

文章编号:1004-9037(2014)01-0121-05

基于距离测度的证据合成方法

刘希亮 陈桂明 李方溪 张 倩

(第二炮兵工程大学装备管理工程系,西安,710025)

摘要:为解决经典证据理论无法处理高度冲突证据的不足,提出了一种基于距离测度的证据合成方法。将证据视为空间向量,定义距离测度空间,计算空间中两两证据之间的距离,并通过距离测度矩阵确立证据之间的一致性测度,在此基础上求得辨识框架对各证据的支持度,通过归一化支持度得到证据一致性因子,以此作为冲突概率的分配权重,最后给出改进的证据合成公式。数值算例证明了改进公式既能处理冲突证据,又能合成非冲突证据,与其他合成方法的比较验证了改进合成公式的有效性。

关键词:证据理论;冲突;距离测度;一致性因子

中图分类号:TP391

文献标志码:A

Approach to Evidence Combination Based on Distance Measurement

Liu Xiliang, Chen Guiming, Li Fangxi, Zhang Qian

(Equipment Management Engineering Department, Second Artillery Engineering University, Xi'an, 710025, China)

Abstract: To deal with the disadvantage that classical evidence theory cannot combine highly conflicting evidence, a new combination approach is presented based on distance measurement. The evidence is regarded as space vector and distance measurement space is defined, in which the distance between evidence is calculated. Then the consistency measurement is established using distance measurement matrix by which the support degree is gained on the frame of discernment. Finally, the evidence consistency coefficient obtained from normalizing support degree is regarded as the weight to distribute conflicting probability, and improved evidence combination formula is proposed. Numerical examples prove that the improved formula can combine the conflicting evidence and the non-conflicting evidence. The contrast to other methods testifies the validity of improved formula.

Key words: evidence theory; conflict; distance measurement; consistency coefficient

引 言

作为贝叶斯理论的推广,Dempster-Shafer(D-S)证据理论具有坚实的数学基础,能在不需要先验知识的情况下,以简单的推理形式,得出较好的融合结果^[1,2]。D-S理论作为一种重要的不确定性推理方法,为不确定性信息的表达和合成提供了强有力的工具,在很多领域取得了广泛应用。

然而,在处理不一致证据时,D-S方法会因合成规则中的归一化过程产生与直觉相悖的结论,即遇到了“在模型 x 上使用方法 y 获得了一个与直

觉相悖的结论 z ”的问题。有人认为是模型 x (证据本身,信息源)不可靠,也有人认为是方法 y (组合规则)出了问题。鉴于此,人们从两个方面对证据理论进行了改进:一方面,基于修正原始证据源的方法。该类方法认为,D-S合成规则本身没有错,当证据高度冲突时,应该首先对冲突证据进行预处理,然后再使用D-S合成规则。Murphy算术平均法^[3]先对所有证据做算术平均处理,再利用D-S公式进行合成;邓勇加权平均法^[4]认为不同的证据对最终决策的影响不一样,把算术平均处理改为加权平均,权重根据证据之间的距离来确定。Murphy算术平均法和邓勇加权平均法对单个证据做

多次合成,违背了证据理论要求各证据相互独立的假设,如果均值证据本身具有不确定性,在没有增加其他证据的情况下人为地使它达到确定性,那么这种确定性是不可靠的;另一方面,基于修改 D-S 合成规则的方法。该类方法认为,冲突合成规则是改进和发展证据推理的关键问题。Smets^[5]认为引起冲突的原因是辨识框架不完备,因此保留冲突概率 k 不做归一化处理;Yager^[6]把冲突概率全部赋给了未知领域。在此基础上,Lefevre^[7]给出了冲突证据合成的一般性框架,文献[8]把冲突概率按各个命题的平均支持程度加权进行分配,缺点是当各命题的平均支持程度相关不大时,可能出现不能收敛到确定性的情况。

本文在前人研究的基础上,遵循第二种改进思路,提出一种基于距离测度的证据合成方法,对证据理论的合成规则进行了修正。

1 经典 D-S 证据理论

在证据理论中,辨识框架 $\Theta = \{H_1, H_2, \dots, H_N\}$ 表示一个由 N 个两两相斥的命题构成的有限集合,用 $P(\Theta) = \{\Phi, \{H_1\}, \{H_2\}, \{H_3\}, \dots, \{H_N\}, \{H_1 \cup H_2\}, \{H_1 \cup H_3\}, \dots, \Theta\}$ 表示 Θ 中 2^N 个元素的幂集。

定义 1 幂集 $P(\Theta)$ 上的集函数 $m: P(\Theta) \rightarrow [0, 1]$ 满足

$$m(\Phi) = 0 \quad (1)$$

$$\sum_{A \subseteq \Theta} m(A) = 1 \quad (2)$$

则称 m 为框架 Θ 上的基本概率赋值函数(Basic probability assignment, BPA), $m(A)$ 反映了证据对辨识框架中命题 A 的支持程度。

定义 2 幂集 $P(\Theta)$ 上所有证据的冲突概率为

$$k = \sum_{\cap A_j = \Phi} \prod_{1 \leq j \leq n} m_i(A_j) \quad (3)$$

式中 $m_i(A_j)$ 表示第 i 条证据第 j 个命题的 BPA。

系数 k 反映了证据之间冲突(不一致)的程度, k 值越大,证据之间的冲突越大,一致性越小。

定义 3 幂集 $P(\Theta)$ 上所有证据的合成规则为

$$m(A) = \frac{1}{1-k} \sum_{\cap A_j = A} \prod_{1 \leq j \leq n} m_i(A_j) \quad (4)$$

式中归一化因子 $1/(1-k)$ 的作用就是避免在合成时将非 0 的概率赋给空集 Φ ^[9,10]。

2 D-S 合成存在的问题

由定义 2 和 3 可知,如果冲突概率 $k=1$,证据

之间矛盾,不能使用合成规则进行融合;而当 $k \rightarrow 1$ 时,证据高度冲突,式(4)会产生有悖常理的结果。

例 1 3 个医生针对同一病人进行诊断,认为病症可能是脑膜炎(M),脑震荡(C)和脑肿瘤(T)中的一种,即辨识框架 $P(\Theta) = \{M, C, T\}$ 。由此获得 3 条诊断证据:

$$\text{证据 1: } m_1(M) = 0.95, m_1(C) = 0.05;$$

$$\text{证据 2: } m_2(C) = 0.05, m_2(T) = 0.95;$$

$$\text{证据 3: } m_3(M) = 0.80, m_3(C) = 0.10, m_3(T) = 0.10.$$

不难发现,证据 1 和 2, 2 和 3 都是高度冲突的,根据式(3)可得 $k=0.99975$,根据式(4)可得 $m(C)=1, m(M)=m(T)=0$ 。由此得出的结论是病人患脑震荡。这显然是有悖常理的,因为 3 个医生都认为病人患脑震荡的可能性很低,合成结果却恰好相反。

由上例可以看出 D-S 证据理论存在的 3 个问题:(1)D-S 合成规则无法合成高度冲突的证据,得出合理的结果;(2)在多证据合成中,由于证据冲突会使合成结果放弃占主导地位的多数意见,而把 100% 的确定性赋予少数意见;(3)只要有一个证据彻底不支持某一命题,那么合成的结果对该命题的支持永远为 0。

3 证据合成公式的改进

为了定量描述证据间的冲突程度,可以建立距离测度空间,将每组证据看作空间中的向量,通过证据向量之间的距离度量各个证据的一致性程度。引入证据一致性因子作为分配冲突概率的依据,若某个证据与其他证据的距离越小,则认为它们之间一致性较好,在分配冲突概率时通过一致性因子赋予该证据较大的权重,得到较多的冲突概率,最终获得较大的支持程度,反之亦然。由此提出了一种改进的证据合成公式。

设辨识框架 Θ 内有 N 个命题,获得 n 条相互独立的证据,以每个证据的 BPA 为坐标,建立一个 \mathbf{R}^{2^N} 维空间(2^N 表示幂集 $P(\Theta)$ 中元素的个数),则可以得到在此空间中的一组证据向量, $\mathbf{p}_i = (m_i(\Phi), m_i(A_1), m_i(A_2), \dots, m_i(A_n), m_i(A_1 \cup A_2), m_i(A_1 \cup A_3), \dots, m_i(\Theta))^T$ ($1 \leq i \leq n$),所有证据组成一个含有 n 个证据向量的集合^[11]。

定义 4 集合 P 上的二元实值函数 $D(\mathbf{x}, \mathbf{y})$ 为 \mathbf{R}^{2^N} 空间中任意两个向量 \mathbf{x} 和 \mathbf{y} 的距离,若 $F(\mathbf{x}, \mathbf{y})$ 满足:(1) $F(\mathbf{x}, \mathbf{y}) = F(\mathbf{y}, \mathbf{x})$; (2) $0 \leq F(\mathbf{x}, \mathbf{y}) \leq 1$; (3) 当且仅当 $\mathbf{x} = \mathbf{y}$, $F(\mathbf{x}, \mathbf{y}) = 0$; 当且仅当 $\mathbf{x} \perp \mathbf{y}$,

$F(x, y) = 1$, 则称 P 为距离测度空间。

定义 5 距离测度空间 P 中任意两个证据向量 p_i 和 p_j 的距离测度为

$$d_{ij} = \| p_i - p_j \| = \sqrt{\left(\sum_{k=1}^{2^N} [m_i(A_k) - m_j(A_k)]^2 \right) / 2} \quad (5)$$

由此可以得到一个距离测度矩阵

$$D = \begin{bmatrix} 0 & d_{12} & d_{13} & \cdots & d_{1n} \\ d_{21} & 0 & d_{23} & \cdots & d_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & d_{n3} & \cdots & 0 \end{bmatrix} \quad (6)$$

定义 6 证据 m_i 和 m_j 之间的一致性测度 Con_{ij} 为

$$Con_{ij} = e^{-(d_{ij} - d_{\min})} \quad i \neq j \quad (7)$$

式中

$$d_{\min} = \min(d_{ij}) \quad (8)$$

表示最小证据距离。

当两个证据之间的距离越小时,它们的一致性测度就越大,表明相互支持的程度越高。

定义 7 辨识框架对证据 m_i 的支持度 Sup_i 为

$$Sup_i = \sum_{j=1, i \neq j}^n Con_{ij} \quad (9)$$

式(9)是将一致性测度中除自身以外的所有元素求和, Sup_i 反映的是证据 m_i 与其他所有证据的支持程度,是一致性测度的函数。如果一个证据与其他证据一致性较“好”,则认为它们相互支持。如果一个证据与其他证据一致性较“差”,则认为它们相互支持的程度也低。

定义 8 证据 m_i 与其他证据的一致性因子 δ_i 为

$$\delta_i = \frac{Sup_i}{\sum_{i=1}^n Sup_i} \quad (10)$$

将一致性因子 δ_i 作为分配冲突概率的权重,可以得到以下的合成公式

$$m(A) = \sum_{\cap A_j = A} \prod_{1 \leq j \leq n} m_i(A_j) + k \cdot \sum_{i=1, \cap A_j = A}^n \delta_i \cdot m_i(A_j) \quad (11)$$

改进的合成公式实际上是一个加权和形式,当证据之间冲突较小时,即 k 值较小, $m(A)$ 合成结果主要由公式第一项 $\sum_{\cap A_j = A} \prod_{1 \leq j \leq n} m_i(A_j)$ 决定,近似于传统 D-S 合成结果;当证据之间高度冲突时,即 k 值较大, $m(A)$ 合成结果主要由公式第二项 $k \cdot$

$\sum_{i=1}^n \delta_i \cdot m_i(A)$ 决定,冲突概率 k 的分配由 $\sum_{i=1}^n \delta_i \cdot m_i(A)$ 决定,而每条证据所占的权重又由一致性因子 δ_i 决定,因此 δ_i 决定了高度冲突证据的合成结果。由上述分析可知,改进的合成公式能够有效平衡多个证据,也不会把 100% 的确定性赋予少数意见,元素较多的命题也能得到一定的信任分配,既能合成一致性证据,又能合成冲突证据。此外,只要有证据支持某一命题,该命题的合成结果就不会为 0,从而有效克服了 D-S 合成公式 3 个方面的不足。

利用改进合成公式进行证据合成的一般步骤如图 1 所示。

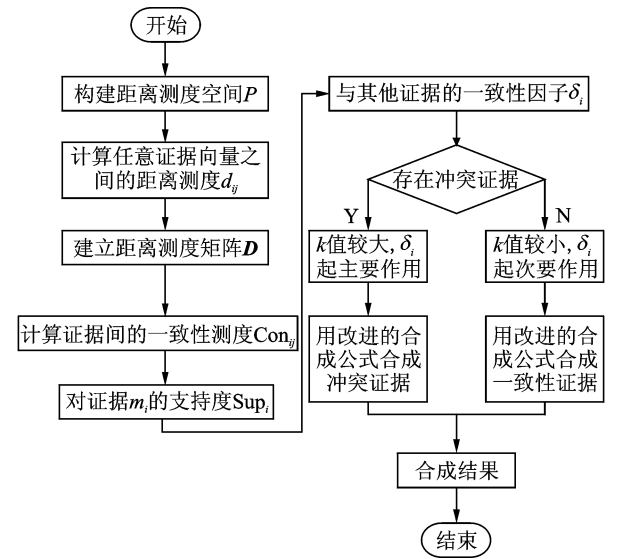


图 1 证据合成流程图

Fig. 1 Flow chart of evidence combination

4 应用实例

4.1 在例 1 中的应用

由例 1 可知,经典证据理论无法合成高度冲突证据,应用本文方法,根据图 1 所示的合成步骤,可得距离测度矩阵

$$D = \begin{bmatrix} 0 & 0.950 0 & 0.132 3 \\ 0.950 0 & 0 & 0.826 1 \\ 0.132 3 & 0.826 1 & 0 \end{bmatrix}$$

因此 $d_{\min} = 0.132 3$, 对证据 m_i 的支持度分别为 $Sup_1 = 1.441 4$, $Sup_2 = 0.941 1$, $Sup_3 = 1.499 6$, 证据 m_i 与其他证据的一致性因子 $\delta_1 = 0.371 3$, $\delta_2 = 0.242 4$, $\delta_3 = 0.386 3$ 。代入式(11)得 $m(M) = 0.651 6$, $m(C) = 0.089 5$, $m(T) = 0.258 9$ 。

由以上结果可知病人患脑膜炎,诊断结果符合

常理(原始证据中有两位医生认为是患脑膜炎),这也证明了本文方法在处理高度冲突证据时的有效性。但由于证据较少,需要更多证据才能得出更为准确可信的结论。

4.2 在冲突证据合成中的应用

例 2 假设辨识框架 $\Theta = \{A, B, C, D\}$, 依次收到如表 1 所示的 8 条证据。

表 1 原始证据值

Table 1 Original evidence value

证据	$m(A)$	$m(B)$	$m(C)$	$m(D)$
1	0.10	0.40	0.30	0.20
2	0.15	0.50	0.15	0.20
3	0.12	0.46	0.21	0.21
4	0.16	0.56	0.12	0.16
5	0.60	0.00	0.20	0.20
6	0.25	0.25	0.25	0.25
7	0.10	0.40	0.30	0.20
8	0.18	0.53	0.23	0.06

为验证改进合成公式的有效性,分别利用 D-S 合成公式、Murphy 合成公式、邓勇合成公式、李弼程合成公式和本文公式进行证据合成,结果如表 2 所示。

表 2 证据合成结果

Table 2 Evidence combination results

合成方法	$m(A)$	$m(B)$	$m(C)$	$m(D)$
D-S 合成法	0.141 5	0.000 0	0.711 8	0.146 7
Murphy 合成法	0.006 0	0.781 8	0.210 5	0.001 7
邓勇合成法	0.002 2	0.889 2	0.107 4	0.001 2
李弼程合成法	0.207 5	0.387 5	0.220 0	0.185 0
本文方法	0.012 4	0.966 5	0.019 2	0.001 9

以命题 B 为例,使用不同合成方法得到的证据值如图 2 所示。

由表 1 可知,证据 5 和 6 与其他证据冲突较大,从而导致各种方法合成结果存在较大差异。从表 2 发现,D-S 合成结果受证据 5 和 6 影响严重,由于证据 5 彻底否定目标 B,导致以后的合成结果 $m(B)$ 始终为 0;其余合成方法均可抑制冲突证据 5 和 6 的干扰,合成结果符合常理,Murphy 和邓勇合成都得到了较高的精度,正如前文所述,这两种方法对修正后的证据做多次合成,违背了 D-S 证据理论要求证据之间相互独立的假设,当证据本身具有不确定性时,在没有增加其他证据的情况下使它达到确定性,那么这种合成是不可靠的;李弼程合成得到的精度较低,不利于决策。从图 1 中

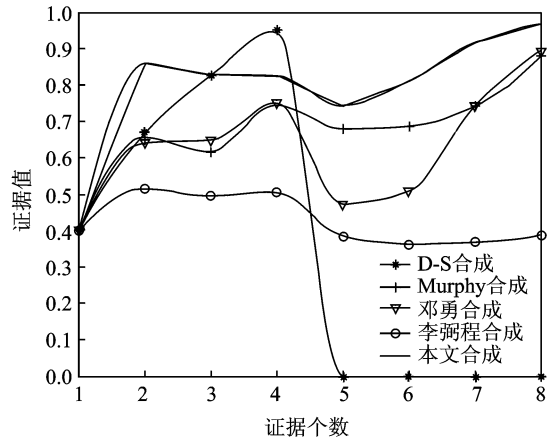


图 2 目标 B 合成结果

Fig. 2 Combination results of B

$m(B)$ 的合成结果可以看出,当出现证据 5 和 6 时,D-S 合成突然变为 0,得到了与事实不符的结果;Murphy 和邓勇合成最终都得到了正确的合成结果,但受证据 5 和 6 影响较大,在图中出现较大的波动;李弼程合成虽然受冲突证据影响较小,但合成精度较低,导致无法判断和决策。而本文方法较好地克服了以上几种方法的不足,受冲突证据影响小,合成精度高,相对于其他方法合成后的证据值变化小,而且随着证据数量的增加, $m(B)$ 值明显增加,有助于做出决策。这也证明了本文方法在合成冲突证据时的有效性。

4.3 在非冲突证据合成中的应用

仍以例 2 为例,发现证据 1~4 之间一致性较好,运用经典 D-S 方法和本文方法分别进行合成,结果见表 3。

表 3 证据合成结果

Table 3 Evidence combination results

合成方法	$m(A)$	$m(B)$	$m(C)$	$m(D)$
D-S 合成法	0.005 3	0.949 0	0.020 9	0.024 8
本文方法	0.011 2	0.971 4	0.010 8	0.006 6

由合成结果可知,对于非冲突证据的合成,本文方法具有和 D-S 方法一致的合成结果。这也证明了本文方法在处理非冲突证据时的有效性。

5 结束语

本文分析了经典 D-S 证据理论存在的 3 方面的问题,阐述了目前证据合成的研究方向,并指出了各种改进方法的不足;改进了证据理论合成公式,通过距离测度定量描述证据之间的一致性程

度,利用证据间的一致性测度和辨识框架对证据的支持度确定一致性因子,以此作为分配冲突概率的权重。改进后的合成公式适用于冲突证据和非冲突证据合成;应用实例表明,改进的证据合成公式克服了经典证据理论存在的问题,相对于其他合成方法具有较好的稳定性和精度。

参考文献:

- [1] Dempster A P. Upper and lower probabilities induced by a multi-valued mapping[J]. *Ann Mathematical Statistics*, 1967, 38: 325-339.
- [2] Shafer G. *A mathematical theory of evidence*[M]. Princeton: Princeton University Press, 1976: 10-40.
- [3] Murphy C K. Combining belief functions when evidence conflicts[J]. *Decision Support Systems*, 2000, 29(1): 1-9.
- [4] 邓勇, 施文康, 朱振福. 一种有效处理冲突证据的组合方法[J]. *红外与毫米波学报*, 2004, 23(1): 27-32.
Deng Yong, Shi Wenkang, Zhu Zhenfu. Efficient combination approach of conflict evidence[J]. *Journal Infrared Millimeter Waves*, 2004, 23(1): 27-32.
- [5] Smets P. The combination of evidence in the transferable belief model[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, 12(5): 447-458.
- [6] Yager R R. On the Dempster-Shafer framework and new combination rules [J]. *Information Sciences*, 1987, 41(2): 93-137.
- [7] Lefevre E, Colot O, Vannoorenberghe P. Belief functions combination and conflict management[J]. *Information Fusion*, 2002, 3(2): 149-162.
- [8] 李弼程, 王波, 魏俊, 等. 一种有效的证据理论合成公式[J]. *数据采集处理*, 2002, 17(1): 31-36.
Li Bicheng, Wang Bo, Wei Jun, et al. An efficient combination rule of evidence theory[J]. *Journal of Data Acquisition and Processing*, 2002, 17(1): 31-36.
- [9] 万建平, 彭天强, 李弼程. 基于证据理论的视频语义概念检测[J]. *数据采集与处理*, 2011, 26(5): 536-542.
Wan Jianping, Peng Tianqiang, Li Bicheng. Video semantic concept detection based on evidence theory [J]. *Journal of Data Acquisition and Processing*, 2011, 26(5): 536-542.
- [10] 赵一兵, 郭烈, 张明恒, 等. 越野环境中无人驾使车的障碍目标识别[J]. *数据采集与处理*, 2011, 26(4): 442-446.
Zhao Yibing, Guo Lie, Zhang Mingheng, et al. Obstacle identification in cross-country environment for unmanned ground vehicles [J]. *Journal of Data Acquisition and Processing*, 2011, 26(4): 442-446.
- [11] 李巍华, 张盛刚. 基于改进证据理论及多神经网络的故障分类[J]. *机械工程学报*, 2010, 46(9): 93-99.
Li Weihua, Zhang Shenggang. Fault classification based on improved evidence theory and multiple neural network fusion [J]. *Journal of Mechanical Engineering*, 2010, 46(9): 93-99.

作者简介:刘希亮(1983-),男,博士研究生,研究方向:信息融合与智能诊断, E-mail: purplelincoln@163.com; 陈桂明(1966-),男,教授,研究方向:机械设备状态监测与故障诊断、信息融合与模式识别; 李方溪(1978-),女,博士研究生,研究方向:机械设备状态监测与故障诊断; 张倩(1984-),女,博士研究生,研究方向:多源信息融合与优化配置。

