

基于证据距离和矛盾因子的加权证据合成法

刘准钊, 程咏梅, 潘 泉, 苗 壮

(西北工业大学 自动化学院, 陕西 西安 710072)

摘要: 针对Dempster-Shafer (DS)理论直接对高冲突证据合成时会有融合结果不合理的问题, 本文提出了一种加权证据合成法. 在多个证据融合时, 该方法综合利用证据距离和矛盾因子表示证据冲突, 并根据各证据与其它证据的冲突程度确定其权重系数, 基于此又分别采用两种适用于不同情况的方法对证据基本置信指派函数修正, 然后用DS规则合成. 算例实验表明了本文方法对高冲突证据融合的有效性, 并且其融合结果还具有好的收敛性.

关键词: 证据理论; 权重系数; 证据距离; 矛盾因子

中图分类号: TP273 **文献标识码:** A

Combination of weighted belief functions based on evidence distance and conflicting belief

LIU Zhun-ga, CHENG Yong-mei, PAN Quan, MIAO Zhuang

(College of Automation, Northwestern Polytechnical University, Xi'an Shaanxi 710072, China)

Abstract: Dempster-Shafer(DS) theory involves counter-intuitive behaviors when evidence highly conflicts. A new approach of combination of weighted belief functions is proposed to solve the problem. If many pieces of evidence are to be combined, the amount of conflict between evidences is at first evaluated by both evidence distance and conflicting belief, and every piece of evidence is given a weight coefficient according to its amount of conflict with the others. Two different methods are separately used to modify the belief function of each piece of evidence based on its weight coefficient. Finally, the modified functions are combined by DS rule. In the last part of this paper, numerical examples are given to illustrate the effectiveness and the good performance of convergence of the proposed approach in case of high conflict between evidences.

Key words: evidence theory; weight coefficient; evidence distance; conflicting belief

1 引言(Introduction)

证据理论^[1]又称Dempster-Shafer(DS)理论可以有效的处理不确定信息, 但是直接利用DS规则对高冲突证据合成时, 很可能得出违背常理的结论. 目前, 解决该问题主要使用的方法包括改变矛盾信息分配方式的改进组合规则法^[2~5]和证据信息修正法^[6~8]. 不过改进后的组合规则很难满足结合律. 而在多个证据一起合成时, 计算量又太大, 不便于实际应用. 为此, 本文提出了一种基于信息修正的加权证据合成法.

本文综合考虑了矛盾因子和证据距离与证据冲突的关系, 提出了由两者共同确定冲突程度的方法. 根据每个证据与其他证据冲突程度的大小来确定其权重系数, 之后分别采用期望证据法和相对加权证

据法对证据修正, 然后用DS规则合成. 通过数字实验将本文方法与其他方法进行了对比分析.

2 DS规则及其改进(DS combination rule and its improvements)

DS理论^[1]是建立在辨识框架 Θ 上的理论, 对于问题域中任何命题 A 都应包含于幂集 2^Θ 中. 在 2^Θ 上定义了基本置信指派函数 $m: 2^\Theta \rightarrow [0, 1]$, m 要满足: 1) $m(\phi) = 0$; 2) $0 \leq m(A) \leq 1$; 3) $\sum_{A \subset 2^\Theta} m(A) = 1$. $\forall A \subset 2^\Theta$, 如果 $m(A) > 0$, 则 A 称为焦元.

假定辨识框架 Θ 上独立的两个证据, 其焦元分别为 B_i 和 C_j ($i = 1, 2, \dots, n; j = 1, 2, \dots, m$), m, n 分别表示这两个证据焦元的个数, 其基本置信指派函数为 m_1 和 m_2 , 按DS规则合成如下:

收稿日期: 2008-10-14; 收修改稿日期: 2009-03-20.

基金项目: 国家自然科学基金重点资助项目(60634030); 航空科学基金资助项目(2007ZC53037); 高等学校博士学科专项科研基金资助项目(20060699032).

$$m(A) = \frac{\sum_{B_i \cap C_j = A} m_1(B_i)m_2(C_j)}{1 - k}, A \neq \Phi; \quad (1)$$

$$m(\Phi) = 0.$$

矛盾因子

$$k = \sum_{B_i \cap C_j = \Phi} m_1(B_i)m_2(C_j). \quad (2)$$

DS规则为了基本置信指派函数的归一化而将矛盾信息按比例分配给了所有焦元,从而导致了直接对高冲突证据合成时会出现结果不合理的问题.为此,许多学者纷纷提出了相应的改进规则^[2~5],其中Dezert和Smarandache提出的PCR5规则^[5]有典型的代表性,在对两个证据合成时其规则如下:

$$\forall x, y \in 2^\Theta,$$

$$m(x) = \sum_{\substack{B_i, C_j \in 2^\Theta \\ B_i \cap C_j = x}} m_1(B_i)m_2(C_j) + \sum_{\substack{y \in 2^\Theta \\ x \cap y = \phi}} \left[\frac{m_1^2(x)m_2(y)}{m_1(x) + m_2(y)} + \frac{m_2^2(x)m_1(y)}{m_2(x) + m_1(y)} \right]. \quad (3)$$

PCR5多个证据一起合成的公式见文^[5].

由式(3)可以看出 $m(x)$ 有两部分相加构成,该规则显然不满足结合律,多个证据合成顺序势必影响最终结果,若按一起合成的公式进行组合,则计算量太大,其他改进规则具有类似特点.为此,本文认为要在利用DS规则合成之前通过对证据信息的修正来解决高冲突证据合成问题.

3 证据冲突程度的确定(Measurement of conflict between belief functions)

证据修正法的思想是与其它证据的冲突程度较高的证据权重较小,冲突程度低的权重较大,以此消减与其他多数证据冲突较大的证据对合成结果的影响.因此,确定证据冲突程度的大小成了关键.

文^[8]中提出了将两个证据的矛盾因子经过变换处理后来确定证据权重的方法.文^[10]提出的证据距离也可以反映证据冲突程度,其定义如下:在相同的辨识框架下,两个证据的基本置信指派函数 m_1 和 m_2 的距离为:

$$d(m_1, m_2) = \sqrt{\frac{1}{2}(\vec{m}_1 - \vec{m}_2)^T \underline{D}(\vec{m}_1 - \vec{m}_2)}, \quad (4)$$

$$\underline{D}[i, j] = |A_i \cap B_j| / |A_i \cup B_j|.$$

其中 $\underline{D}$ 为 $2^{|\Theta|} \times 2^{|\Theta|}$ 正定系数矩阵, A_i, B_j 分别 m_1 和 m_2 幂集 2^Θ 中的元素.

矛盾因子和证据距离分别单独使用时却不能全

面地描述证据冲突程度.

例1 设 $\Theta = \{a, b, c\}$,

$$E1: m_1(a) = 0.6, m_1(b) = 0.2, m_1(c) = 0.2;$$

$$E2: m_2(a) = 0.6, m_2(b) = 0.2, m_2(c) = 0.2.$$

这两个证据是一样的,直觉判断其冲突应为零.计算得 $k = 0.56, d = 0$,根据距离判断其没有冲突,与分析一致;根据矛盾因子判断其有显著冲突,因此单纯利用矛盾因子来表征冲突程度是不合适的.

例2 设 $\Theta = \{a, b, c\}$

$$E1: m_1(a) = 0.5, m_1(b) = 0.3, m_1(c) = 0.2;$$

$$E2: m_2(a) = 0.2, m_2(b) = 0.3, m_2(c) = 0.5;$$

$$E3: m_3(a) = 0, m_3(b) = 0.1, m_3(c) = 0.9.$$

证据 $E1$ 支持的是 a , $E2$ 和 $E3$ 支持 c ,一般推断 $E2$ 和 $E1$ 的冲突要大于 $E2$ 和 $E3$,经计算得 $k_{12} = 0.71$ (k_{ij} 表示第 i 个和第 j 个证据的矛盾因子), $d_{12} = 0.3$ (d_{ij} 表示第 i 个和第 j 个证据的距离), $k_{23} = 0.52$, $d_{23} = 0.35$,故根据矛盾因子判断 $E2$ 和 $E1$ 的冲突大于 $E2$ 和 $E3$,与分析一致,而按距离判断却相反,故单纯利用证据距离表征冲突程度也不合适.

矛盾因子表示了证据合成时不相容焦元(交集为空集)结合产生的矛盾信息大小,证据距离表示了证据间相容焦元基本置信值分配的差异,这两者具有一定的互补性.因此,本文认为证据冲突程度应由矛盾因子和证据距离来联合反映,共同确定.

4 加权证据合成法(Combination of weighted belief functions)

4.1 证据权重系数(Weight coefficient of belief functions)

证据权重系数取决于该证据与其他证据的冲突程度.所以,在对 n 个证据赋权重时,首先要求得第 i 个与第 j ($j \neq i$)个证据的矛盾因子 k_{ij} 和距离 d_{ij} ,由于一般反映冲突程度的量都位于0到1之间,在应用中还要简便易求,故本文用证据距离和矛盾因子的几何均值来表示冲突的大小,即

$$\text{conf}(i, j) = \sqrt{k_{ij}d_{ij}}. \quad (5)$$

获得各证据间冲突程度后,可构造一个 $n \times n$ 的证据冲突程度方阵:

$$\text{CONF} = \begin{bmatrix} 0 & \text{conf}_{1,2} & \cdots & \text{conf}_{1,n} \\ \text{conf}_{2,1} & 0 & \cdots & \text{conf}_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \text{conf}_{n,1} & \text{conf}_{n,2} & \cdots & 0 \end{bmatrix}_{n \times n}, \quad (6)$$

则第 i 个证据与其他证据的总冲突程度为冲突方阵的第 i 行或第 i 列之和:

$$\text{conf}_i = \sum_{j=1}^n \text{conf}(i, j). \quad (7)$$

那么第*i*个证据获得其他*n* - 1个证据的总支持度, 也可称为第*i*个证据的众信度为

$$\text{tru}_i = (n - 1) - \text{conf}_i. \quad (8)$$

显然, 与其他证据的冲突程度越高, 获得众信度越小. 然后, 可以得到*n*个证据的众信度向量

$$\text{TRU} = [\text{tru}_1 \cdots \text{tru}_i \cdots \text{tru}_n]. \quad (9)$$

将众信度向量归一化后可得各证据的权重系数

$$w_i = \text{tru}_i / \sum_{j=1}^n \text{tru}_j. \quad (10)$$

故得到证据的权重系数向量为

$$W = [w_1 \cdots w_i \cdots w_n]. \quad (11)$$

4.2 证据信息的修正(Modification of belief functions)

1) 期望证据法.

根据权重系数向量式(11)可以求得期望证据

$$M = \sum_{i=1}^n w_i \vec{m}_i. \quad (12)$$

DS规则对*M*迭代组合*n* - 1次后的结果作为这*n*个证据的合成结果.

2) 相对加权证据法.

首先要确定权重系数最大的证据为关键证据, 其权重系数为

$$w_{\max} = \max(w_1, \cdots, w_i, \cdots, w_n). \quad (13)$$

然后得到各证据相对权重向量

$$W^* = [w_1 \cdots w_n] / w_{\max}. \quad (14)$$

由此可确定证据基本置信指派函数的“折扣因子”

$$\alpha_i = \frac{w_i}{w_{\max}}. \quad (15)$$

根据“折扣因子”对证据进行修正,

$$\begin{aligned} m_i^*(A_k) &= \alpha_i m_i(A_k), A_k \neq \Theta; \\ m_i^*(\Theta) &= 1 - \sum m_i^*(A_k). \end{aligned} \quad (16)$$

修正后的证据利用DS规则逐个合成.

5 算例实验与讨论(Numerical examples and discussion)

实验 1 基于多传感器的空中目标识别是一个重要的研究方向, 现设有5个不同性质的传感器对空中目标进行观测识别.

识别算法如下: 设辨识框架 $\Theta = \{a: \text{民航机}, b: \text{轰炸机}, c: \text{战斗机}\}$, 则所用的模版库中存有以上5种传感器预先测得的这3类目标的特征数据. 当传感器对目标观测时, 首先分别提取这5种观测信息的特征,

然后计算提取特征与模版库中3类目标相应特征的欧氏距离, 接着构造映射求取观测目标与辨识框架目标的相似度, 映射要满足距离越大, 相似度越小. 这样可得到5种对观测目标在辨识框架下的隶属度分配. 对5种隶属度分配归一化处理后就得到了5个证据, 而后采用证据理论融合算法对证据融合, 根据融合结果做出判断识别.

为便于对比, 本例采用与文[7]中算例相同的证据基本置信指派函数.

$$E1: m_1(a) = 0.5, m_1(b) = 0.2, m_1(c) = 0.3;$$

$$E2: m_2(a) = 0, m_2(b) = 0.9, m_2(c) = 0.1;$$

$$E3: m_3(a) = 0.55, m_3(b) = 0.1, m_3(c) = 0.35;$$

$$E4: m_4(a) = 0.55, m_4(b) = 0.1, m_4(c) = 0.35;$$

$$E5: m_5(a) = 0.55, m_5(b) = 0.1, m_5(c) = 0.35.$$

不同方法的合成结果如表1所示.

表 1 合成结果对比

Table 1 Comparison between combination results

		$E1 \sim E3$	$E1 \sim E4$	$E1 \sim E5$
DS	$m(a)$	0	0	0
	$m(b)$	0.6316	0.3288	0.1288
	$m(c)$	0.3684	0.6712	0.8722
文[6]	$m(a)$	0.35	0.6027	0.7958
	$m(b)$	0.5224	0.2627	0.0932
	$m(c)$	0.1276	0.1346	0.1110
文[7]	$m(a)$	0.4861	0.7773	0.8909
	$m(b)$	0.3481	0.0628	0.0086
	$m(c)$	0.1657	0.1600	0.1005
期望证据法	$m(a)$	0.6013	0.8193	0.8954
	$m(b)$	0.2214	0.0354	0.0056
	$m(c)$	0.1774	0.1452	0.0990

从5个证据可以看出, 有4个证据一致认为真实目标是*a*的可能性最大, 据常理推断, 第2个证据可能是相应传感器因受干扰导致信息有误, 真实目标应该是*a*: 民航机.

2个证据合成时无法体现权重系数给出方法的优劣. 在3个证据合成时, 文[6]得到的平均证据中 $m(b) > m(a)$, 导致结果认为*b*的基本置信值 $m(b)$ 最大, 以此易做出不合理的判断, 文[7]虽然得出 $m(a)$ 最大, 但 $m(a)$ 的绝对值以及 $m(a)$ 与 $m(b)$ 差值都不够大, 不利于做出正确决策, 而期望证据法已可明确得出正确的决策识别. 随着第4、5个证据加入, 文[6,7]和期望证据法的合成结果都得出了 $m(a)$ 最大, 可以做出有效决策, 这明显优于DS规则, 但期望证据法下各目标的基本置信值向期望值收敛($m(a) \rightarrow 1, m(b) \rightarrow 0, m(c) \rightarrow 0$)要快于其它方法, 有利于及时做出正确识别.

实验 2 由于本例是希望对比相对加权证据法

与文[8]方法给出权重因子方法的优劣,故分别求得权重系数后皆用DS规则合成.PCR5按证据的标号顺序依次逐个合成.不同方法对实验1证据算例的合成结果如表2所示.

表2 合成结果对比

Table 2 Comparison between combination results

		$E1 \sim E3$	$E1 \sim E4$	$E1 \sim E5$
PCR5	$m(a)$	0.3700	0.5101	0.6154
	$m(b)$	0.4482	0.2579	0.1247
	$m(c)$	0.1818	0.2319	0.2598
文[8]	$m(a)$	0.1691	0.2970	0.3947
	$m(b)$	0.4913	0.2121	0.0683
	$m(c)$	0.3395	0.4909	0.5369
文[9]	s	[1, 0.089]	[1, 1]	[1, 1]
	$m(a)$	0	0	0
	$m(b)$	0.5000	0.3288	0.1288
相对加权 证据法	$m(c)$	0.5000	0.6712	0.8722
	$m(a)$	0.5794	0.7598	0.8383
	$m(b)$	0.1493	0.0224	0.0045
	$m(c)$	0.2713	0.2179	0.1572

PCR5规则直到前4个证据合成后方可得出 $m(a)$ 最大,但各目标基本置信值向期望值收敛太慢,不利于及时做出正确识别.文[8]方法的融合结果一直未得出 $m(a)$ 最大,无法做出正确识别.利用文[9]方法时,由于证据 $E2$ 与其他证据距离明显较大,故采用双层合成法,先选取参数 $s(1)$ (与表中 s 的第1维向量对应)对除 $E2$ 外的其他证据合成,再选取参数 $s(2)$ 利用合成结果与 $E2$ 合成,其结果中 $m(a)$ 一直为零,与DS规则相比基本没有改善,表明该方法不适合对高冲突证据合成.相对加权证据法从3个证据合成开始一直可得出 $m(a)$ 最大,并且随着证据的增加,各目标的基本置信值向期望值收敛较快,根据合成结果可以较容易的快速得出正确识别.

从实验结果来看,期望证据法和相对加权证据法都可以有效解决高冲突证据的合成问题.但在应用中对这两种方法的选择要注意的是:期望证据法是将期望证据按DS规则迭代合成,故在多数证据的基本置信指派函数较为一致时,通过确定合理的权重系数得到准确的期望证据,则合成结果中各目标基本置信值会很快向期望值收敛,有利于及时做出正确决策.但是,当多数证据中基本置信指派函数不一致,对不同目标支持度的大小区别不明显时,该方法经过对期望证据的多次迭代后,合成结果会做出较为绝对的判断,可能导致决策错误,此时应慎用.相对加权证据法通过对证据打折扣增大其不确定性来消减彼此的冲突,由于其增大了不确定性信息,目标基本置信值的收敛速度会稍慢一些,但受权重系数的影响没有期望证据法那样严重,是一种较保守的合成方法.故要根据实际情况来做出合适的选择.

6 结论(Conclusion)

本文将矛盾因子和证据距离综合起来更全面地表示了证据间冲突程度,从而更好的确定出了各证据的权重系数,并采用了两种方法对证据信息修正,然后用DS规则合成.实验结果表明,在按照本文方法给出各证据权重系数的基础上,相对加权证据法和期望证据法分别与其他方法相比,可以更加有效地解决高冲突证据的合成问题,使各目标的基本置信值更快地向期望值收敛,有利于快速及时地做出正确的识别决策.

参考文献(References):

- [1] SHAFER G A. *Mathematical Theory of Evidence*[M]. Princeton, New Jersey: Princeton University Press, 1976.
- [2] 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, 8(8): 117 – 119.
(SUN Quan, YE Xiuqing, GU Weikang. A new combination rules of evidence theory[J]. *Acta Electronica Sinica*, 2000, 8(8): 117 – 119.)
- [3] 张山鹰, 潘泉, 张洪才. 证据推理冲突问题研究[J]. 航空学报, 2001, 22(4): 369 – 372.
(ZHANG Shanying, PAN Quan, ZHANG Hongcai. Conflict problem of DS evidence theory[J]. *Acta Aeronautica et Astronautica Sinica*, 2001, 22(4): 369 – 372.)
- [4] YAGER R R. On the dempster shafer framework and new combination rules[J]. *Information System*, 1989, 41(2): 93 – 137.
- [5] SMARANDACHE F, DEZERT J. Proportional conflict redistribution rules for information fusion[M] // *Advances and Applications of DSMT for Information Fusion (Collected works)*, Florentin Smarandache and Jean Dezert, Editors, Rehoboth: American Research Press, 2006, 2: 3 – 68.
- [6] MURPHY C K. Combining belief functions when evidence conflicts[J]. *Decision Support Systems*, 2000, 29(1): 1 – 9.
- [7] DENG Y, SHI W K, ZHU Z F, et al. Combining belief functions based on distance of evidence[J]. *Decision Support Systems*, 2004, 38(3): 489 – 493.
- [8] 叶清, 吴晓平, 宋业新. 基于权重系数与冲突概率重新分配的证据合成方法[J]. 系统工程与电子技术, 2006, 28(7): 1014 – 1016.
(YE Qing, WU Xiaoping, SONG Yexin. Evidence combination method based on the weight coefficients and the conflict probability distribution[J]. *Systems Engineering and Electronics*, 2006, 28(7): 1014 – 1016.)
- [9] BENJAMIN Q, MASSON M H, DENOUEUX T. Refined classifier combination using belief functions[C] // *Proceedings of the 11th International Conference on Information Fusion (FUSION '08)*. Piscataway: IEEE Press, 2008, 7: 1 – 7.
- [10] JOUSSELME A L, GRENIER D, BOSSE E. A new distance between two bodies of evidence[J]. *Information Fusion*, 2001, 2(2): 91 – 101.

作者简介:

刘准钊 (1984—), 男, 博士研究生, 主要研究方向为信息融合、图像识别, E-mail: liuzhunga@gmail.com;

程咏梅 (1960—), 女, 教授, 主要研究方向为人工智能、信息融合;

潘泉 (1961—), 男, 教授, 主要研究方向为估计与控制、信息融合、图像处理;

苗壮 (1978—), 男, 博士生, 主要研究方向为信息融合、目标识别.