

基于证据分类的冲突证据合成方法

权文, 王晓丹, 史朝辉

(空军工程大学防空反导学院, 陕西三原, 713800)

摘要 证据分类策略能够很好地避免冲突证据融合问题。借鉴证据分类的思想, 定义证据-属性支持度、分类门限、证据可信度, 将证据分类, 保证同一类别的证据具有较好的一致性。在同一类别内采用 D-S 进行证据合成, 并根据各类证据的个数及证据不确定度给出每类证据合成结果的权重, 并采用加权合成方法给出最终的合成结果。

关键词 冲突证据融合; 证据分类; 合成方法

DOI 10.3969/j.issn.1009-3516.2012.03.017

中图分类号 TP212.9 **文献标识码** A **文章编号** 1009-3516(2012)03-0080-05

证据理论(Evidential Theory)^[1] 又称 Dempster-Shafer 理论(简称 DST), 作为一种不确定推理方法, 具有坚实的理论基础, 适合于无先验信息的融合, 具有利用证据积累以缩小假设集合的特殊能力。

然而在实际应用中, DST 在处理高冲突问题时, 常常得出有悖常理的结论。对此, 国内外众多学者对其进行改进, 提出了许多解决方法^[2-11]。关于改进算法的分类并没有一个统一方法。大体来说, 按系统的辨识框架可分为闭世界算法与开世界算法^[2], 按证据间的关系可分为利用证据相关性算法和去相关性算法^[3], 按组合规则分为加性策略和乘性策略^[4], 按冲突分配分为全局^[6-7]和局部分配算法^[8-10], 按修改方法分为修改证据源、修改组合规则、修改证据源结合修改组合规则方法^[11]等。

其中, 文献[4-5]中何兵提出的证据分类策略, 结合加性策略和乘性策略, 从全局考虑, 先进行乘性策略, 再用加性策略得出最终结论, 既利用了乘性策略的收敛性又利用了加性策略的可靠性, 是一种较理想的融合策略。但文献[4-5]方法在加性融合前需进行约减, 增加了计算量而且分类门限根据主观经验确定, 缺乏依据。张月玲等^[11]对何兵方法进行了改进, 提出一种基于证据相似性的证据分类策略, 以证据集内证据的平均相似性来确定分类门限, 不需要在加性融合前约减, 有效降低了分类数。

本文主要针对冲突证据的合成问题展开讨论, 针对文献[5]、[11]存在的缺陷, 根据定义的证据类别标签将证据分类, 使同一类别内的证据具有较好的一致性, 采用 D-S 进行证据合成, 并根据各类证据的个数及证据间的距离对每类的证据合成结果衡量可信度, 最后采用加权合成方法给出最终的合成结果。

1 D-S 基本理论

D-S 基本理论在相关文献中均有详述, 本文不再赘述。这里只给出 Dempster 组合规则。

定义 1 设 $m_1(\cdot)$ 和 $m_2(\cdot)$ 是 2^θ 上的 2 个相互独立的基本概率分配函数 BPA (Basic Probability Assignment, BPA), 定义组合后的 BPA: $m(\cdot) = [m_1 \oplus m_2](\cdot)$ 为:

$$\begin{cases} m(\varphi) = 0 \\ m(A) = \sum_{\substack{X, Y \subset 2^\theta \\ X \cap Y = A}} m_1(X) m_2(Y) / (1 - \sum_{\substack{X, Y \subset 2^\theta \\ X \cap Y = \varphi}} m_1(X) m_2(Y)), \forall (A \neq \varphi) \in 2^\theta \end{cases} \quad (1)$$

* 收稿日期: 2011-11-15

基金项目: 国家自然科学基金资助项目(60975026)

作者简介: 权文(1983-), 女, 陕西蒲城人, 博士生, 主要从事智能信息处理和机器学习等研究。

E-mail: 937182228@qq.com

式中 $\sum_{\substack{X, Y \subseteq \Theta \\ X \cap Y = \varnothing}} m_1(X)m_2(Y)$ 为2证据间的冲突度,记为 k_{12} ,若 $k_{12} \neq 1$,则确定一个基本概率分配函数;若 $k_{12} = 1$ 时,则认为 m_1, m_2 矛盾,不能对 BPA 进行组合。

式(1)所给出的组合规则满足结合律和交换律,适用于多个证据的组合。

2 基于证据分类的合成方法

2.1 证据分类策略

设证据 m_1, m_2, \dots, m_N 所对应属性的并集为 $\{A_i\}, i = 1, 2, \dots, m$, 属性的基元是 $\{\theta_j\}, j = 1, 2, \dots, k$ 。依据文献[5]中的关于证据-属性间支持度定义,证据 m_i 不反对属性 B_j 的度量:

$$N_i(B_j) = \sum_{A_i \cap B_j \neq \varnothing} m_i(A_i) \tag{2}$$

为了降低合成结果对证据变化的敏感度,定义门限 G (如取 0.5),并依此定义证据 m_i 所支持的属性集为:

$$\Omega_i = \left\{ B_l \mid \frac{N_i(B_l)}{N_i(B_{i,\max})} \right\}, l = 1, 2, \dots, K \tag{3}$$

从而,支持属性 $B_j, j = 1, 2, \dots, K$ 的证据构成的集合为:

$$T_{B_j} = \{m_i \mid B_j \in \Omega_i\}, i = 1, 2, \dots, n \tag{4}$$

定义集类:

$$\Gamma_T = \{T_{B_j}\}, j = 1, 2, \dots, k \tag{5}$$

从 Γ_T 删除 Γ_D 得到类集: $\Gamma = \Gamma_T / \Gamma_D$ 。

虽然文献[5]方法将相似的证据归为一类进而避免冲突证据合成的思想很新颖,但一些情况下其应用受到限制,如例 1、2 所示。

例 1 设识别框架 $\Theta = \{a, b\}$, 基本可信度分配函数见表 1。

表 1 基本可信度分配表

Tab. 1 Basic probability assignment table

证据	a	b
m_1	1	0
m_2	0	1
m_i	0.9	0.1

使用文献[5]方法得 $N(a) = 0, N(b) = 0, B_{i,\max} = \max_j \{N_i(B_j)\} = 0, \Omega_i = \left\{ B_l \mid \frac{N_i(B_l)}{N_i(B_{i,\max})} \right\}$ 中 $\frac{N_i(B_l)}{N_i(B_{i,\max})}$ 为 $\frac{0}{0}$, 无法判断 m_1, m_2 是否为一类证据。

例 2 设识别框架 $\Theta = \{a, b, c, d\}$, 基本可信度分配函数见表 2。

表 2 基本可信度分配表

Tab. 2 Basic probability assignment table

证据	a	b	c	d	$abcd$
m_1	0.95	0	0	0.05	0
m_2	0	0.8	0	0.2	0

使用文献[5]得 $N(a) = N(b) = N(c) = N(d) = 0, B_{i,\max} = \max_j \{N_i(B_j)\} = 0, \Omega_i = \left\{ B_l \mid \frac{N_i(B_l)}{N_i(B_{i,\max})} \right\}$ 中 $\frac{N_i(B_l)}{N_i(B_{i,\max})}$ 为 $\frac{0}{0}$, 无法判断 m_1, m_2 是否为一类证据。

由例 1、例 2 可得,文献[5]方法对于全集基本可信度为 0 的情况,无法判断证据是否为一类、是否可以进行乘性融合。

2.2 基于证据相似性的证据分类策略

文献[11]针对基于基元属性的证据分类方法分类数较多,分类门限根据主观经验确定的问题,提出一种基于证据相似性的证据分类策略并构造了这种分类策略下证据加性融合时的加权系数。

定义 2 m_1 和 m_2 是在辨识框架 Θ 上的 2 个 BPA, 则 m_1 和 m_2 的距离可以表示为:

$$d_{\text{BPA}}(m_1, m_2) = \sqrt{\frac{1}{2} (m_1 - m_2)^T D (m_1 - m_2)} \tag{6}$$

式中 D 为一个 $2^N \times 2^N$ 矩阵, D 中的元素为:

$$D(A, B) = \frac{|A \cap B|}{|A \cup B|}, \quad A, B \in P(\Theta) \quad (7)$$

此改进方法根据证据间距离计算出相似性,将相似性大的证据归为一类。门限值通过计算得出而非主观任意,具有加性策略之前无需约减的优势。但对于例3、例4情形,此方法会产生误判。

例3 设识别框架 $\Theta = \{a, b\}$, 基本可信度分配函数见表3。

表3 基本可信度分配表

Tab. 3 Basic probability assignment table

证据	a	b	$a \cup b$
m_1	0.6	0.3	0.1
m_2	0	1	0

文献[5]方法 $N(a) = 0, N(b) = 0, \frac{N_i(B_l)}{N_i(B_{i, \max})}$ 为 $\frac{0}{0}$, 无法判断 m_1, m_2 是否为一类证据; 文献[11]方法距离 $d = 0.6519, \text{sim} = 0.3481, G = 0.3481$, 则将证据 m_1, m_2 误判归为一类, 使用 D-S 方法: $m_{12}(a) = 0, m_{12}(b) = 1$, 无论后续证据如何支持某一焦元, 结果始终为 $m(a) = 0, m(b) = 1$ 不变, 得到与人类直觉相悖的结论。

例4 设有识别框架 $\Theta = \{\theta_1, \theta_2, \theta_3\}, i = 2, 3, \dots, l$, 基本可信度分配函数见表4。

表4 基本可信度分配表

Tab. 4 Basic probability assignment table

证据	θ_1	θ_2	$\theta_2 \cup \theta_3$	Θ
m_1	0	0.35	0.65	0
m_i	0.8	0	0	0.2

文献[5]方法 $N_1(a) = 0, N_1(b) = 0.65, N_1(c) = 0.65, N_2(a) = 0, N_2(b) = 0.2, N_2(c) = 0.2, \Omega_1 = \{b, c\}, \Omega_2 = \{b, c\}$, 则将证据 m_1, m_2 归为一类, 使用 D-S 方法: $m_{12}(a) = 0, m_{12}(b) = 1$, 得到与人类直觉相悖的结论; 文献[11]方法距离 $d = 0.8183, \text{sim} = 0.1817 > G$, 则将证据 m_1, m_2 归为一类, 无论后续证据如何支持某一焦元, 结果始终为 $m_{12}(\theta_2) = 0.35, m_{12}(\theta_2 \cup \theta_3) = 0.65$ 不变, 得到与人类直觉相悖的结论。

3 改进的证据分类策略

对文献[5]、[11]优缺点分析基础上, 本文提出一种改进的证据分类策略。

首先给出证据-属性间支持度定义:

定义3 设证据 m_1, m_2, \dots, m_N 所对应属性为 $\{A_i\}, i = 1, 2, \dots, m$, 属性的基元是 $\{\theta_j\}, j = 1, 2, \dots, k$ 。

证据 m_i 支持属性 B_j 的度量:

$$S_i(B_j) = \sum_{A_i \cap B_j \neq \varnothing} \frac{m_i(A_i)}{|A_i|} \quad (8)$$

式中 $|A_i|$ 是属性 A_i 中所包含的基元属性个数。

$G = \frac{1}{|\Theta|}$ 为门限, 支持度大于门限的属性, 为证据支持的属性。即 m_i 支持的属性集为 $\Omega_i = \{B_l \mid S_i(B_l) \geq G\}, l = 1, 2, \dots, k$, 从而, 支持属性 $B_j, j = 1, 2, \dots, k$ 的证据构成的集合为:

$$T_{B_j} = \{m_i \mid B_j \in \Omega_i, i = 1, 2, \dots, N\}$$

定义集类:

$$\Gamma_T = \{T_{B_j}\}, j = 1, 2, \dots, k \quad (9)$$

从 Γ_T 删除 Γ_D 得到类集: $\Gamma = \Gamma_T / \Gamma_D$ 。

定义4 设证据集 $\Gamma_T = (m_1, m_2, \dots, m_n)$ 中共有 n 个证据, 则证据集的可信度 C_{redi} 为:

$$C_{\text{redi}} = n e^{-\sum \frac{m_i(\Theta)}{|\Theta|}} \quad (10)$$

证据集的可信度 C_{redi} , 根据证据集内证据的个数和证据集不确定程度决定, 对数为底的形式主要是为避免 $m(\Theta)$ 或 $|\Theta|$ 为零的情况, 式中 $\sum \frac{m_i(\Theta)}{|\Theta|}$ 表示不确定性, 不确定性越小, $e^{-\sum \frac{m_i(\Theta)}{|\Theta|}}$ 越大, 确定性越大。证据集内证据越多, 确定性越大, 则可信度越高。

根据不同分类证据集的可信度 C_{redi} 确定分类合成证据的加权系数: $\omega_i = C_{\text{redi}} / \sum_{i=1}^l C_{\text{redi}}$ 。对所有分类合成

证据做加权平均可得最终的合成证据: $m_F = \sum_l \omega_l m_f^l$ 。

4 数值算例

对例1, 本文判决结果为 m_1, m_2 为不同类证据, $\text{Cred}_1 = \text{Cred}_2 = 1, \omega_1 = \omega_2 = 0.5$, 对其进行加性融合结果为: $m(a) = m(b) = 0.5$ 。

对例2, 本文判决结果为 m_1, m_2 为不同类证据, 对其进行加性融合结果为: $m(a) = 0.475, m(b) = 0.4, m(c) = 0.125$ 。

对例3, 本文判决结果为 m_1, m_2 为不同类证据, 对其进行加性融合结果为: $m(a) = 0.2925, m(b) = 0.6588, m(c) = 0.0487$ 。

对例4, 本文判决结果为 m_1, m_2 为不同类证据, 对其进行加性融合结果为: $m(a) = 0.3800, m(b) = 0.1838, m(bc) = 0.3412, m(abc) = 0.095$ 。

由例1-4 融合结果看出, 对于证据全集的基本可信度为零的情况(例1、例2)和文献[11]无法处理的仅有2 矛盾证据时误判其为一类的问题(例3、例4), 本文算法均能够有效地解决。

为了说明本算法对多个证据同样有效, 引用文献[11]的算例进行分析。

例5 设有识别框架 $\Theta = \{a, b, c, d, e, f\}$, 共有5 个证据, 分别为:

$$\begin{aligned} m_1: m_1(a, b) = 1, m_2: m_2(a, b, c) = 1, m_3: m_3(a, b, c, d) = 1, \\ m_4: m_4(f) = 1, m_5: m_5(e, f) = 1 \end{aligned}$$

按照本文的方法:

$S_1(a) = 0.5 = S_1(b) \geq 1/6, m_1$ 支持 $\{a, b\}$; $S_2(a) = 0.3333 = S_2(b) = S_2(c) \geq 1/6, m_2$ 支持 $\{a, b, c\}$; $S_3(a) = 0.3333 = S_3(b) = S_3(c) = S_3(d) \geq 1/6, m_3$ 支持 $\{a, b, c, d\}$; $S_4(a) = 1 \geq 1/6, m_4$ 支持 $\{e\}$; $S_5(e) = S_5(f) = 0.5 \geq 1/6, m_5$ 支持 $\{e, f\}$ 。

$T_{B1} = \{a, b, c, d\}, T_{B2} = \{e, f\}$ 。对 m_1, m_2, m_3 使用乘性策略 D-S 组合规则得: $m_{123}(ab) = 1, m_{45}(e) = 1$;
 $\text{Cred}_1 = 3 \times \frac{1}{e^0} = 3, \text{Cred}_2 = 2 \times \frac{1}{e^0} = 2, w_1 = \frac{3}{3+2} = 0.6, w_2 = 0.4$ 。

则最终融合结果 $m(e) = 0.4, m(ab) = 0.6$ 。

按何兵的分类策略, 初次分类 Γ_T 中共为7 种分类结果:

$$T_{B1} = \{m_1, m_2, m_3\}, T_{B2} = \{m_1, m_2, m_3\}, T_{B3} = \{m_2, m_3\}, T_{B4} = \{m_3\}, T_{B5} = \{m_4, m_5\}, T_{B6} = \{m_5\}。$$

分类后的最终合成结果为: $m(e) = 0.4746, m(ab) = 0.5254$ 。

按文献[11]的分类策略, $m(e) = 0.4502, m(ab) = 0.5498$ 。

由上例可得, 3 种方法约减后的分类结果一致, 验证了本文算法对多证据融合同样有效。且本文算法证据分类时, 只需一次计算即得证据分类结果, 不用迭代再分, 计算量小。

5 结束语

本文改进现有的证据分类策略, 给出证据一属性间支持度、分类门限定义, 在得到若干类证据集后, 证据集的可信度 C_{red} 根据证据集内证据的个数和证据集不确定程度决定, 融合结果由证据集加权合成。在同类证据间使用乘性策略, 证据集间使用加性策略, 克服了无法处理属性并集的基本可信度为零的情况和将两类证据误判为一类的问题。

参考文献 (References):

- [1] Dempster A P. Upper and lower probabilities induced by a multivalued mapping [J]. Annals of mathematical statistics, 1967, 38(2): 325 - 339.
- [2] Smets P, Kennes R. The transferable belief model [J]. Artificial intelligence, 1994, 66: 191 - 243.
- [3] 孙怀江, 杨静宇. 一种相关证据合成方法 [J]. 计算机学报, 1999, 22(3): 1004 - 1007.

- SUN Huaijiang, YANG Jingyu. A combination method for dempster evidence[J]. Chinese journal of computers, 1999,22(3): 1004 - 1007. (in Chinese)
- [4] 何兵, 毛士艺, 张有为, 等. 基于证据分类的 DS 证据合成及判决方法[J]. 电子与信息学报, 2002,24(7):894 - 899.
HE Bing, MAO Shiyi, ZHANG Youwei, et al. Evidence combination and decision based on DS evidence theory and evidence classification[J]. Journal of electronics & information technology, 2002,24(7):894 - 899. (in Chinese)
- [5] 何兵. 基于分类及不确定熵的 DS 证据合成及判决方法[J]. 北京航空航天大学学报, 2003,29(10):927 - 930.
HE Bing. Modified DS evidence combination strategy based on evidence classification and uncertain entropy[J]. Journal of Beijing university of aeronautics and astronautics, 2003,29(10):927 - 930. (in Chinese)
- [6] 邓勇, 施文康. 一种改进的证据推理组合规则[J]. 上海交通大学学报, 2003,37(8):1275 - 1278.
DENG Yong, SHI Wenkang. A modified combination rule of evidence theory[J]. Journal of Shanghai jiaotong university, 2003,37(8):1275 - 1278. (in Chinese)
- [7] Yager R R. On the dempster - shafer framework and new combination rules[J]. Information science, 1987(41):93 - 137.
- [8] 高社生, 倪龙强, 杨凯, 一种新的基于局部冲突分配的证据合成规则[J]. 西北工业大学学报, 2009, 27(1): 43 - 46.
GAO Shesheng, NI Longqiang, YANG Kai. A new and better rule for combination sharply conflicting evidence[J]. J of north-western polytechnical university, 2009, 27(1):43 - 46. (in Chinese)
- [9] Smarandache F, Dezert J. Proportional conflict redistribution rules for information fusion [M]. Rehoboth: American research press, 2006.
- [10] Dubois D, Prade H. Representation and combination of uncertainty with belief functions and possibility measures[J]. Computer intelligence, 1988(4):244 - 264.
- [11] 郭华伟, 施文康, 邓勇, 等. 一种通用的不确定性推理和决策模型[J]. 传感技术学报, 2006, (4):1176 - 1180.
GUO Huawei, SHI Wenkang, DENG Yong, et al. A modified combination rule of evidence theory[J]. Chinese Journal of sensors and actuators, 2003, 37(8): 1275 - 1278. (in Chinese)
- [12] 张月玲, 雷英杰, 王睿, 等. 一种基于证据相似性的证据分类策略[J]. 空军工程大学学报: 自然科学版, 2009, 10(3): 55 - 58.
ZHANG Yueling, LEI Yingjie, WANG Rui, et al. An evidence classification strategy based on similarity of evidence[J]. Journal of air force engineering university: natural science edition, 2009, 10(3):55 - 58. (in Chinese)

(编辑: 田新华)

Conflict Evidence Combination Strategy Based on Evidence Classification

QUAN Wen, WANG Xiao - dan, SHI Zhao - hui

(School of Air and Missile Defense, Air Force Engineering University, Sanyuan 713800, Shaanxi, China)

Abstract: Evidence classification strategy can resolve the conflict evidence combination problem. To classify the evidences and make sure of the uniformity of the evidences in one class, this paper gives the definitions of evidence - attribute support degree, classification threshold, evidence believe degree. When a series of beliefs are available, D - S rule in inner class is adopted, and the weights of every class according to the number of evidence in each class and the uncertainty of evidences are obtained, then all the combination results are integrated with weight rule to form the final results

Key words: conflict evidence combination; evidence classification combination method