

一种基于证据相似性的证据分类策略

张月玲¹, 雷英杰², 王睿², 姬云³

(1. 空军工程大学 电讯工程学院, 陕西 西安 710077; 2. 空军工程大学 导弹学院, 陕西 三原 713800; 3. 西安市阎良区建设局, 陕西 西安 710089)

摘要:证据分类策略是一种有效的冲突证据推理融合方法,但是在实际应用中会产生大量重复分类,而且分类门限根据主观经验确定,缺乏依据。为了降低分类数,提高证据分类策略的准确性,提出一种基于证据相似性的证据分类策略。首先以证据推理中得到的初始证据作为理想证据,通过衡量各个证据与理想证据的证据距离,按照证据的相似性将系统内证据分为2类证据集,然后继续按照以上的分类策略分类,直到不可再分。最后对分类结果采用 Dempster 组合规则合成,衡量各个分类的可信度,对合成结果加权平均组合。通过算例对该策略进行了验证,结果表明:基于证据相似性的证据分类策略可以有效降低分类数。

关键词:证据理论;分类策略;证据相似性

DOI:10.3969/j.issn.1009-3516.2009.03.012

中图分类号: TP391 **文献标识码:** A **文章编号:** 1009-3516(2009)03-0055-04

D-S 证据理论是一种有效的不确定性推理与决策方法,常应用于决策级数据融合^[1]。但证据组合规则在融合高度冲突证据时,会得出有悖于常理的结论^[2]。对此国内外学者提出了大量的改进方法,这些方法可以概括为2大类,一是基于修改 Dempster 组合规则的方法^[3-4],二是基于修改原始证据源的方法^[5-7]。第2类方法由于在实际应用中可以通过局域计算降低系统的运算负载,因而受到广泛关注^[8]。目前这类方法主要有 Shafer 提出的折扣运算^[5],Murphy 提出的证据平均组合规则^[6],何兵提出的基于基元属性的证据分类方法^[7]。

何兵提出的证据分类策略,由于结合了概率分布中常用的加权平均算法,在保持证据推理收敛性的同时具有良好的可靠性。尤其适用于解决冲突证据推理中的“一票否决”问题。但是分类结果随着门限变化而变化。而且当系统内基元属性较接近时,采用该分类方法易产生大量重复分类,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,分类结果随着门限变化而变化。为此本文提出一种基于证据相似性的证据分类策略。

1 证据理论基础

设 Θ 为一个包含全部可能命题的封闭识别框架。如果 Θ 的幂集 2^Θ 上的一个映射 $m:2^\Theta \rightarrow [0,1]$,且满足: $\sum_{A \in 2^\Theta} m(A) = 1, m(\emptyset) = 0$,称 $m(A)$ 为基本概率指派函数。若 $m(A) > 0$ 则称命题 A 为焦元。

假定在识别框架 Θ 上,有性质不同的2个证据,其基本概率指派函数分别为 m_1 和 m_2 ,焦元为 A_1, A_2, \dots, A_k 和 B_1, B_2, \dots, B_r ,则可按照如下的 Dempster 组合规则对 m_1 和 m_2 进行组合:

* 收稿日期:2008-12-11

基金项目:国家自然科学基金资助项目(60773209)

作者简介:张月玲(1963-),女,陕西蒲城人,副教授,主要从事数据库与信息安全研究;E-mail:Zylwz@163.com

雷英杰(1956-),男,陕西渭南人,教授,博士生导师,主要从事智能信息处理数据融合等研究。

$$\begin{cases} m(C) = \frac{1}{1 - k_{A,B_j=C}} \sum_{A_i, B_j=C} m_1(A_i) m_2(B_j), C \neq \emptyset, C \subset \Theta \\ m(\emptyset) = 0 \end{cases} \quad (1)$$

式中:

$$k = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j) \quad (2)$$

为冲突因子,表示证据的冲突程度。

2 基于证据相似性的证据分类策略

为度量系统中各个证据间的相似性程度,引入 Jousselme 等人给出的距离函数^[9]。

定义1 Θ 为一包含 N 个两两不同命题的完备识别框架, $E_{p(\Theta)}$ 是 Θ 所有子集生成的空间。一个基本概率指派函数在 $E_{p(\Theta)}$ 中的坐标为 $m(A_i)$, 可以看成为一个向量 \mathbf{m} :

$$\sum_{i=1}^{2^N} m(A_i) = 1 \quad m(A_i) \geq 0, i = 1, 2, \dots, 2^N, A_i \in P(\Theta) \quad (3)$$

定义2 Θ 为一包含 N 个两两不同命题的完备的识别框架, m_1 和 m_2 是在识别框架 Θ 上的 2 个基本概率指派函数, 则 m_1 和 m_2 的距离可以表示为:

$$d_{\text{BPA}}(m_1, m_2) = \sqrt{\frac{1}{2}(\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{D}(\mathbf{m}_1 - \mathbf{m}_2)} \quad (4)$$

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

$$D(A, B) = \frac{|A \cap B|}{|A \cup B|}, \quad A, B \in P(\Theta) \quad (5)$$

具体的计算方法是:

$$d_{\text{BPA}}(m_1, m_2) = \sqrt{\frac{1}{2}(\|\mathbf{m}_1\|^2 + \|\mathbf{m}_2\|^2 - \langle \mathbf{m}_1, \mathbf{m}_2 \rangle)} \quad (6)$$

式中 $\|\mathbf{m}\|^2 = \langle \mathbf{m}, \mathbf{m} \rangle$, $\langle \mathbf{m}_1, \mathbf{m}_2 \rangle$ 为 2 个向量的内积:

$$\langle \mathbf{m}_1, \mathbf{m}_2 \rangle = \sum_{i=1}^{2^N} \sum_{j=1}^{2^N} m_1(A_i) m_2(A_j) \frac{|A_i \cap A_j|}{|A_i \cup A_j|} \quad A_i, A_j \in P(\Theta) \quad (7)$$

定义3 证据 m_i 和 m_j 之间的相似性测度 $S_{\text{im}}(m_i, m_j)$ 为^[10]:

$$S_{\text{im}}(m_i, m_j) = 1 - d_{\text{BPA}}(m_i, m_j) \quad i, j = 1, 2, \dots, N \quad (8)$$

假设系统共得到 N 个证据 m_1, m_2, \dots, m_N , 将证据推理中得到的初始证据 m_1 作为理想证据, 首先计算证据 m_1 和 m_i ($i = 2, 3, \dots, N$) 之间的证据距离 $d_i(m, m_i)$, 得出证据之间的相似度 $S_{\text{im}}(m_i, m_j)$, 按照证据的相似性将系统内证据分为 2 类证据集 T_1 和 T_2 。

定义和 m_1 相似的证据集为 T_1 :

$$T_1 = \{m_i \mid S_{\text{im}}(m_i, m_j) \geq G, i = 1, 2, \dots, N\}$$

$$G = \frac{\sum_{j=2}^N S_{\text{im}}(m_i, m_j)}{N - 1} \quad (9)$$

式中 G 为门限, 通过证据集 Γ 内证据的平均相似性确定。

系统证据中 T_1 以外和 m_1 差异较大的证据集为 T_2 :

$$T_2 = \{m_i \mid S_{\text{im}}(m_i, m_j) < G, i = 1, 2, \dots, N\}$$

将证据集 T_2 中的第 1 个证据 m_k 作为证据集 Γ_2 内的理想证据, 对证据集 Γ_2 继续按照以上的分类方法分类, 直到不可再分。

当证据集不可再分时, 得到集类 $\Gamma = \{T_1, T_2, \dots, T_l\}$, 集类 Γ 中的每一个元素都表示一类分类结果。计算集类 Γ 中证据集 T_i 的可信度。

定义4 设证据集 $T_i = (m_1, m_2, \dots, m_n)$ 中共有 n 个证据, $S_{\text{im}}(m_i, m_j)$ 为证据集 T_i 中证据 m_i 和 m_j 的相

似性,则证据集的可信度 C_{red_i} 为:

$$C_{\text{red}_i} = \frac{\sum_{i=1, i < j}^n S_{\text{im}}(m_i, m_j)}{n(n-1)/2} \quad (10)$$

证据集的可信度 C_{red_i} 根据证据集内证据的个数和证据集整体相似度确定,证据集内证据越多,相似性越大则可信度越高。

对分类结果 $\Gamma = \{T_1, T_2, \dots, T_l\}$ 采用 Dempster 组合规则合成,得到一组合成证据 m_f^l 。

根据不同分类证据集的可信度 C_{red_i} 确定分类合成证据的加权系数:

$$\omega_l = \frac{C_{\text{red}_l}}{\sum_{i=1}^l C_{\text{red}_i}} \quad (11)$$

对所有分类合成证据做加权平均可得最终的合成证据:

$$m_F = \sum_l \omega_l m_f^l \quad (12)$$

3 算例验证

假设在识别框架 $\Theta = \{a, b, c, d, e, f, g, h\}$ 下系统中共有 6 个证据,分别为:

$$m_1: m_1(a, b) = 1, \quad m_2: m_2(a, b, c) = 1, \quad m_3: (a, b, c, d) = 1,$$

$$m_4: m_4(a, b, c, d, e) = 1; \quad m_5: m_5(f) = 1, \quad m_6: m_6(f, g) = 1$$

$$S_{\text{im}}(m_1, m_2) = 0.67, S_{\text{im}}(m_1, m_3) = 0.5, S_{\text{im}}(m_1, m_4) = 0.4,$$

$$S_{\text{im}}(m_1, m_5) = S_{\text{im}}(m_1, m_6) = 0, G = \frac{\sum_{j=2}^N S_{\text{im}}(m_1, m_j)}{N-1} = 0.314$$

按照本文的方法根据门限 G 将系统内证据分为 2 类证据集 T_1 和 T_2 :

$$T_1 = \{m_1, m_2, m_3, m_4\}, \quad T_2 = \{m_5, m_6\}$$

计算 T_1 和 T_2 的可信度为:

$$C_{\text{red}_1} = 0.52, C_{\text{red}_2} = 0.3$$

最终合成结果为:

$$m_F(a, b) = 0.63, m_F(f) = 0.37$$

采用何兵的分类策略,假设门限 $G = 0.5$,初次分类 Γ_T 中共有 7 种分类结果,分别为:

$$T_{B_1} = \{m_1, m_2, m_3, m_4\}; T_{B_2} = \{m_1, m_2, m_3, m_4\}; T_{B_3} = \{m_2, m_3, m_4\};$$

$$T_{B_4} = \{m_3, m_4\}; T_{B_5} = \{m_4\}; T_{B_6} = \{m_6, m_7\}; T_{B_7} = \{m_7\}$$

约减后的分类结果与本文一致。

分类后的最终合成结果为:

$$m_F: m_F(a, b) = 0.5962, m_F(f) = 0.4038$$

通过以上算例可以看出,何兵提出的分类策略,将不存在冲突的 m_1, m_2, m_3, m_4 按照基元支持度分为 7 类,在加性融合前需进行约减。本文方法和何兵的方法相比在加性融合前不需要约减,而且根据证据间的相似性给出了确定分类门限的方法。

4 结束语

针对基于基元属性的证据分类方法,在实际应用中分类较多,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。该分类策略通过考虑证据间的相似性,有效降低了实际应用中的分类数,而且根据证据间的相似性给出了确定分类门限的

方法。算例表明本文提出的方法在降低分类数的同时,融合结果仍旧保持了一定的收敛性和可靠性。

参考文献:

- [1] 何 友,王国宏. 多传感器信息融合及应用[M]. 北京:电子工业出版社,2000.
HE You,WANG Guohong. Multi - sensor Fusion and Its Application [M]. Beijing: Electronic Industry Press,2000. (in Chinese)
- [2] Zadeh L A. A Simple View of the Dempster - shafer Theory of Evidence and Its Implication for the Rule of Combination [J]. AI Magasine,1986, 7(2):85 - 90.
- [3] Yager R. On the Dempster - Shafer Framework and New Combination Rules [J]. Information Science,1989,41(2):93 - 137.
- [4] Lefevre E, Colot O, Vannoorenberghe P. Belief Functions Combination and Conflict Managent [J]. Inf Fusion,2002, 3(2): 149 - 162.
- [5] Shafer G. A Mathematical Theory of Evidence [M]. Princeton:Princeton University Press, 1976.
- [6] Murphy C K. Combining Belief Functions When Evidence Conflicts [J]. Decision Support System,2000,29(1):1 - 9.
- [7] 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,2003,10(24):927 - 930.
HE Bing. Modified DS Evidence Combination Strategy Based on Evidence Classification and Uncertain Entropy [J]. Journal of Beijing University of Aeronautics and Astronautic, 2003,10(24):927 - 930. (in Chinese)
- [8] Haenni R. Are Alternative to Dempster's Rule of Combination Real Alternative? Comments on About the Belief Function Combination and the Conflict Management Problem - E [J]. Inf Fusion,2002, 3(3): 237 - 239.
- [9] Jousselme A L, Grenier D, Bosse E. A New Distance between Two Bodies of Evidence [J]. Inf Fusion,2001, 2(2): 91 - 101.
- [10] 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,2004,1(23):27 - 32.
DENG Yong, SHI Wenkang, ZHU Zhenfu. Efficient Combination Approach of Conflict Evidence [J]. Journal Infrared Millimeter Wave,2004,1(23):27 - 32. (in Chinese)

(编辑:田新华)

An evidence classification strategy based on similarity of evidence

ZHANG Yue - ling¹, LEI Ying - jie², WANG Rui², JI Yun³

(1. Telecommunication Engineering Institute, Air Force Engineering University, Xi'an 710077, China; 2. Missile Institute, Air Force Engineering University, Sanyuan 713800, Shaanxi, China; 3. Infrastructure Construction Bureau Yanliang District, Xi'an 710089, China)

Abstract: The evidence classification strategy is a useful conflict evidence reasoning and fusion method, but in practice it would bring a great deal of repetitive class. In order to reduce the number of class and improve the efficiency of evidence classification strategy, an evidence classification strategy based on similarity of evidence is proposed. First by taking the initially evidence in evidence reasoning as the ideality evidence, scaling the distance between each evidence and the ideality evidence, the evidences in the system are divided into two classes according to the similarity of them. Then the classification will continue with the above classification strategy till it will be impossible. Finally the following workings such as combining the classified results with Dempster combination rule, scaling the reliability of the classification results and weighting the Dempster combination result are done. The strategy is validated through an example. The result shows that the evidence classification strategy based on similarity of evidence can reduce the number of class effectively.

Key words: evidence theory; classification strategy; similarity of evidence