

文章编号: 1001-0920(2008)02-0229-04

# 基于模糊集的证据组合方法及其应用

郭惠昕

(长沙大学 机电工程系, 长沙 410003)

**摘要:** 为解决高度冲突证据的合成问题, 提出一种基于模糊集的证据组合方法. 对证据集中的所有证据对的冲突权进行计算, 构造了冲突权矩阵, 并提出了一种根据冲突权大小甄别冲突源的方法. 引入模糊集理论, 根据证据所引起的冲突权大小定义了证据的模糊可信度. 按模糊可信度对证据进行预处理, 然后用 Dempster 法则进行组合. 实例表明, 该方法能有效处理冲突证据的合成问题, 保证了冲突证据合成的合理性和可靠性.

**关键词:** 证据理论; 冲突证据; 模糊可信度; 组合规则

中图分类号: TP274

文献标识码: A

## Approach to evidence combination based on fuzzy theory and its applications

GUO Hui-xin

(Department of Mechanical and Electronic Engineering, Changsha University, Changsha 410003, China. E-mail: xinhuiguo@126.com)

**Abstract:** A combination rule is proposed in order to efficiently combine highly conflicting evidences. The conflict weights of every two evidences in evidence set are calculated, and a conflict matrix is composed of the calculated conflict weights. A method for distinguishing the evidences bringing on conflicts is presented according to the conflict matrix. Based on fuzzy theory, the fuzzy credibility degree of evidence is defined and formulated as a function of the conflict weights. Each of evidences is preprocessed according to its fuzzy credibility degree, and then the Dempster's rule of combination is used to combine the preprocessed evidences. Examples show that the proposed approach can efficiently eliminate the influence of conflicts between evidences on evidence combination and can achieve a reliable combination result with rationality.

**Key words:** Evidence theory; Conflict evidence; Fuzzy credibility; Combination rule

### 1 引言

证据理论由 Dempster 于 1967 年首先提出, 后由 Shafer 进行了进一步研究与发展, 故又称为 D-S 证据理论. 证据理论具有较强的处理不确定信息的能力, 已经成为不确定信息的表达和处理的有力工具, 在多目标识别、多传感器信息融合、不确定多属性决策等领域中获得了广泛应用. 然而, 在证据具有较大冲突的情况下, D-S 证据组合可能会产生与直觉相反的结论. 因此, 如何有效地处理高度冲突的证据融合问题, 已经成为证据理论研究的热点之一<sup>[1-4]</sup>. 在研究证据体冲突原因的基础上, 对冲突源进行识别, 并引入模糊集理论对证据的可信性进行判决, 进而提出了一种新的冲突证据组合方法.

### 2 D-S 证据理论及其缺点

以下仅简介 D-S 证据理论的基本概念与证据组合的基本原理<sup>[5]</sup>. 设  $\Theta$  为辨识框架, 如果集函数  $m: 2^\Theta \rightarrow [0, 1]$  ( $2^\Theta$  为  $\Theta$  的幂集) 满足

$$m(\emptyset) = 0, \quad m(A) = 1, \quad A \in \mathcal{A}$$

则称  $m$  为框架  $\Theta$  上的基本可信度分配;  $\forall A \in \mathcal{A}$ ,  $m(A)$  称为  $A$  的基本可行数. 设  $\Theta$  为辨识框架,  $m: 2^\Theta \rightarrow [0, 1]$  为辨识框架  $\Theta$  上的基本可信度分配, 则由

$$\text{Bel}(A) = \sum_{B \subset A} m(B), \quad \forall A \in \mathcal{A}$$

所定义的函数  $\text{Bel}: 2^\Theta \rightarrow [0, 1]$  为  $\Theta$  上的信度函数. 设  $\text{Bel}_1$  和  $\text{Bel}_2$  是同一辨识框架  $\Theta$  上的两个信度函数,  $m_1$  和  $m_2$  分别是其对应的基本可信度分配, 焦点元分别为  $A_1, \dots, A_k$  和  $B_1, \dots, B_j$ . 设

收稿日期: 2006-10-28; 修回日期: 2007-03-19.

基金项目: 湖南省自然科学基金项目(00JJ Y2050); 湖南省教育厅高等学校科研重点项目(03A031).

作者简介: 郭惠昕(1962—), 男, 湖南桃源人, 教授, 硕士, 从事智能信息处理、智能优化设计等研究.

$$c = \frac{m_1(A_i) m_2(B_j)}{A_i \cap B_j = \phi} < 1, K = 1 - c,$$

那么由下式定义的函数  $m: 2^{\Omega} \rightarrow [0, 1]$  是两个证据合成后的基本可信度分配:

$$m(A) = \begin{cases} 0, & A = \phi; \\ K^{-1} \frac{m_1(A_i) m_2(B_j)}{A_i \cap B_j = A}, & A \neq \phi \end{cases} \quad (1)$$

式(1)就是两个证据合成的 Dempster 法则,  $m(A)$  为两个证据合成后  $A$  的信度大小. 当  $A = \phi$  时,  $c$  为分配到空集  $\phi$  上的一部分信质, 但式(1)丢弃了这部分信质, 定义此时  $A$  的信度为零. 为了保证总信质为 1, 式(1)中引入了归一化因子  $K^{-1}$ .  $c$  是由于所融合的各个证据间存在冲突引起的, 冲突越大,  $c$  越大. 当  $c = 1$  时, 将不能直接用 Dempster 法则进行证据合成. 当  $c < 1$  时, 由于证据间冲突严重, 将导致与直觉相悖的结果.

Zadeh 引用了关于两个医生的例子来说明这个问题, 称为 Zadeh 悖论<sup>[6]</sup>. 对于同一个患者, 医生诊断的可能性即识别框架  $\Omega = \{A, B, C\} = \{\text{脑膜炎}, \text{脑瘤}, \text{脑震荡}\}$ . 两个医生经过诊断后分别提出如下的基本可信度分配:  $m_1(A) = 0.99, m_1(B) = 0.01, m_1(C) = 0; m_2(A) = 0, m_2(B) = 0.01, m_2(C) = 0.99$ . 若直接按照 Dempster 合成规则, 则有  $m(A) = m(C) = 0, m(B) = 1$ . 这意味着 Dempster 合成规则将 100% 的信任分配给了脑瘤, 可以被确定患有脑瘤. 但是, 从医生 1 和医生 2 各自的诊断结果来看, 他们都认为该患者患有脑瘤的可能都很小, 仅为 0.01, 显然合成结果与这种直观分析是相悖的.

### 3 基于模糊集理论的证据合成方法

在多目标识别、多传感器数据融合等实际的决策系统中, 由于各种不确定因素的干扰, 某些传感器的输出信息具有较大的不确定性, 甚至与实际情况相悖, 导致与其他传感器输出的信息有较大的冲突, 如果用式(1)进行证据合成将导致决策失误. 为了克服 Dempster 法则的缺点, 国内外学者提出了不同的改进思路与改进方法. 第 1 种改进思路: 认为 Dempster 法则是概率论中贝叶斯方法的推广, 具有坚实的数学基础, 故坚持 Dempster 法则的正确性, 在证据严重冲突时先对冲突证据进行预处理, 然后用 Dempster 法则进行融合<sup>[1,3]</sup>. 第 2 种改进思路: 认为 Dempster 法则中的归一化处理比较勉强甚至不合理, 分配到空集  $\phi$  上的一部分信质  $c$  不应丢弃, 应按一定的法则分配给其他命题, 并提出了不同于式(1)的多证据合成公式<sup>[2,4,7]</sup>.

设决策系统的信息源大多数是可靠的, 否则该

系统已经失效. 证据间的冲突应来源于干扰因素对信息源(证据)可靠性的不良影响, 如果多信息源的可靠性高, 获得的证据反映了真实情况, 则多证据间应该不存在冲突, 完全可以用 Dempster 法则进行融合. 证据之间的冲突是因为某个(或某些)证据的不可靠引起的, 而且其可靠性越差, 引起的证据间的冲突也将越大. 如果把可靠性高的证据称为强证据, 则可靠性差的证据可称为弱证据. 由于 Dempster 法则本身不能识别强证据和弱证据, 要得到合理的证据合成结果, 有必要在合成之前对证据集的冲突及冲突源进行分析, 并按证据的可靠性对证据集进行预处理.

**定义 1** 设  $\Omega$  为辨识框架,  $Bel_i, Bel_j$  为  $\Omega$  上的 2 个信度函数, 对应的基本可信度分配为  $m_i, m_j$ , 焦化元分别为  $A_1, \dots, A_k$  和  $B_1, \dots, B_k$ . 证据  $i$  与证据  $j$  之间的冲突权值  $c_{ij}$  定义为

$$c_{ij} = \frac{m_i(A_k) \cdot m_j(B_k)}{A_k \cap B_k = \phi} \quad (2)$$

显然,  $c_{ij} = c_{ji}, c_{ij} \in [0, 1]$ .  $c_{ij}$  越大, 两个证据间的冲突越大.

**定义 2** 设  $\Omega$  为辨识框架,  $Bel_1, \dots, Bel_M$  为  $\Omega$  上的  $M$  个信度函数, 对应的基本可信度分配为  $m_1, \dots, m_M$ . 定义由  $M$  个证据所组成的证据集的冲突权矩阵为

$$C = \begin{bmatrix} c_{11} & \dots & c_{1M} \\ \dots & \ddots & \dots \\ c_{M1} & \dots & c_{MM} \end{bmatrix}_{M \times M} \quad (3)$$

在此基础上, 可以计算出第  $i$  个证据所引起的证据相互间的冲突量为

$$C_i = \sum_{t=1}^M C_{it} - C_{ii}, \quad i = 1, 2, \dots, M. \quad (4)$$

令

$$C_{\max} = \max\{C_i, i = 1, 2, \dots, M\},$$

$$C_{\min} = \min\{C_i, i = 1, 2, \dots, M\}.$$

#### 3.1 证据冲突分析与证据集的相容性判断

##### 3.1.1 证据个数 $M$ 焦化元个数 $k$

分析冲突权矩阵(3), 令

$$C_i^{\min} = \min\{c_{ij}, c_{ji} \mid j = 1, 2, \dots, M, i \neq j\}, \quad i = 1, 2, \dots, M.$$

判断

$$\frac{1}{1 - C_i^{\min}} \leq C_{\max}, \quad \text{或} \quad C_i^{\min} = 1 \quad (5)$$

是否成立, 其中  $C_{\max}$  为给定的域值, 根据多次仿真计算的统计分析, 取 3.5 比较合适. 若式(5)对所有的  $i$  均不成立 ( $i = 1, 2, \dots, M$ ), 则证据集中的所有证据是相容的, 证据间不存在较大冲突, 可称之为相容证据

集;若式(5)对某个或部分*i*成立,则第*i*个证据至少与证据集中的另外一个证据存在严重冲突,该证据即为一个冲突源,此时的证据集可称之为部分相容证据集;若式(5)对所有的*i*均成立,则证据集的所有证据两两互不相容,存在最严重的冲突,可称之为不相容证据集.显然,证据冲突最极端的情况发生在证据个数  $M =$  焦化元个数  $k$ ,且证据集的各个证据分别明确支持识别框架中的不同焦化元的情形.

### 3.1.2 证据个数 $M >$ 焦化元个数 $k$

不难分析,此时证据集中至少有一对证据同时较大地支持了识别框架中的同一个焦化元,即至少有一对证据是相容或基本相容的.此时,可进一步用上述方法判断证据集属于相容证据集还是部分相容证据集.

### 3.2 基于模糊集的证据可信性判断

证据冲突来源于信息的不可靠,而导致证据可靠性不高的原因很复杂,如传感器性能的变异、环境因素的干扰等.某证据引起的冲突大,则其可靠性差,含有的不确定因素就较多,该证据的可信度就低.由于系统不确定因素的复杂性,信息源的可靠性很难被确定地描述和量化.以下引入模糊集理论,用模糊集的隶属函数来定量描述证据的可信度.

对于部分相容证据集,根据模糊集理论,与证据集中的其他证据冲突最大的证据含有的不确定信息最多,故该信息的模糊度最大,它对“可信”的隶属度应该为 0.5.而对于与其他证据冲突最小的证据,它含有的不确定信息最少,可以认为它最可信,它对“可信”的隶属度应该为 1.因此,第  $i$  个证据( $i = 1, 2, \dots, M$ )的模糊可信度定义为

$$\mu_i = \begin{cases} 1 - 0.5 \left( \frac{c_i - c_{\min}}{c_{\max} - c_{\min}} \right)^2, & c_i < c_{\max}; \\ 0.5, & c_i = c_{\max}. \end{cases} \quad (6)$$

对于相容的证据集,每个证据的模糊可信度定义为 1;而对于不相容证据集,因为证据集中的任意一对证据均存在较大冲突,在没有其他补充证据的情况下,无法判断哪个证据较为可信,故每个证据的模糊可信度均定义为 0.5.对于部分相容证据集,其中各证据的模糊可信度由式(6)计算.

### 3.3 基于模糊可信度的证据预处理

对于模糊可信度  $< 1$  的证据,含有一定的不确定因素,应该把原来分配给各焦化元的信质向不确定元素  $U$  ( $U$  为辨识框架中所有焦化元的并集)转移.设  $\Omega$  为辨识框架,  $Bel$  为  $\Omega$  上的信度函数,对应的基本可信度分配为  $m$ ,焦化元分别为  $A_1, \dots, A_k$ .若某证据  $m_i$  的模糊可信度为  $\mu_i$ ,则经过预处理后的基本可信度分配为

$$m_i(\Phi) = 0, \quad m_i(A_j) = \mu_i \cdot m_i(A_j), \quad j = 1, 2, \dots, k, \\ m_i(U) = \mu_i \cdot m_i(U) + (1 - \mu_i).$$

### 3.4 基于模糊集理论的证据组合算法

Step1: 根据已获得的证据集,由式(2) ~ (4) 计算证据集的冲突权矩阵.

Step2: 计算并判断式(5)是否满足,并根据  $M$  和  $k$  的关系按 3.1 节方法判断证据集的相容性.

Step3: 证据集的预处理:对于部分相容证据集,按式(6)计算各证据的模糊可信度;对于不相容证据集,置各证据的模糊可信度为 0.5;然后用 3.3 节方法进行证据预处理.相容证据集不需预处理.

Step4: 对预处理后的证据集,用 Dempster 法则进行证据组合.

根据上述证据合成新方法的原理与算法,编写了多证据合成的计算机程序,程序用 Matlab 语言实现.

## 4 应用实例

例 1 用上述方法来研究 Zadeh 悖论.因 2 个医生的诊断严重冲突,在没有其他医生的补充诊断的情况下,无法判断哪个医生的诊断结果可信度较高,故 2 个诊断的模糊可信度均为 0.5.经过预处理后用 Dempster 法则组合,结果为

$$m(A) = 0.3300, \quad m(B) = 0.0067, \\ m(C) = 0.3300, \quad m(U) = 0.3333.$$

从组合情况分析,患者患脑瘤的信度仍然很小,患脑膜炎和脑震荡的信度相同,但不能确定患 {脑膜炎,脑瘤,脑震荡} 中何种疾病的信度  $m(U)$  较大,因此不能确诊.要对患者确诊,还需进一步补充诊断.该组合结果与人的直觉推理相容,较好地解释了 Zadeh 悖论.

例 2 为了验证和比较本文方法的多信息融合能力,取文献[8]中的实例进行验证.设有一个由多传感器组成的自动目标识别系统,辨识框架为  $\Omega = \{A, B, C\}$ ,系统已经获得来源于 5 个传感器产生的信息(即证据),分别为

$$m_1: m_1(A) = 0.6, m_1(B) = 0.2, m_1(C) = 0.3; \\ m_2: m_2(A) = 0, m_2(B) = 0.9, m_2(C) = 0.1; \\ m_3: m_3(A) = 0.55, m_3(B) = 0.1, m_3(C) = 0.35; \\ m_4: m_4(A) = 0.55, m_4(B) = 0.1, m_4(C) = 0.35; \\ m_5: m_5(A) = 0.55, m_5(B) = 0.1, m_5(C) = 0.35.$$

取不同数目的证据组成证据集,用所提出的算法进行证据合成,结果列于表 1.为便于比较,表中还列出了用其他方法进行目标识别的结果.

分析表 1 合成结果,可以得到如下结论:

1) Dempster 合成法则和 Yager 合成法则均不能直接处理有较大冲突的多证据合成问题,两种方

表 1 用不同证据合成方法进行目标识别的结果及其对比

方 法	证 据 集			
	$m_1, m_2$	$m_1, m_2, m_3$	$m_1, m_2, m_3, m_4$	$m_1, m_2, m_3, m_4, m_5$
Dempster 证据合成 法则	$m(A) = 0$	$m(A) = 0$	$m(A) = 0$	$m(A) = 0$
	$m(B) = 0.857\ 1$	$m(B) = 0.631\ 6$	$m(B) = 0.328\ 8$	$m(B) = 0.122\ 8$
	$m(C) = 0.142\ 9$	$m(C) = 0.364\ 8$	$m(C) = 0.671\ 2$	$m(C) = 0.877\ 2$
Yager 证据 合成法则	$m(A) = 0$	$m(A) = 0$	$m(A) = 0$	$m(A) = 0$
	$m(B) = 0.18$	$m(B) = 0.018$	$m(B) = 0.001\ 8$	$m(B) = 0.000\ 2$
	$m(C) = 0.03$	$m(C) = 0.010\ 5$	$m(C) = 0.003\ 7$	$m(C) = 0.001\ 3$
Murphy 的 证据合成 方法 <sup>[1]</sup>	$m(A) = 0.154\ 3$	$m(A) = 0.350\ 4$	$m(A) = 0.602\ 7$	$m(A) = 0.795\ 8$
	$m(B) = 0.746\ 9$	$m(B) = 0.523\ 1$	$m(B) = 0.262\ 7$	$m(B) = 0.093\ 2$
	$m(C) = 0.098\ 8$	$m(C) = 0.126\ 5$	$m(C) = 0.134\ 6$	$m(C) = 0.111\ 0$
文献[8]的 证据合成 方法	$m(A) = 0.154\ 3$	$m(A) = 0.462\ 6$	$m(A) = 0.741\ 9$	$m(A) = 0.882\ 7$
	$m(B) = 0.746\ 9$	$m(B) = 0.384\ 5$	$m(B) = 0.112\ 0$	$m(B) = 0.014\ 2$
	$m(C) = 0.098\ 8$	$m(C) = 0.152\ 9$	$m(C) = 0.146\ 1$	$m(C) = 0.103\ 1$
基于模糊 集的证据 合成方法	$m(A) = 0.155\ 8$	$m(A) = 0.613\ 5$	$m(A) = 0.771\ 4$	$m(A) = 0.851\ 2$
	$m(B) = 0.398\ 8$	$m(B) = 0.116\ 6$	$m(B) = 0.020\ 7$	$m(B) = 0.003\ 9$
	$m(C) = 0.134\ 0$	$m(C) = 0.269\ 9$	$m(C) = 0.207\ 9$	$m(C) = 0.144\ 9$
	$m(U) = 0.315\ 5$	$m(U) = 0.0$	$m(U) = 0.0$	$m(U) = 0.0$

法都不能正确识别目标  $A$ 。由于证据  $m_2$  否定了  $A$ ，不管以后积累的证据多么支持  $A$ ，始终有  $m(A) = 0$ 。用 Yager 合成法则时，随着证据的增加，不确定项  $m(\ )$  不断增加，这与人的直觉推理相悖。

2) 当只有 2 个证据，且它们存在较大冲突时（如实例中的  $m_1, m_2$ ），不论用哪种方法合成，均不能正确识别目标。要正确识别目标，必须补充其他证据，这和人的直觉推理是相容的。

3) 当证据积累达到 4 个时，文献[1]和文献[8]方法均能正确识别目标  $A$ ；而当证据积累达到 3 个时，本文方法就能正确识别目标  $A$ ，说明本文证据组合方法对目标的识别效率以及识别的准确性更高，处理冲突证据更可靠、更合理。

4) 当证据累计达到 3 个及以上时，基于模糊集的证据可信性判断方法能准确地判断证据集中的冲突源，赋予冲突源较小的模糊可信度，并通过有利证据的不断积累逐步降低分配给不确定事件的基本可信度  $m(U)$ ，提高了证据推理的合理性。

## 5 结 语

在多信息融合系统中，由于传感器本身的物理变异、使用环境的影响、敌方人为的干扰等原因，可能导致少数信息失真，从而引起所收集的证据之间存在冲突甚至较大冲突。如何甄别这些失真的证据并降低对系统决策带来的不确定影响，确保决策的正确性，是具有现实意义且值得深入研究的课题。所提出的基于模糊集的证据组合方法，能有效判断证据集中的冲突源，通过证据的模糊可信度对证据进行预处理，然后用 Dempster 合成法则进行合成。实例表明，该方法可以有效地处理高度冲突的证据，提高了证据合成的合理性和可靠性。

## 参考文献(References)

- [1] Murphy C K. Combining belief functions when evidence conflicts[J]. Decision Support Systems, 2000, 29(1): 1-9.
- [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, 28(8): 117-119.  
(Sun Quan, Ye Xiu-qing, Gu Wei-kang. A combination rule of evidence[J]. Acta Electronica Sinica, 2000, 28(8): 117-119.)
- [3] Jousselme A L, Grenier D, Bosse E. A new distance between two bodies of evidence[J]. Information Fusion, 2001, 2(1): 91-101.
- [4] 张山鹰, 潘泉, 张洪才. 一种新的证据推理组合规则[J]. 控制与决策, 2000, 15(5): 540-544.  
(Zhang Shan-ying, Pan Quan, Zhang Hong-cai. A new kind of combination rule of evidence theory[J]. Control and Decision, 2000, 15(5): 540-544.)
- [5] 段新生. 证据理论与决策、人工智能[M]. 北京: 中国人民大学出版社, 1993.  
(Duan Xin-sheng. Evidence theory & decision and artificial intelligence [M]. Beijing: China Renmin University Press, 1993.)
- [6] Joseph Garratano, Gary Riley. Principle of expert system and programming[M]. Beijing: China Machine Press, 2001.
- [7] Yager R R. On the dempster-shafer framework and new combination rules[J]. Information Sciences, 1987, 41(2): 93-137.
- [8] Chen Liang-zhou, Shi Wen-kang, Deng Yong, et al. A new fusion approach based on distance of evidences[J]. J of Zhejiang University Science, 2005, 6A(5): 476-482.