(x_i, y_j) = \frac{e_{ij} \sqrt{\lambda_j}}{\sigma_i} \quad (3)$$

其中,  $\sigma_i$  为  $x_i$  的标准差,  $e_{ij}$  为特征向量  $e_j$  的第  $j$  个分量。

**定理 2** 设  $(y_1, y_2, \dots, y_k)$  为总体  $X$  的前  $k$  个主成分, 则  $x_j$  与  $(y_1, y_2, \dots, y_k)$  的复相关系数的平方<sup>[8]</sup>

$$\tau_j = \rho^2[x_j, (y_1, y_2, \dots, y_k)] = \sum_{i=1}^k \rho^2(y_i, x_j) \quad (4)$$

因为  $\tau_j$  表示  $k$  个主成分  $(y_1, y_2, \dots, y_k)$  对原变量  $x_j$  的贡献率, 而这  $k$  个主成分  $(y_1, y_2, \dots, y_k)$  概括了原变量  $X$  的 80% 以上的信息, 所以在某种程度上,  $\tau_j$  可以看成是变量  $x_j$  在总体  $X$  的信息中所占的比重, 可以利用  $\tau_j$  来定义各传感器的综合支持程度。

**定义 3** 定义第  $i$  个传感器的综合支持程度

$$z_i = \frac{\tau_i}{\sum_{l=1}^n \tau_l} \quad (5)$$

最后得到所有传感器的融合数据为:

$$x = \sum_{i=1}^n z_i x_i \quad (6)$$

式(6)即为该文给出的多传感器数据融合公式。

### 3 实验

以文献[5]中的实验来说明该文的方法。实验采用三个热电偶对恒温箱温度检测, 经过 6 次测量, 得到如表 1 所示的观测值:

表 1 文献[5]中的测量值(真值 900°C)

传感器	观测次数					
	1	2	3	4	5	6
1	899.5	905.3	901.9	900.6	889.9	899.4
2	898.3	875.9	888.1	886.2	907.5	904.4
3	896.7	906.8	898.2	904.0	896.4	891.6

在此  $n=3$ , 总体  $X=(x_1, x_2, x_3)$ 。由于总体的协方差矩阵  $\Sigma$  未知, 可以由样本的协方差矩阵  $S$  或相关系数矩阵  $\hat{R}$  代替。利用 Matlab6.5.1, 得到样本的协方差和相关系数矩阵如下:

$$S = \begin{pmatrix} 26.55 & & \\ -51.72 & 146.00 & \\ 14.90 & -59.52 & 30.66 \end{pmatrix}, \hat{R} = \begin{pmatrix} 1 & & \\ -0.84 & 1 & \\ 0.52 & -0.89 & 1 \end{pmatrix}$$

计算相关系数矩阵的特征值和对应的单位正交化特征向量, 再利用式(3)得到特征值的累计贡献率, 如表 2 所示。

表 2 特征根和特征向量

特征值	特征向量	累计贡献率/(%)
2.605 7	(0.540 7, -0.628 6, 0.559)	83.52
0.479 7	(-0.743 5, -0.046 3, 0.667 2)	99.51
0.014 7	(0.393 5, 0.776 3, 0.492 4)	100

由表 2 可看到前两个特征值的累计贡献率已达 99.51%,

故选取两个样本主成分, 分别为  $y_1=0.5, y_2=1.5$ 。

由式(4)得到变量  $x_1, x_2, x_3$  与主成分  $y_1, y_2$  的复相关系数的平方, 由式(5)得到各传感器的综合支持程度, 如表 3 所示。

表 3 相关系数、复相关系数、支持度

$\rho(y_i, x_j)$	$x_1$	$x_2$	$x_3$
$y_1$	0.166 1	-0.082 3	0.159 8
$y_2$	-0.099 9	-0.002 7	0.083 5
$\tau_j$	0.037 6	0.006 8	0.032 5
$z_j$	0.488 9	0.088 4	0.422 6

最后, 利用式(5)和(6)得到融合后的数据如表 4 所示。同时, 在表 4 中也列出了平均值法和文献[5]可靠性法的融合结果以及绝对误差的比较。

表 4 三种方法融合结果的比较

观测次数	1	2	3	4	5	6
主成分融合	898.21	903.33	899.12	900.76	894.20	896.55
绝对误差	1.79	3.33	0.88	0.76	5.80	3.45
平均值融合	898.17	896.00	896.07	896.93	897.93	898.34
绝对误差	1.73	4.00	3.93	3.07	2.07	1.66
可靠性融合	898.24	896.13	896.06	896.86	897.92	898.48
绝对误差	1.76	3.87	3.94	3.14	2.08	1.52

从表 4 可知, 该文的主成分融合法分别在测量次数  $k=1, 2, 3, 4$  时比平均值法融合的结果更接近真值 900, 相应的绝对误差较小; 同时主成分融合分别在  $k=2, 3, 4$  时比可靠性融合的结果更接近真值, 相应的绝对误差较小。三种方法 6 次融合结果的总绝对误差反映在表 5 中。

表 5 三种方法的总绝对误差

	主成分法	平均值法	可靠性法
总绝对误差	16.01	16.56	16.31

表 5 表明主成分法比平均值法、可靠性法融合的总绝对误差都小。为详细比较三种方法的融合效果, 将实验次数提高到 100 次, 仿真结果如图 1 所示。

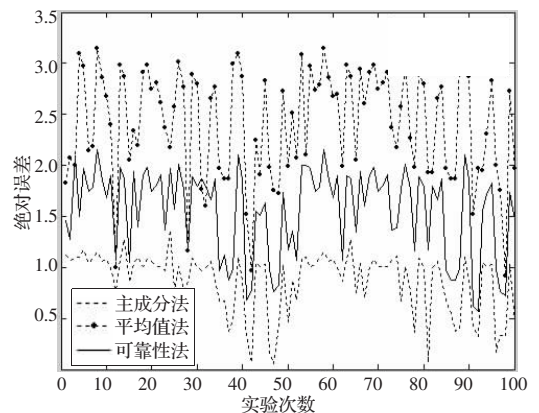


图 1 100 次实验的仿真结果

从图 1 的绝对误差变化曲线来看, 与平均法、可靠性法相比, 该文的主成分法仍显示出较好的精确性和稳健性, 并经多次的重复仿真仍表现出较好的鲁棒性。

由上述分析可知, 该文的主成分法更精确、稳健。这主要是因为采用主成分不仅能避免信息的损失, 而且各变量与主成分的复相关系数度量了各变量在所有测量信息中所占的不同权

重。平均值法不分优劣地利用数据,没有考虑各传感器之间融合的不同权重(而是采用相等权重)。可靠性法虽然在融合过程中考虑到观测值在整个时间轴上的可靠性,但由于其衰减系数和可调参数的选取均具有非常大的主观因素,影响了算法的稳健性<sup>[7]</sup>。

#### 4 与已有方法的比较

提出的主成分融合方法同已有文献相比,具有如下优点:

(1)不需要定义基本概率指派函数,不同于文献[3-4]的证据理论方法,事实上基本概率指派较难确定。

(2)不需要定义受主观因素影响的关系矩阵,不同于文献[5-6]。也不需要定义贴近度函数,不同于文献[7]。

(3)利用主成分不仅可以避免有效数据的损失,而且可以大大减少信息融合过程中的计算量。例如,在上述实验中只选取两个样本主成分就可以概括了原变量的 99.51% 以上的信息,无需再选取第 3 个主成分。

#### 5 结束语

针对多个传感器对某一特性指标进行测量实验的数据融合问题,提出了一种基于主成分分析的数据融合方法。该方法视各传感器的测量值为一变量,利用各测量值与主成分的复相

(上接 104 页)

其中,系数矩阵  $A$  如式(13)所示。解方程(16),并利用约束条件式(8),可得稳态解为:

$$w_{s1} = \dots = w_{sn_1} = \frac{\alpha_1(1-\beta_2)}{n_1\alpha_1(1-\beta_2) + n_2\alpha_2(1-\beta_1)} (\gamma C t_p - 0.5q) \quad (17)$$

$$w_{sn_1+1} = \dots = w_{sn} = \frac{\alpha_2(1-\beta_1)}{n_1\alpha_1(1-\beta_2) + n_2\alpha_2(1-\beta_1)} (\gamma C t_p - 0.5q) \quad (18)$$

由式(17)、(18)可知,在稳态时,同一权限的各 VCP 流达到公平收敛;不同权限的各 VCP 流按照其权限收敛,即  $w_{si}(i=1, 2, \dots, n_1)$  与  $w_{sj}(j=n_1+1, \dots, n)$  的比值为:

$$\frac{w_{si}}{w_{sj}} = \frac{\alpha_i(1-\beta_1)}{\alpha_j(1-\beta_2)} \quad (19)$$

从式(19)知,若选择  $\alpha_1 > \alpha_2, 0 < \beta_2 < \beta_1 < 1$ , 则拥塞窗口为  $w_{si}(i=1, 2, \dots, n_1)$  的 VCP 流比拥塞窗口为  $w_{sj}(j=n_1+1, \dots, n)$  的 VCP 流,在稳态时占用更多的网络资源。

对于系统矩阵  $A$  的  $(n_1-1)$  个重特征值  $\beta_1$ 、 $(n_2-1)$  个重特征值  $\beta_2$  以及单特征值  $\beta^*$ , 由于  $0 < \beta_2 < \beta^* < \beta_1 < 1$ , 所以,它们都在单位圆内,相应的 AIMD 算法的动态过程是渐近稳定的,系统收敛于唯一的稳态解式(17)、(18)。AIMD 算法的动态过程的收敛特性由最大的特征值  $\beta_1$  所界定。很明显,  $\beta_1$  越接近于 1, 则收敛的动态过程越长。

#### 5 仿真

利用网络仿真软件 ns2<sup>[8]</sup>验证 VCP 算法中参数  $\alpha, \beta$  对加权公平性的影响。仿真采用单瓶颈的网络拓扑结构,瓶颈链路带宽为  $C=45$  Mb/s。有五个 VCP 数据流,其单向延迟为 40 ms,数据包大小为 1 000 Bytes。五个数据流分为两个等级:一个数据流属于高优先级,设置  $\alpha_1=1.5, \beta_1=0.9375$ ;四个数据流属于低优先级,设置  $\alpha_2=1.0, \beta_2=0.875$ 。低优先级的四个数据流首先启动,达到稳定状态后,高优先级的一个数据流加入系统,观察该高优先级数据流拥塞窗口的收敛过程。仿真结果见图 2。由图 2 可知,系统稳定后,高优先级数据流的拥塞窗口值约为低优

关关系,定义各传感器的综合支持程度,得到传感器数据的融合公式。可以充分利用测量信息,避免有效数据的损失,提高了融合结果的精度。

#### 参考文献:

- [1] 舒红平,王运锋,徐振明,等.基于  $K$ -均值聚类的多雷达数据融合算法研究[J].计算机工程与应用,2007,43(16):183-186.
- [2] Hong L, Lynch A. Recursive temporal-spatial information fusion with applications to target identification[J]. IEEE Transactions on Aerospace and Electronic Systems, 1993, 29: 435-445.
- [3] Bogler L P. Shafer-dempster reasoning with application to multi-sensor target identification system[J]. IEEE Trans on System, Man and Cybernetics, 1987, SMC-17(6): 968-977.
- [4] 李茹,李弼程. D-S 证据理论的改进算法在时-空信息融合中的应用[J]. 数据采集与处理, 2005, 20(4): 407-410.
- [5] 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): 537-539.
- [6] 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.
- [7] 万树平. 基于信噪比的多传感器数据融合方法[J]. 传感技术学报, 2008, 21(1): 178-181.
- [8] 孙文爽,陈兰祥. 多元统计分析[M]. 北京:高等教育出版社,1994: 345-367.

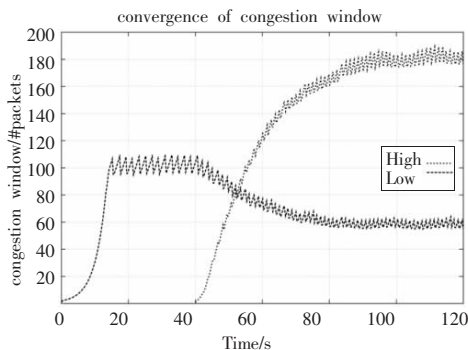


图2 高优先级数据流收敛过程

优先级数据流拥塞窗口值的三倍左右,这验证了式(19)的正确性。

#### 6 结语

对 VCP 协议加权公平性的动态过程进行了分析,指出只要合适地选择 AIMD 算法中的加性参数和乘性参数,就可使 VCP 流按权限不同而占用不同的网络资源。

应指出,在建立 VCP 协议加权公平性收敛的动态模型时,对 MD 的回退过程的细节作了一些简化处理<sup>[3-4]</sup>,但这不影响加权分配网络资源的主要结论。

#### 参考文献:

- [1] Jacobson V. Congestion avoidance and control[C]//ACM SIGCOMM'88, 1988. New York: ACM Press, 1988: 314-329.
- [2] Katabi D. Decoupling congestion control and bandwidth allocation policy with application to high bandwidth-delay product networks[D]. Massachusetts Institute of Technology, 2003.
- [3] Xia Y, Subramanian L, Stoica I, et al. One more bit is enough[C]//ACM SIGCOMM'05, 2005. New York: ACM Press, 2005: 37-48.
- [4] Xia Y, Subramanian L, Stoica I, et al. One more bit is enough[R]. UC Berkeley, 2005-06.
- [5] The network simulator ns-2.30[EB/OL]. <http://www.isi.edu/nsnam/ns>.