

不同精度冗余数据的融合¹⁾

唐 璘^{1,2} 张闻捷¹ 高 琰¹ 付明玉³ 蔡自兴¹

¹(中南大学信息科学与工程学院 长沙 410083)

²(国防科技大学通信与信息系统博士后流动站湘计算机工作站 长沙 410007)

³(哈尔滨工程大学动力与核能工程学院 哈尔滨 150001)

(E-mail: tjin@mail.csu.edu.cn)

摘 要 针对融合误差的最大值和数学期望, 提出了一个评判数据融合方法优劣的标准. 随后, 提出了一种新的数据融合方法, 扩展加权平均法. 当待融合数据为两个时, 通过理论分析得到了计算融合参数的公式. 当有更多的数据参与融合时, 通过数值仿真得到了该方法的各个融合参数. 该方法能解决最大似然估计法所难以解决的均匀分布数据的融合问题, 且具有比包括最大似然估计法在内的其它三种有代表性的数据融合方法更高的精度.

关键词 数据融合, 加权平均, 均匀分布, 最小期望, 蒙特卡洛法

中图分类号 TP212

Fusion of Different Accuracy Redundant Data

TANG Jin^{1,2} ZHANG Wen-Jie¹ GAO Yan¹ FU Ming-Yu³ CAI Zi-Xing¹

¹(School of Information Science and Engineering, Central South University, Changsha 410083)

²(Hunan Computer Co. Ltd, National University of Defense Technology, Changsha 410007)

³(Power and Nuclear College of Harbin Engineering University, Harbin 150001)

(E-mail: tjin@mail.csu.edu.cn)

Abstract This paper presents criteria to evaluate different approaches of redundant data fusion; the criteria mostly concern the expectation and the maximum of the fusion error. A new fusion approach for multiple data is also presented, as the data have different accuracies. This approach is an extension of the weighted average. Through theoretic analysis, we obtain the formula to calculate the parameters for fusion of two uniform distribution data. Through Monte Carlo method, we get the parameters when we fuse more data or other distribution data. Our extended weighted method can fuse uniform data that cannot be handled by maximum likelihood approach. Its result is compared with other three representative fusion algorithms: maximum likelihood, optimal weighted average, and HILARE method. Comparison shows that our approach is better than all weighted average approaches; it has the smallest expectation and smallest maximum of the fusion error in all the four approaches.

Key words Data fusion, weighted average, uniform distribution, minimum expectation, Monte Carlo

1) 国家自然科学基金 (60234030), 教育部留学回国人员科研启动基金项目资助

Supported by National Natural Science Foundation of P. R. China (60234030), and the Project Sponsored by SRF for ROCS, SEM

收稿日期 2004-11-11 收修改稿日期 2005-7-30

Received November 11, 2004; in revised form July 30, 2005

1 前言

数据融合有机地组合了多个传感器的数据, 给出更为合理的结果^[1,2]. 它有许多的优点, 如改善观察能力和观察过程, 更精确地估计观察结果等, 正被应用在越来越多的领域^[3,4], 如遥感^[5~8] 和医学^[9]. 数年前, 不同精度的冗余数据的融合已有一些研究结果^[10]. 如, HILARE 项目采用了加权平均法, 如果所有数据的方差在同一个数量级就对数据采用加权平均法, 否则就直接使用方差最小的数据. 文献 [11] 则记载了一种利用最大似然估计推出的直接对待融合数据进行加权平均的方法.

近来, 研究者大多致力于互补数据的融合上, 文献 [12] 中有很多的方法. 文献 [13,14] 则讨论了一些基于概率、最大似然估计、加权平均法等对冗余数据进行融合的方法. 当待融合数据是正态分布的随机变量时, 最大似然估计可很好地应用. 但当数据为其它的分布(如: 均匀分布)时, 最大似然估计就不一定能很好工作了. 更为重要的是, 尽管能找到不少数据融合方法, 但难以找到一个评判各种方法优劣的标准.

因此本文第 2 节提出了一个评判不同方法优劣的简单标准; 第 3 节提出了一种扩展加权平均法, 它能融合两个均匀分布的冗余数据, 通过理论分析获得了其融合参数; 第 4 节通过比较, 显示该方法要好于其它三种加权平均法; 第 5 节把本文提出的扩展加权平均法推广到其它分布和多个待融合数据的情形, 通过蒙特卡洛法确定了融合参数, 并与其它方法进行了对比.

2 融合方法优劣的评判标准

设要把 n 个冗余数据融合为一个数据. 设原数据为 X'_1, X'_2, \dots, X'_n , 它们是不同的传感器对同一物体的同一物理量进行测量得到的. X'_i ($i = 1, 2, \dots, n$) 均为随机变量. 融合后的数据为 Y' . 显然, Y' 也是一个随机变量. 我们希望 Y' 比所有的 X'_i 都更精确.

所有传感器都可以在测量之前经过仔细校准以消除系统误差, 于是 X'_i ($i = 1, 2, 3, \dots, n$) 都没有系统误差, 即: $E[X'_i] = \theta$, θ 是被测量的真实值. 令 $X_i = X'_i - \theta$ ($i = 1, 2, 3, \dots, n$) 以及 $Y = Y' - \theta$. 不失一般性, 只需要把 X_i ($i = 1, 2, 3, \dots, n$) 融合成 Y 即可. 此时有 $E[X_i] = 0$ ($i = 1, 2, 3, \dots, n$) 以及 $E[Y] = 0$.

有许多方法能把 X_i ($i = 1, 2, 3, \dots, n$) 融合成 Y ^[10~12], 但缺乏一个简单的标准来判别这些方法的优劣. 本文提出两条标准用于判别, 它们可同时或分别使用. 两条标准如下: 1) $|Y|$ 的期望值 $E[|Y|]$ 越小, 方法越好. 2) $|Y|$ 的最大值 $\text{Max}|Y|$ 越小, 方法越好.

3 两个数据的融合

设只有 X_1 和 X_2 被融合, 它们均为随机变量. X_1 是 $[-1, 1]$ 上的均匀分布, X_2 是 $[-h, h]$ 上的均匀分布, $h > 1$. 任何两个均匀分布的数据都可以简化成为这种形式.

通常的融合两个数据的加权平均法的形式是 $Y = kX_1 + (1 - k)X_2$, 这是独立于上下文的定常 (Context independent constant behavior, CICB) 融合算子^[15], 其融合参数不随待融合数据的变化而变化. 而独立于上下文的异动 (Context independent variable behavior, CIVB) 融合算子可能有更好的性能^[15], 该类算子的融合参数随待融合数据的变化而变化. 本文融合方法属于 CIVB, 这里称之为扩展加权平均法. 其融合策略如下

$$\begin{cases} Y = X_1 & \text{if } |X_1 - X_2| < C \\ Y = kX_1 + (1 - k)X_2 & \text{if } |X_1 - X_2| > C \end{cases} \quad C > 0, \text{ 且 } 0 < k < 1$$

参见图 1, 这类似于 X_1 和 X_2 的概率联合分布图. C 和 k 为两个变量, 与 h 的大小有关. 它们将通过最小化 $|Y|$ 的期望值 $E[|Y|]$ 来被确定.

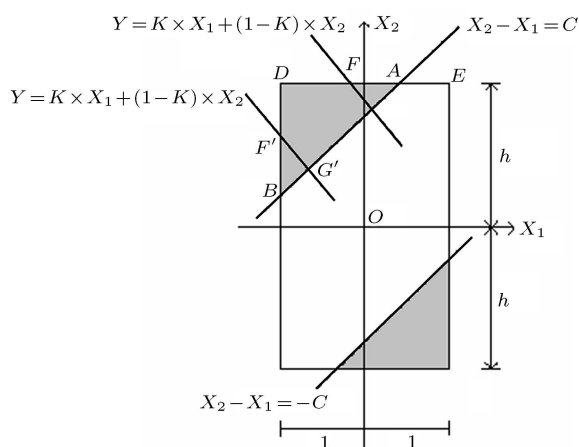


图 1 两个数据的融合

Fig. 1 Fusion of two data

首先应计算出整个区域的 $E[|Y|]$. 此时假设直线 $X_2 - X_1 = C$ 位于点 E 的左边. 后面将会看到使 $E[|Y|]$ 达到极小值的直线 $X_2 - X_1 = C$ 确实位于点 E 的左边.

3.1 白色区域的 $E[|Y|]$

记 $E_1[|Y|]$ 为白色区域的 $|Y|$ 的数学期望值. 由于对称性, 只需讨论 $X_2 > 0$ 的情形. 即使当点 B 位于 X_1 轴下面时, 也是如此. 点 A 坐标为 $(h - c, h)$, 点 B 坐标为 $(-1, c - 1)$. 记 $S = 4h$, 它是整个区域的面积. 当 $-1 < Y < h - c$ 时, 概率密度函数为 $f(Y) = (C + Y)/4h$. 当 $h - C < Y < 1$ 时, 概率密度函数为 $f(Y) = h/4h = 1/4$.

1) 当 $h < C$ 时: 此时意味着点 A 在 X_2 轴的左边, 有

$$E_1[|Y|] = 2\left(-\int_{-1}^{h-c} Y \frac{C+Y}{4h} dY - \int_{h-c}^0 \frac{Y}{4} dY + \int_0^1 \frac{Y}{4} dY\right) = \frac{(h-c)^3}{12h} + \frac{C}{4h} + \frac{1}{4} - \frac{1}{6h} \quad (1)$$

2) 当 $h > C$ 时: 此时意味着点 A 在 X_2 轴的右边, 有

$$E_1[|Y|] = 2\left(-\int_{-1}^0 Y \frac{C+Y}{4h} dY - \int_0^{h-c} \frac{Y}{4} dY + \int_{h-c}^1 \frac{C+Y}{4h} dY\right) = -\frac{(h-C)^3}{12h} + \frac{C}{4h} + \frac{1}{4} - \frac{1}{6h} \quad (2)$$

3.2 灰色区域的 $E[|Y|]$

记 $E_2[|Y|]$ 为灰色区域的 $|Y|$ 的数学期望值. 在点 A、点 B、点 D 处, 分别有: $Y(A) = k(h - C) + (1 + k)h = h - kC$, $Y(B) = k \times (-1) + (1 - k)(C - 1) = C - kC - 1$, $Y(D) = h(1 - k) - k = h - (h + 1)k$, 易得, $Y(B) < Y(D) < Y(A)$.

点 F' 的坐标为 $(-1, (Y + k)/(1 - k))$, 点 G 和点 G' 的坐标均为 $(Y - (1 - k)C, Y + kC)$, 点 F 坐标为 $((Y - h)/k + h, h)$. $\Delta BG'F'$ 、 ΔBAD 、 ΔGAF 的面积分别可计算得到如下

$$S_{\Delta BG'F'} = \frac{1}{2(1-k)}(Y + kC + 1 - C)^2 \quad (3)$$

$$S_{\Delta BAD} = \frac{1}{2}(h - c + 1)^2 \quad (4)$$

$$S_{\Delta GAF} = \frac{1}{2k}(Y - kC - h)^2 \quad (5)$$

于是, 当 $Y \in [C - kC - 1, h - k(h + 1)]$ 以及当 $Y \in [h - k(h + 1), h - kC]$ 时, 概率分布函数分别为

$$P_1(Y) = \frac{1}{4h} S_{\Delta BG'F'} = \frac{1}{8(1-k)h} (Y + kC + 1 - C)^2 \quad (6)$$

$$P_1(Y) = \frac{1}{4h} (S_{\Delta BAD} - S_{\Delta GAF}) = \frac{1}{8h} [(h - C + 1)^2 - \frac{1}{k} (Y + kC - h)^2] \quad (7)$$

于是可得概率密度函数如下

$$f_1(Y) = \frac{1}{4(1-k)h} (Y + kC + 1 - C) \quad (8)$$

$$f_2(Y) = \frac{1}{4kh} (h - Y - kC) \quad (9)$$

$E2[|Y|]$ 由 k 和 C 共同决定, $E1[|Y|]$ 仅由 C 决定.

直线 $Y = k \times X_1 + (1 - k) \times X_2$ 把三角形区域 ΔBAD 划分为两部分, 让该直线通过点 D , 参见图 2. 可得

$$k = \frac{h}{h + 1} \quad (10)$$

及

$$\begin{cases} S_{\Delta BGD} = \frac{1}{2(1-k)} (kC + 1 - C)^2 \\ S_{\Delta DGA} = \frac{1}{2k} (kC - h)^2 \end{cases} \quad (11)$$

把 (10) 式代入 (11) 式, 可得 $S_{\Delta DGA} = hS_{\Delta BGD}$, 因 $h > 1$, 故

$$S_{\Delta DGA} > S_{\Delta BGD} \quad (12)$$

由于两边面积不等, 可以推测为了使 $E2[|Y|]$ 达到极小, 直线 $Y = k \times X_1 + (1 - k) \times X_2 = 0$ 应当位于点 D 和点 A 之间. 此时意味着

$$\begin{cases} C - kC - 1 < 0 \\ h - k(h + 1) < 0 \\ h - kC > 0 \end{cases} \quad \text{即} \quad \begin{cases} k > \frac{C-1}{C} \\ k > \frac{h}{h+1} \\ k < \frac{h}{C} \end{cases} \quad (13)$$

于是有

$$E2[|Y|] = 2 \left[\int_{C-kC-1}^{h-k(h+1)} \frac{1}{4(1-k)h} (Y + kC + 1 - C)(-Y) dY + \int_{h-k(h+1)}^0 \frac{1}{4kh} (h - Y - kC)(-Y) dY + \right]$$

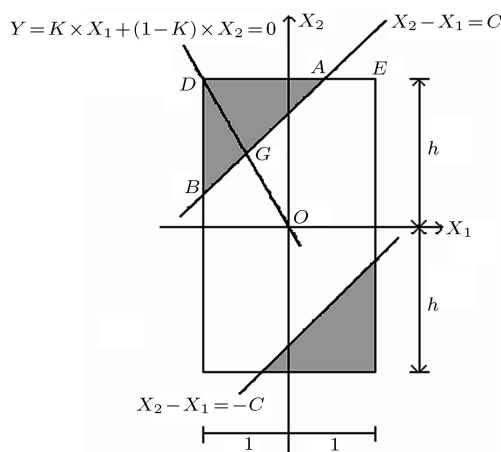


图 2 直线 OD 分割三角形区域 ΔDAB
Fig. 2 Triangle DAB divided by line OD

$$\int_0^{h-hC} \frac{1}{4hk} (h-Y-kC)Y dY = \frac{(h-kC)^3}{12kh} + \frac{(kC+1-C)^3}{12(1-k)h} + \frac{(h-(h+1)k)^3}{12k(1-k)h} \quad (14)$$

3.3 如何确定 C 和 k

可以通过对整个区域的 $E[|Y|]$ 取极小值来确定 C 和 k . 即

$$\begin{cases} \frac{d(E_1[|Y|] + E_2[|Y|])}{dk} = 0 \\ \frac{d(E_1[|Y|] + E_2[|Y|])}{dC} = 0 \end{cases} \quad (15)$$

因为 $E_1[|Y|]$ 与 k 无关, 因此当 $h > C$ 时以及 $h < C$ 时, (15) 式的第一式相同. 即

$$\frac{d(E_1[|Y|] + E_2[|Y|])}{dk} = \frac{d(E_2[|Y|])}{dk} = 0$$

于是

$$\begin{aligned} & k^2(2C^3k^3 + (-6C^3 + 3C^2)k^2 + (6C^3 - 6C^2)k - 2C^3 + 3C^2 - 1) + 2C^3k^5 + \\ & k^4(-4C^3 - 3hC^2) + k^3(2C^3 + 6hC^2) + k^2(-3hC^2 + h^3) - 2h^3k + h^3 = \\ & k^4(h^3 + 3h^2 + 3h + 1) - k^3(2h^3 + 6h^2 + 6h + 2) + k^2(3h^2 + 3h) + 2h^3k - h^3 \end{aligned}$$

解这个关于 k 的 5 阶方程, 可以得到出两个解 $k_1 = k_2 = 1$ 和一个关于 k 的三阶方程

$$4C^3k^3 + k^2(-2C^3 + 3C^2 - 3hC^2 - h^3 - 3h^2 - 3h - 1) + 2h^3 = 0 \quad (16)$$

现在来关注 (15) 式的第二个方程, 此时需要区分 $h > C$ 和 $h < C$ 两种情形:

1) 当 $h > C$ 时, 整个区域的 $E[|Y|]$ 的极小值最有可能产生于这种情形. (15) 式的第二个方程为

$$\frac{d(E[|Y|])}{dC} = \frac{(h-C)^2}{4h} + \frac{1}{4h} - \frac{(h-kC)^2}{4h} - \frac{(C-kC-1)^2}{4h} = 0$$

上面方程的解为

$$\begin{cases} C_1 = 0 \\ C_2 = \frac{h-1}{k} \end{cases} \quad (17)$$

C_1 没有意义, 不予考虑. 在只考虑 $C = C_2$ 的情形下, 因为 $k < 1$, 因此 $C > h-1$.

当直线 $X_2 - X_1 = C$ 通过点 E 时, $C = h-1$, 参见图 1. $C > h-1$ 意味着直线位于点 E 的左边, 在前面计算 $E[|Y|]$ 之前, 实际上已经假设了这一点. 也即, 得到的解答 (17) 满足该假设. 此外, 根据 (13) 式 $k > (C-1)/C$ 和 (17) 式 $k = (h-1)/C$, 可以推得 $(C-1)/C < (h-1)/C$. 于是可得 $h > C$. 这就是上面求得 (17) 这个解答时假设的情形.

这些意味着如果 k 满足条件 (13), 则解答 $C = (h-1)/k$ 将满足所有的条件和假设. 换言之, (17) 式中的 C 是合理的. 不过, 这里没有办法断定 k 是否一定满足所有条件和假设, 只有在进行数据融合时, 根据具体的待融合数据来进行判断.

把 $C = (h-1)/k$ 代入 (16) 式, 得

$$4\left(\frac{h-1}{k}\right)^3k^3 + k^2\left(-2\left(\frac{h-1}{k}\right)^3 + 3\left(\frac{h-1}{k}\right)^2 - 3hc^2 - h^3 - 3h^2 - 3h - 1\right) + 2h^3 = 0$$

则, $k^3(h+1)^3 - k((h-1)^3 + 2h^3) + 2(h-1)^3 = 0$, 于是可以通过下面的方程组来确定 k 和 C ,

$$\begin{cases} k^3(h+1)^3 - k((h-1)^3 + 2h^3) + 2(h-1)^3 = 0 \\ C = \frac{h-1}{k} \end{cases} \quad (18)$$

2) 当 $h < C$ 时

$$\frac{d(E[|Y|])}{dC} = -\frac{(C - kC - 1)^2}{4h} - \frac{(h - kC)^2}{4h} - \frac{(h - C)^2}{4h} + \frac{1}{4h} = 0$$

于是

$$C_{1,2} = \frac{(1 - k + hk + h) \pm \sqrt{-3h^2k^2 + 1 + k^2 - 3h^2 - 2hk^2 + 6h^2k - 2k + 2h}}{2(k^2 - k + 1)} \quad (19)$$

此时可用 (16) 式、(19) 式求得 k 和 C , 但无法断定此时的 C 是否为实数, 更无法确定 C 和 k 是否有意义.

可对第 3 节总结如下: 给定 h , 可以通过 (18) 式或者结合 (19) 式和 (16) 式求出最优的 k 和 C . 实际上, 在后面所有的计算例子中有意义的解总是来自 (18) 式.

4 各种方法的比较

这里把扩展加权平均法与其它三种方法进行比较: 1) 第 1 节中描述的 HILARE 方法; 2) 通常的最优加权平均法; 3) 强迫最大似然估计法^[11]. 当数据的概率为均匀分布时, 最大似然估计法难以被直接应用, 这里只是用了文献 [11] 中的公式. 它不是真正的最大似然估计, 这里姑且称之为强迫最大似然估计法.

现在来最优化通常的加权平均法. 其公式如下: $Y = kX_1 + (1 - k)X_2$, 需要求出当 k 为何值时, 该方法达到最优. 其中 Y, k, X_1 和 X_2 的定义与第 3 节中的定义一样. 这里通过最小化 $E[|Y|]$ 来求得 k . 首先计算概率分布函数, 然后计算 $E[|Y|]$, 再令 $\frac{d(E[|Y|])}{dk} = 0$, 即得到最优的 k . 这个过程与第 3 节中的计算过程很类似. 最后可得结果如下

$$\begin{cases} k = \frac{h}{\sqrt{h^2 + 3}} \\ E[|Y|] = \frac{1}{12hk(1-k)}((k + h - kh)^3 - (k + kh - h)^3) \end{cases} \quad (20)$$

h 的定义与第 3 节中的定义相同.

各种方法的融合精度和融合参数的比较如表 1 所示. R 为其它方法的 $E[|Y|]$ 与扩展加权平均法的 $E[|Y|]$ 之比

表 1 各种方法的比较
Table 1 Comparison of some approaches

	h	1.0	1.5	2.0	2.5	3.0	3.5
扩展加权平均法	$E[Y]$	0.3333	0.3923	0.4217	0.4387	0.4497	0.4574
	$\text{Max} Y $	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	K	0.5000	0.6443	0.7264	0.7780	0.8132	0.8388
	C	0.0000	0.7760	1.3767	1.9281	2.4594	2.9803
HILARE 法	$E[Y]$	0.3333	0.5000	0.5000	0.5000	0.5000	0.5000
	$\text{Max} Y $	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	R	1.0000	1.2745	1.1857	1.1397	1.1119	1.0931
最优加权平均法	$E[Y]$	0.3333	0.3956	0.4305	0.4512	0.4641	0.4726
	$\text{Max} Y $	1.0000	1.1727	1.2441	1.2670	1.2679	1.2594
	K	0.5000	0.6547	0.7559	0.8220	0.8660	0.8963
	R	1.0000	1.008	1.0209	1.0285	1.0320	1.0332
强迫最大似然估计	$E[Y]$	0.3333	0.3974	0.4333	0.4540	0.4667	0.4748
	$\text{Max} Y $	1.0000	1.1538	1.2000	1.2069	1.20007	1.1887
	K	0.5000	0.6923	0.8000	0.8621	0.9000	0.9245
	R	1.0000	1.0130	1.0275	1.0349	1.0378	1.0380

从表 1 中可以得出如下一些结论:

- 1) 扩展加权法是四个方法中最好的, 无论是 $E[|Y|]$ 还是 $\max |Y|$.
- 2) 最优加权平均法与强迫最大似然估计法差不多具有同样的精度.
- 3) 当两个待融合的数据具有相同的精度时, 四种方法的融合策略和融合精度都完全一样, 实际上就是直接取平均值.

从 (17) 式还可以推出一个有趣的性质: 在扩展加权平均法中, $|Y|$ 在灰色区域中的最大值与其在白色区域中的最大值相等, 两者均为 1.0000. 参见图 1.

实际上也可以把上述融合策略应用于两个待融合数据均为正态分布的情形. 以 $X_1 \sim N(0, 1)$, $X_2 \sim N(0, 4)$ 为例. 这里只比较 $E[|Y|]$, 因为 $|Y|$ 的最大值均为无限大. 此时最大似然估计法能很好地应用, 它的融合参数为 $k = 0.8$, 其 $E[|Y|] = 0.7136$. 本文的扩展加权平均法也可以很好的工作, 它比最大似然估计法效果更好, 尽管效果不是十分显著. 它的融合参数为 $k = 0.7975$, $C = 0.2850$, 其 $E[|Y|] = 0.7103$. 如果忽略随机误差, 这两种方法的融合精度几乎相等.

5 三个数据的融合

设有三个数据 X_1, X_2 和 X_3 将被融合, 它们均为随机变量. X_1 是 $[-1, 1]$ 上的均匀分布, X_2 是 $[-h_2, h_2]$ 上的均匀分布, $h_2 > 1$, X_3 是 $[-h_3, h_3]$ 的均匀分布, $h_3 > 1$. 所有融合均匀分布的三个数据的情形均可以简化为这样. X_1 必须是最精确的数据, 因为当 X_1 与其它数据差别较大时, 本文提出的扩展加权平均融合策略实际上让融合结果直接等于 X_1 .

融合三个数据的方法与前面融合两个数据的方法很相像. 描述如下

$$\begin{cases} Y_1 = X_1 & \text{if } |X_1 - X_2| < C_1 \\ Y_1 = k_1 X_1 + (1 - k_1) X_2 & \text{if } |X_1 - X_2| > C_1 \end{cases} \quad C_1 > 0, \text{ 且 } 0 < k_1 < 1$$

$$\begin{cases} Y = Y_1 & \text{if } |Y_1 - X_3| < C_2 \\ Y = k_2 Y_1 + (1 - k_2) X_3 & \text{if } |Y_1 - X_3| > C_2 \end{cases} \quad C_2 > 0, \text{ 且 } 0 < k_2 < 1 \quad (21)$$

因为推导过于复杂, 几乎不可能按照第 3 节的方法通过公式推导来得到此时的融合参数. 因此这里将通过蒙特卡洛法来确定 (21) 式中的各个融合参数.

本节使用 VC 来进行数值仿真. 此时一个重要的工作就是获得一个质量较高的伪随机数. 通常的软件包中的伪随机数是通过如下 (22) 式所示的同余序列产生的,

$$I_{n+1} = (aI_n + c) \bmod m \quad (22)$$

其中 I_0 为随机数种子, $m > I_0, a, c, a > 0, c > 0, a, c, m$ 均为整数.

这样产生的随机数的性质并不好, 当使用这样的随机数来产生随机向量时, 所产生的随机向量会聚集在一些超平面附近. 这会导致蒙特卡洛法的结果非常低劣. 例如, 如果在蒙特卡洛法中使用真正的随机数, $E[|Y|]$ 的计算结果应该没有什么变化. 而使用 (22) 式来产生随机数时, 只是改变随机数种子的值, 就导致 $E[|Y|]$ 的值改变了 10% 以上.

为了得到较好性质的伪随机数, 这里使用改进后的方法. 此时需要两个种子, 方法如下所述:

$$I_n = (a \times I_{n-1} + b \times I_{n-2}) \bmod m \quad (23)$$

当使用 (23) 式时, 改进的效果非常明显.

这里不详细介绍蒙特卡洛法. 通过计算得到了几种情形下的融合参数 C_i 和 k_i , 如表 2 所示. 其它情形的融合参数可以通过插值得到.

表 2 三个数据融合时的融合参数和融合精度

Table 2 Parameters and accuracies when fusing three pieces of data

h_2	h_2	C_1	k_1	C_2	k_2	$E[Y]$	$\text{Max} Y $
1.5	2.0	0.7575	0.6490	1.6985	0.7770	0.3483	1.0080
1.5	2.5	0.7646	0.6452	2.2290	0.8175	0.3574	1.0019
1.5	3.0	0.7607	0.6460	2.7407	0.8464	0.3635	1.0026
1.5	3.5	0.7662	0.6460	3.2514	0.8672	0.3675	1.0018
1.5	4.0	0.7680	0.6456	3.7680	0.8827	0.3706	1.0018
2.0	2.5	1.3591	0.7307	2.1642	0.8068	0.3809	1.0015
2.0	3.0	1.3732	0.7292	2.7055	0.8345	0.3878	1.0000
2.0	3.5	1.3657	0.7297	3.1980	0.8598	0.3927	1.0000
2.0	4.0	1.3662	0.7285	3.7085	0.8760	0.3963	1.0000
2.5	3.0	1.9151	0.7807	2.6358	0.8322	0.4016	1.0011
2.5	3.5	1.9164	0.7816	3.1652	0.8542	0.4070	1.0000
2.5	4.0	1.9202	0.7810	3.6652	0.8724	0.4110	1.0008

如果改变 X_2 和 X_3 的融合顺序, 即先把 X_3 和 X_1 融合, 再把所得结果与 X_2 融合. 此时为了仍然达到最佳融合效果, 需要改变各个融合参数, 对于具体的每一组融合数据, 最后的融合结果 Y 当然会改变, 但如果忽略蒙特卡洛法中的随机误差, 可以认为 $E[|Y|]$ 和 $|Y|_{\max}$ 并没有改变. 表 3 给出了两个例子.

表 3 融合顺序的影响

Table 3 Influence of fusion sequence

h_1	h_2	C_1	k_1	C_2	k_2	$E[Y]$	$\text{Max} Y $
4.0	1.5	3.4564	0.8702	0.8859	0.6710	0.3636	1.0000
1.5	4.0	0.7680	0.6456	3.7680	0.8827	0.3706	1.0018
2.0	3.0	1.3732	0.7292	2.7055	0.8345	0.3878	1.0000
3.0	2.0	2.4442	0.8197	1.5405	0.7505	0.3893	1.0000

可以从表 2 得出结论, 如果忽略蒙特卡洛法中的随机误差, $|Y|$ 的最大值总是 1.0. 与其它方法相比较, 这是本文的扩展加权平均法较为突出的优点之一.

在融合三个数据时, 这里再把本文的方法与通常的最优加权平均法做比较. 最优加权平均法可描述为:

$$Y = k_1 X_1 + k_2 X_2 + (1 - k_1 - k_2) X_3 \quad (24)$$

$$0 < k_1 < 1, 0 < k_2 < 1, 0 < (1 - k_1 - k_2) < 1.$$

实际上, 大多数加权平均法都可以用 (24) 式来表示. 按照本文在第 2 节提出的评判标准, 可以通过推导得到最优的 k_1 和 k_2 . 由于推导过程和结果过于复杂, 这里使用蒙特卡洛法通过最小化 $E[|Y|]$ 来确定最优的 k_1 和 k_2 .

当 $h_1 = 1.0, h_2 = 2.0, h_3 = 3.0$ 时, 比较结果如表 4 所示. 显然, 扩展加权平均法要好于最优加权平均法.

表 4 扩展加权平均法与最优加权平均法的融合精度比较

Table 4 Comparison of fusion accuracy between extended average and optimal average

h_2	C_1	k_1	C_2	k_2	$E[Y]$	$\text{Max} Y $
扩展加权平均法	1.3732	0.7292	2.7055	0.8345	0.3878	1.0000
最优加权平均法	—	0.6941	—	0.2120	0.4078	1.3828

6 结语

本文提出了一种扩展加权平均法来融合多个精度不同的原始数据, 这些数据可以是均匀分布, 也可以是正态分布. 融合结果与其它方法进行了对比. 按照本文第 2 节中提出的评判标准, 扩展加权平均法要明显好于其它方法. 将来的工作要着眼于把该方法推广到更多的情形, 同时把该方法应用于实际工程中.

References

- 1 Zied Elouedi, Khaled Mellouli, Philippe Smets. Assessing sensor reliability for multi-sensor data fusion within the transferable belief model. *IEEE Transactions on Systems, Man, and Cybernetics – Part B*, 2004, **34**(1): 782~787
- 2 Su Jian-Bo, Wang Jun, Xi Yu-Geng. Incremental learning with balanced update on receptive fields for multi-sensor data fusion. *IEEE Transactions on Systems, Man, and Cybernetics – Part B*, 2004, **34**(1): 659~665
- 3 Solaiman B, Debon R, Pipelier F, Cauvin J M, Roux C. Information fusion, application to data and model fusion for ultrasound image segmentation. *IEEE Transactions on Biomedical Engineering*, 1999, **46**(10): 1171~1175
- 4 Ben-Yacoub S, Abdeljaoued Y, Mayoraz E. Fusion of face and speech data for person identity verification. *IEEE Transactions on Neural Networks*, 1999, **10**(5): 1065~1074
- 5 Quartulli M, Datcu M. Information fusion for scene understanding from interferometric SAR data in urban environments. *IEEE Transactions on Geoscience and Remote Sensing*, 2003, **41**(9): 1976~1985
- 6 Solberg A H S. Contextual data fusion applied to forest map revision. *IEEE Transactions on Geoscience and Remote Sensing*, 1999, **37**(3): 1234~1243
- 7 Solaiman B, Pierce L E, Ulaby F T. Multisensor data fusion using fuzzy concepts: application to land-cover classification using ERS-1/JERS-1 SAR composites. *IEEE Transactions on Geoscience and Remote Sensing*, 1999, **37**(3): 1316~1326
- 8 McKeown D M Jr, Cochran S D, Ford S J, McGlone J C, Shufelt J A, Yocum D A. Fusion of HYDICE hyperspectral data with panchromatic imagery for cartographic feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 1999, **37**(3): 1261~1277
- 9 Azuaje F, Dubitzky W, Black N, Adamson K. Improving clinical decision support through case-based data fusion. *IEEE Transactions on Biomedical Engineering*, 1999, **46**(10): 1181~1185
- 10 Luo R C, Kay M G. Multisensor integration and fusion in intelligent systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 1989, **19**(5): 901~931
- 11 Hackett J K, Shah M. Multi-sensor fusion: a perspective. *IEEE International Conference Proceedings on Robotics and Automation*, 1990, **2**: 1324~1330
- 12 Hall D L, Llinas J. Handbook of Multisensor Data Fusion. Boca Raton, Florida, USA: CRC Press, 2001
- 13 Baek W, Bommareddy S. Optimal m-ary data fusion with distributed sensors. *IEEE Transactions on Aerospace and Electronic Systems*, 1995, **31**(3): 1150~1152
- 14 Yifeng Z, Leung H, Yip P C. An exact maximum likelihood registration algorithm for data fusion. *IEEE Transactions on Signal Processing*, 1997, **45**(1): 1560~1573
- 15 Bloch. Information combination operators for data fusion: a comparative review with classification. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 1996, **26**(1): 52~67

唐 进 博士, 副教授, 主要研究方向为图象处理、机器人、嵌入式系统和力学。

(TANG Jin Ph.D., associate professor. His research interests include image processing, robot, embedded system, and mechanics.)

张 闯 捷 硕士生, 主要研究方向为机器人。

(ZHANG Wen-Jie Graduate student. His research interest includes robot.)

高 琰 博士生, 主要研究方向为智能信息处理。

(GAO Yan Ph.D. candidate. Her research interests include intelligent information processing.)

付 明 玉 硕士, 教授, 主要研究方向为系统仿真和自动控制。

(FU Ming-Yu Professor. Her research interests include system simulation and automatic control.)

蔡 自 兴 教授, 主要研究方向为机器人、人工智能和智能控制。

(CAI Zi-Xing Professor. His research interests include robot, artificial intelligence, and intelligent control.)