

余鑫, 吴根秀, 蔡奥丽. 基于双重中位数绝对偏差的证据加权组合方法 [J]. 江西师范大学学报(自然科学版) 2022 46(5): 523-532.

YU Xin, WU Genxiu, CAI Aoli. The improved evidence weighted combination method based on double median absolute detection [J]. Journal of Jiangxi Normal University(Natural Science) 2022 46(5): 523-532.

文章编号: 1000-5862(2022) 05-0523-10

基于双重中位数绝对偏差的证据加权组合方法

余 鑫, 吴根秀*, 蔡奥丽

(江西师范大学数学与统计学院, 江西 南昌 330022)

摘要: Dempster-Shafer 证据理论目前已被广泛应用于大数据时代的各行各业, 但是当该理论应用在高度冲突的证据源进行融合时, 往往会产生一些有悖直觉的结果, 具有一定的局限性. 为了解决这一问题, 该文提出了一种基于双重中位数绝对偏差(MAD) 检测和一种新的证据加权组合改进方法. 首先通过 MAD 算法检测出异常证据, 再使用这组证据的平均值对异常证据进行修正, 然后使用新的证据加权组合方法对证据加权平均后得到最终结果, 最后通过随机模拟实验和具体的算例实验, 并与其他几种经典的证据合成规则进行比较. 仿真结果表明该文提出的方法能有效地解决高度冲突的证据融合问题.

关键词: D-S 证据理论; 中位数绝对偏差; 高度冲突

中图分类号: TP 391 文献标志码: A DOI: 10. 16357/j. cnki. issn1000-5862. 2022. 05. 13

0 引言

证据理论是 20 世纪 60 年代末至 70 年代初由 A. P. Dempster^[1] 提出的, 并由他的学生 G. Shafer^[2] 进一步修改并发展, 因此也称该理论为 D-S 证据理论. D-S 证据理论在证据合成中是迄今为止使用最频繁的一种证据融合规则, 被广泛地应用于信息融合、神经网络、故障诊断等领域^[3-9]. 由于证据会受到不确定性因素的影响, 所以当证据融合时往往也会得到不确定性的结果. D-S 合成规则需要解决的核心问题是: 在面对高度冲突的证据融合时会产生不符合直觉的结果, 如 L. A. Zadeh^[10] 提出的悖论、“一票否决”问题等. 基于此, 近几十年来, 国内外许多学者针对 D-S 证据理论在进行证据融合时产生的问题提出了许多改进的方法. 这些方法主要分为 2 类. 一类是修改 D-S 证据理论的合成规则. 此类方法认为 D-S 本身存在一定的局限性, 在面对高度冲突的证据融合时没有做到合理地分配而产生了有悖直觉的结果; 其中有 R. R. Yager^[11] 的合成规则、孙全等^[12] 的合成规则、吴根秀^[13] 的合成规则等. R. R.

Yager 的合成规则是直接将冲突的信息分配给识别框架. 这种方法认为冲突是由客观的无知造成的. R. R. Yager 的合成规则虽然解决了在证据完全冲突时的 D-S 合成规则无法合成的问题, 但高度冲突证据的加入会导致合成结果难以识别目标和无法做出相应的决策. 孙全等的合成规则是引入了可信度来度量证据之间的可信程度, 并认为证据之间存在冲突部分也是可用的, 然而该方法的合成效率不高. 吴根秀的合成规则是在引入自冲突概念的基础上提出了一种权重分配算法, 最后通过算例计算得到了较好的融合结果. 另一类是修改原始证据源. 这种方法认为高度冲突证据是受到不确定因素的干扰导致的, 因此在证据融合之前需要将证据预处理, 即修改原始的证据源; 其中有 C. K. Murphy^[14] 的合成规则、邓勇等^[15] 的合成规则. C. K. Murphy 的合成规则是先将原始证据的基本概率分配(BPA) 函数值取算术平均修改成新的证据值, 然后再将修改后的证据值进行 D-S 合成, 该方法没有充分考虑不同证据之间的影响与联系. 邓勇等的合成规则是为了考虑证据之间的联系, 引入一个度量证据与证据之间的距离的 Jousseleme 函数^[16], 并进一步获得证据之间

收稿日期: 2022-07-01

基金项目: 国家自然科学基金(61876074) 资助项目.

通信作者: 吴根秀(1965—), 女, 江西南丰人, 教授, 主要从事不确定性推理与信息融合研究. E-mail: wgx_nc@sina.com

的相似度和由其他证据产生的各个证据的支持度,最后将证据进行加权后再利用 D-S 合成规则进行证据融合.该方法虽然改进了 C. K. Murphy 的合成规则的不足,但是忽略了证据自身的影响.

以上几种方法在一定程度上都是对经典 D-S 证据理论合成规则做出的改进,但对证据融合效率有待提高.基于此,本文提出了一种既考虑证据源对融合的影响,又考虑 D-S 合成规则自身缺陷的新合成方法.该方法首先认为证据融合效果不好是证据源本身的问题(即存在异常证据),在证据融合之前需要先对异常证据进行修改,认为冲突较高的证据是受到某些不确定的因素影响导致的;然后通过引入双重中位数绝对偏差(MAD)检测出异常证据,将异常证据使用平均值法进行修正;最后将修改的证据进行新的加权组合得到最终结果.

1 预备知识

1.1 基础知识

定义 1 设 $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ 为一个有限识别框架,映射 $m: 2^\Theta \rightarrow [0, 1]$ 满足

$$\begin{cases} m(\emptyset) = 0, \\ \sum_{A \subseteq \Theta} m(A) = 1, \end{cases}$$

则称 m 为 Θ 上的基本概率分配(basic probability assignment, BPA)函数,简称 BPA 函数.称 $m(A)$ 为 A 的基本概率分配; $\forall A \subseteq \Theta$, $m(A)$ 表示对 A 的支持度,若 $m(A) > 0$, 则称 A 为 m 的焦点.所有焦点的集合构成基本概率分配的核(core),记为 $C(m)$.

1.2 D-S 合成规则及其存在的问题

为了将 n 个独立信息源的证据进行融合, Dempster-Shafer 提出一种合成规则,其本质是将这些证据进行直交和计算. D-S 合成规则如下: 设 m_1, m_2, \dots, m_n 为幂集 2^Θ 上的 n 个基本概率分配函数, 则 D-S 合成公式为

$$\begin{cases} m(\emptyset) = 0, \\ m_{D-S}(A) = \frac{1}{1-K} \sum_{\bigcap_{i=1}^n A_i = A} \prod_{i=1}^n m_i(A_i), \end{cases} \quad (1)$$

其中 $K = \sum_{\bigcap_{i=1}^n A_i = \emptyset} \prod_{i=1}^n m_i(A_i)$ 为证据的冲突程度.

在经典的 D-S 合成规则中,当在 2 条证据融合后 K 值为 1 时,这 2 条证据完全冲突,此时 D-S 合成规则就失效了;当 $K \rightarrow 1$ 时,该 2 条证据之间高度冲突,此时在进行证据的融合后将会产生与直觉不符

的结果,例 1 和例 2 说明了这种情况.

1.2.1 Zadeh 悖论问题

例 1 识别框架 $\Theta = \{a, b, c\}$ 2 条证据 m_1 和 m_2 对该识别框架的各个目标基本概率分配分别为

$$\begin{aligned} m_1: m_1(a) &= 0.99, m_1(b) = 0.01, m_1(c) = 0; \\ m_2: m_2(a) &= 0, m_2(b) = 0.01, m_2(c) = 0.99. \end{aligned}$$

由式(1)~(2)计算得 $K = 0.9999$, $m(a) = m(c) = 0$, $m(b) = 1$. 虽然这 2 条证据对目标 b 的基本概率分配值都很低,但是在进行 D-S 合成后却对目标 b 的基本概率分配值为 1,完全肯定了目标 b ,此结果显然是有悖直觉的,这个问题最早是由 L. A. Zadeh^[10] 提出的,称之为 Zadeh 悖论.

1.2.2 一票否决问题

例 2 识别框架 $\Theta = \{a, b, c\}$ 2 条证据 m_1 和 m_2 对该识别框架的各个目标基本概率分配分别为

$$\begin{aligned} m_1: m_1(a) &= 0.8, m_1(b) = 0.1, m_1(c) = 0.1; \\ m_2: m_2(a) &= 0, m_2(b) = 0.2, m_2(c) = 0.8; \\ m_3: m_3(a) &= 0.7, m_3(b) = 0.2, m_3(c) = 0.1. \end{aligned}$$

由式(1)~(2)计算得 $K = 0.9880$, $m(a) = 0$, $m(b) = 0.33$, $m(c) = 0.67$. 从合成结果来看,虽然很多证据对目标 a 的支持度都较高,但是只要出现 1 条证据的 $m(a) = 0$,合成的结果就是 $m(a) = 0$,即对目标 a 始终持否定态度,出现了一种“一票否决”的现象.

1.3 已有的几种解决冲突证据合成规则的修正方法

1.3.1 R. R. Yager 的合成规则 R. R. Yager^[11] 认为冲突的信息是无用的信息,基于此提出了对 D-S 合成规则的改进,改进公式为

$$\begin{cases} m_Y(\emptyset) = 0, \\ m_Y(A) = \sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j), \\ m_Y(\Theta) = \sum_{A_i \cap B_j = \Theta} m_1(A_i) m_2(B_j) + K, \end{cases}$$

其中 Θ 为其他的未知命题, K 为冲突因子.该方法的最大的优势是解决了 D-S 合成规则可能带来的悖论.当冲突因子 K 不是很大时,该方法合成的效果具有一定的价值;不过该方法同样也存在的问题,当冲突因子 K 接近于 1 时,冲突的部分会分配给未知命题,从而导致焦点获得的概率很小而无法做出最终的决策.当证据较多时,也可能产生“一票否决”现象.

1.3.2 C. K. Murphy 的合成规则 C. K. Murphy^[14] 则在平均信度优先情况下提出了一种对证据冲突进行合理简便处理的改进方法,改进公式为

$$m_M(A) = (m_1(A) + m_2(A) + \dots + m_n(A)) / n,$$

$\forall A \subseteq \Theta$.

1.3.3 孙全等的合成规则 孙全等^[12]进一步改进了 R. R. Yager^[11]的方法,改进公式为

$$\begin{cases} m_s(\emptyset) = 0, \\ m_s(A) = p(A) + K\varepsilon q(A) \quad A \neq \emptyset, \Theta, \\ m_s(\Theta) = p(\Theta) + K\varepsilon q(\Theta) + K(1 - \varepsilon), \end{cases}$$

其中 $\varepsilon = e^{-\tau}$, $\tau = \sum_{i < j} k_{ij} / (n(n-1)/2)$, $k_{ij} =$

$$\sum_{A_i \cap A_j = \emptyset} m_i(A_i) m_j(A_j) p(A) = \sum_{\bigcap_{i=1}^n A_i = A} \prod_{i=1}^n m_i(A_i) q(A) =$$

$$\sum_{i=1}^n m_i(A) / n.$$

孙全等^[12]提出了一种加权求和的方法,同时引入证据之间可信度的概念,将冲突的证据根据可信度分成2类,该方法提升了融合结果的可靠性.然而该方法在需要快速识别并做出决策的领域中应用的效果是有待提高的.

1.3.4 邓勇等的合成规则 邓勇等^[15]的合成规则是在 C. K. Murphy 的合成规则基础上考虑证据之间的联系,引入了一个度量证据与证据之间的距离的 Jousseleme 函数^[16],并通过计算得到证据之间的支持度 $s(m_i)$,然后通过归一化处理得到证据 m_i 的可信度 $r(m_i)$,并将它作为证据 m_i 的权重,最后在对证据进行加权平均后再利用 D-S 合成规则进行证据的融合,其改进部分的公式为

$$m_D(A) = \sum_{i=1}^n r(m_i) m_i(A),$$

其中 $r(m_i) = s(m_i) / \sum_{j=1}^n s(m_j)$.

1.3.5 刘康等的合成规则 刘康等^[17]认为数据的融合不仅要遵循算法的客观规则,而且要考虑决策者的主观判断.基于此,提出一种基于本征向量和 Jousseleme 距离^[18]的权重分配方法,该方法有效地提高了高度冲突证据融合的准确度和可靠性.

2 基于双重中位数绝对偏差(MAD)的证据加权方法

2.1 基于MAD的异常证据检测方法

2.1.1 基本概念 双重中位数绝对偏差被简称为 MAD^[19],该统计方法近几年来在异常值检测领域^[20]中备受青睐.下面定义了这一方法的基本概念.

设有一列数据: x_1, x_2, \dots, x_n , $M_d(x_j)$ ($j = 1, 2, \dots, n$) 为这列数据的中位数,用 M 表示,其表达

式为

$$M = M_d(x_j) \quad (j = 1, 2, \dots, n). \quad (3)$$

MAD 算法是在中位数基础上发展而来的^[21-22],称 M_σ 为双重中位数绝对偏差值,其表达式为

$$M_\sigma = M_d(|x_j - M|) \quad (j = 1, 2, \dots, n). \quad (4)$$

为了对数据中的异常值进行检测,给出判定范围 S ,其表达式为

$$S = [M - \varepsilon M_\sigma, M + \varepsilon M_\sigma], \quad (5)$$

其中 $\varepsilon = \sqrt{3}e^M$. 若 $x_i \in S$,则认为 x_i 为正常数据;若 $x_i \notin S$ 时,则认为 x_i 为异常数据.

2.1.2 基于MAD的异常证据检测方法 本文将 MAD 在数理统计中对数据检测的方法引入证据推理中.下面首先给出异常焦元和异常证据的定义.

定义2 设 $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ 为有限识别框架, m_1, m_2, \dots, m_n 为 Θ 上的基本概率分配函数,若 $\forall A_j \subseteq \Theta$, 有 $m_i(A_j) > 0$ 且 $m_i(A_j) \notin S$ ($i = 1, 2, \dots, n$), 则称 A_j 为证据 m_i 的异常焦元(abnormal focal element).

定义3 设 $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ 为有限识别框架, m_1, m_2, \dots, m_n 为 Θ 上的基本概率分配函数,若存在异常焦元 $A_j \subseteq \Theta$, 且 $A_j \in C(m_i)$ ($i = 1, 2, \dots, n$), 则称证据 m_i 为异常证据(abnormal evidence). 即若在1条证据中存在异常焦元,则该证据一定是异常证据.

2.1.3 随机模拟实验检测该方法的有效性 为了检测该方法的有效性,下面做3组随机模拟实验.

根据式(3)~(5)可以计算出图1~图3在5条证据融合时的各个焦元的正常范围(见表1).从随机产生的数据和通过式(3)~(5)计算出来的 S , BPA 值不属于 S 的证据被认定为异常证据也是比较符合直觉的.下面给出检测异常证据的一般步骤与3个具体的实例.

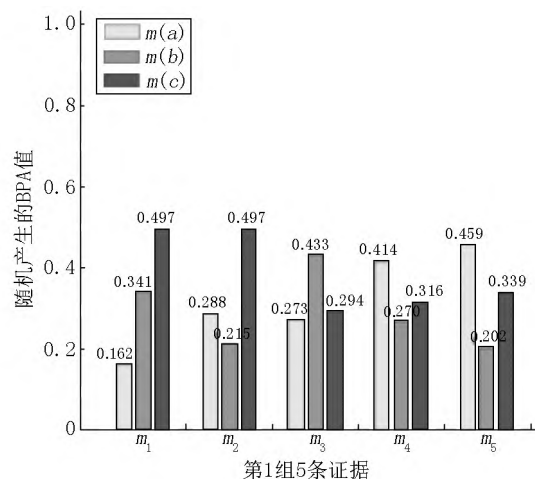


图1 随机产生第1组5条证据的BPA值

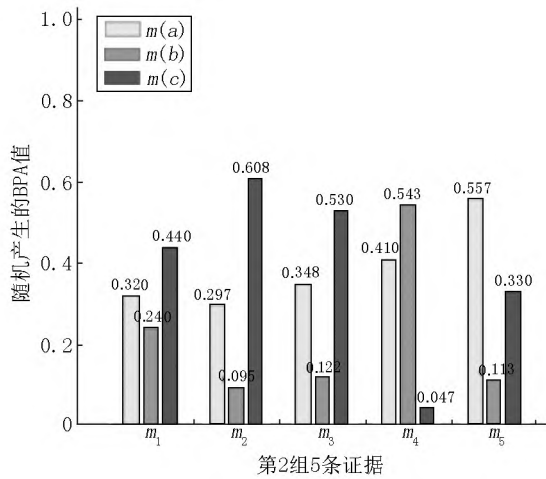


图2 随机产生第2组5条证据的BPA值

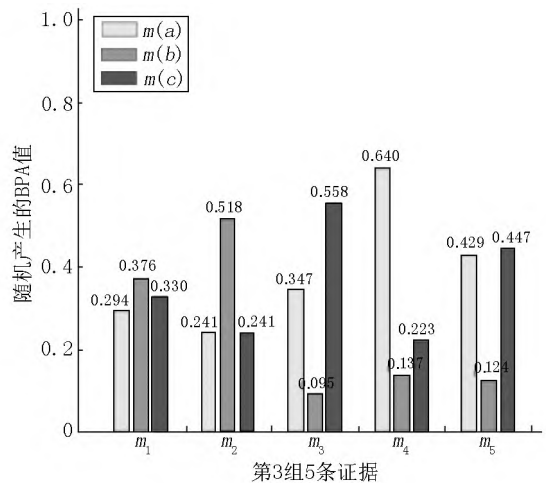


图3 随机产生第3组5条证据的BPA值

表1 在5条证据融合时各个焦元的正常范围

焦元	S		
	图1	图2	图3
a	[0.00 0.58]	[0.12 0.42]	[0.23 0.45]
b	[0.22 0.47]	[0.07 0.17]	[0.14 0.74]
c	[0.15 0.55]	[0.05 0.22]	[0.07 0.59]

2.1.4 检测异常证据的一般步骤

Step 1 通过式(3)~(4)可以计算出每个焦元的双重中位数绝对偏差 $M_\sigma(m(A_j))$ ($j = 1, 2, \dots, l$) 其中 $M_\sigma(m(A_j))$ 表示在 n 个证据源中同一焦元的 M_σ 值.

Step 2 通过式(5)计算出焦元 A_j 的判定范围 $S(A_j)$ ($j = 1, 2, \dots, l$).

Step 3 判断各个焦元是不是异常焦元. 若 $m_i(A_j) \in S(A_j)$ ($i = 1, 2, \dots, n$) 则 A_j 为证据 m_i 的正常焦元; 若 $m_i(A_j) \notin S(A_j)$ ($i = 1, 2, \dots, n$) 则 A_j 为证据 m_i 的异常焦元.

Step 4 判断各个证据是不是异常证据. 若 $\exists A_j \subseteq \Theta$ 使得 $m_i(A_j) \notin S(A_j)$ ($i = 1, 2, \dots, n$) 且

$A_j \in C(m_i)$ ($i = 1, 2, \dots, n$) 则称证据 m_i 为异常证据; 若 $\forall A_j \subseteq \Theta$, 使得 $m_i(A_j) \in S(A_j)$ ($i = 1, 2, \dots, n$) 且 $A_j \in C(m_i)$ ($i = 1, 2, \dots, n$) 则称证据 m_i 为正常证据.

在证据源的检测中,对于含有异常焦元的异常证据, MAD 算法对证据源的数量并不敏感,即使在较少的证据源($n \geq 3$)中也依然可行,且该算法检测出的异常证据是符合直觉的.下面通过3个例子检测的数据说明这个算法的合理性.

例3 设在1个盒子中装有1个球,已知球的颜色为绿色、红色、黄色之一,令 a, b, c 分别表示红色、黄色、绿色,则可以给出它的识别框架为 $\Theta = \{a, b, c\}$,有2条证据 m_1 和 m_2 对 Θ 的各个目标的基本概率分配为

$$m_1: m_1(a) = 0.2, m_1(b) = 0, m_1(c) = 0.8;$$

$$m_2: m_2(a) = 0.8, m_2(b) = 0, m_2(c) = 0.2.$$

通过 MAD 算法检测异常焦元.表2给出了在 m_1, m_2 2条证据合成时各个焦元的正常范围.

表2 在 m_1, m_2 证据合成时各个焦元的正常范围

焦元	a	b	c
S	[0.1, 0.995]	0	[0.1, 0.995]

例4 在例3的基础上加1条证据 m_3 , 其对各个目标的基本概率分配为

$$m_3: m_3(a) = 0.2, m_3(b) = 0, m_3(c) = 0.8.$$

通过 MAD 算法检测异常焦元,在 m_1, m_2, m_3 3条证据合成时各个焦元的正常范围与例3一样.表3给出了在 m_1, m_2, m_3 3条证据合成时各个焦元的正常范围.

表3 在 m_1, m_2, m_3 证据合成时各个焦元的正常范围

焦元	a	b	c
S	0.2	0	0.8

例5 在例4的基础上加1条证据 m_4 , 其对各个目标的基本概率分配为

$$m_4: m_4(a) = 0.4, m_4(b) = 0, m_4(c) = 0.6.$$

通过 MAD 算法检测异常焦元,在 m_1, m_2, m_3, m_4 4条证据合成时各个焦元的正常范围与例4一样.表4给出了在 m_1, m_2, m_3, m_4 4条证据合成时各个焦元的正常范围.

表4 在 m_1, m_2, m_3, m_4 证据合成时各个焦元的正常范围

焦元	a	b	c
S	[0.109 1, 0.490 9]	0	[0.415 2, 0.984 8]

从上述3个例子可知: MAD 算法检测的焦元的正常范围会随着证据的加入而发生改变,具有自适应性.通过本文给出的算法可以得出:当证据源 $n \leq 2$ 时,无法检测出异常证据.这是因为在仅有1条或2

条证据时证据的各个焦元的基本概率分配值一定在正常范围内。

该算法当 $n \geq 3$ 时就可以检测出异常证据,即该方法更适用于证据数量多的情形,而这些检测出来的异常证据与人们的直觉相符。

2.2 基于 MAD 检测证据加权组合的新方法

如前所述,D-S 合成规则被广泛地应用于证据的融合,且具有一些良好的性质,然而在面对高度冲突的证据源时该方法对数据的融合会产生一些悖论。为了避免这种现象发生,C. K. Murphy^[14] 通过提出一种证据的算术平均方法来对证据进行预处理,此方法在一定程度上改进了 D-S 合成规则面对高度冲突证据的不足,但此方法没有充分考虑各个证据之间的联系。为了解决这个问题,邓勇等^[15] 提出了基于 Jusselme 距离的证据加权平均的方法,该方法继承了 C. K. Murphy 方法的所有优点,并且考虑了证据之间的联系,在一定程度上有较好的合成效果,但该方法忽略了证据本身对数据的影响。本文基于这 3 种方法提出了一个新的证据加权组合方法。

本文方法的核心思想是:将收集到的多源证据通过 MAD 算法检测是否存在异常证据,若不存在异常证据则就直接使用证据加权组合方法得出最终的合成结果;若含有异常证据,则先将异常证据用这组证据的平均值证据进行修正,再将修改后的证据使用 D-S 合成规则、C. K. Murphy 的合成规则、邓勇等的合成规则 3 种方法进行加权组合,其改进公式为

$$m(A) = \beta_1 m_1(A) + \beta_2 m_2(A) + \beta_3 m_3(A), \quad (6)$$

其中 $\sum_{i=1}^3 \beta_i = 1$, m_1 为 D-S 合成证据, m_2 为 C. K. Murphy 的合成证据, m_3 为邓勇等的合成证据。现在的核心问题就是如何确定证据 m_i 的权重 β_i ,下面将介绍该权重的确定方法。为了度量证据之间的相关性程度,引入 Jusselme 距离函数,其公式为

$$d_{\text{BPA}}(m_k, m_l) = \sqrt{(m_k - m_l)^T D (m_k - m_l) / 2}, \quad (7)$$

其中 m_k, m_l 为证据的向量形式, D 为 $2^N \times 2^N$ 矩阵,矩阵 D 中的元素 $d_{ij} = |A_i \cap A_j| / |A_i \cup A_j|, \forall A_i, A_j \subseteq \Theta, N$ 为 Θ 中元素的个数。2 条证据的距离越大说明这 2 条证据之间的相关性程度越小,故本文在 $\alpha_i \neq 0 (i = 1, 2, 3)$ 时使用

$$1/\alpha_i (i = 1, 2, 3) \quad (8)$$

来表示各个证据之间的相关性系数。 $\alpha_i = 0$ 表示这组证据都完全相同,此时就使用 D-S 合成规则来计

算而无须使用本文的证据加权组合方法。这里

$$\alpha_1 = \sum_{i=1}^n d_{\text{BPA}}(m_1, m_i), \alpha_2 = \sum_{i=1}^n d_{\text{BPA}}(m_2, m_i), \alpha_3 =$$

$\sum_{i=1}^n d_{\text{BPA}}(m_3, m_i)$ 分别表示修正后的证据采用 D-S 合成规则、C. K. Murphy 的合成规则、邓勇等的合成规则融合后的结果与修正后的证据之间的 Jusselme 距离,其中 n 为独立证据的个数。在利用 Jusselme 距离函数求出 m_i 的相关性系数 $1/\alpha_i$ 后,将其归一化就可以获得最后的证据 m_i 的权重 β_i ,其表达式为

$$\beta_i = (1/\alpha_i) / (1/\alpha_1 + 1/\alpha_2 + 1/\alpha_3) (i = 1, 2, 3, \alpha_i \neq 0), \quad (9)$$

这样就确定了权重分配的系数,从而可以通过式(6)计算出最终的结果。为了验证该方法的有效性,下面通过随机模拟实验和仿真实验来说明本文方法可以较好地解决证据源高度冲突融合问题。

2.3 本文方法流程图

图 4 为基于 MAD 检测的证据加权组合方法流程图。

2.4 对本文算法时间复杂度进行分析

2.4.1 对本文的检测算法进行分析 本文通过 MAD 算法计算出 M_σ 和 S_1) 先将对同一焦元的赋值大小进行排序,其次数为 n^2 ; 2) 进行 1 次取中位数操作,共进行 n 次。后面的步骤的执行次数均不超过 $O(n^2)$ 。故该算法的时间复杂度为 $O(n^2)$ 。

2.4.2 对本文加权组合新方法进行分析 本文需要通过 Jusselme 距离函数计算出 $\alpha_1, \alpha_2, \alpha_3$ 的值。1) 先计算出其他 3 种合成方法的结果; 2) 计算出权重 $\alpha_1, \alpha_2, \alpha_3$ 的值。前 2 步算法的时间复杂度为 $O(n^2)$ 。后面的步骤的执行次数均不超过 $O(n^2)$ 。故该算法的时间复杂度为 $O(n^2)$ 。

2.5 随机模拟实验检测本文合成方法的有效性

为了检测本文合成方法的是否比已有的方法有效,本文用合成后的最终结果与原始证据之间的 Jusselme 距离 d 来评价本文方法的优劣。合成结果的评价标准为: d 越小说明合成结果与原始证据越接近,即该方法合成结果变形程度越小,合成的效果越好; d 越大说明合成结果与原始证据越远,即该方法合成结果变形程度越大,合成的效果越差。现使用 Matlab 随机模拟 100 次,每次随机生成 100 条基本概率分配函数,计算出合成结果与原始证据之间的距离并绘制出图 5。

从图 5 可知:本文方法计算出的结果与原始证据之间的距离较小,这说明本文方法计算的最终结

果变形程度小,合成的效果较好.其他3种方法计算出来的结果与原始证据之间的距离较大,特别是经典的 D-S 合成规则计算出的最终结果变形程度较大

且不稳定.这也解释了近几十年来许多学者^[23-24]对经典的 D-S 合成规则要改进的原因.

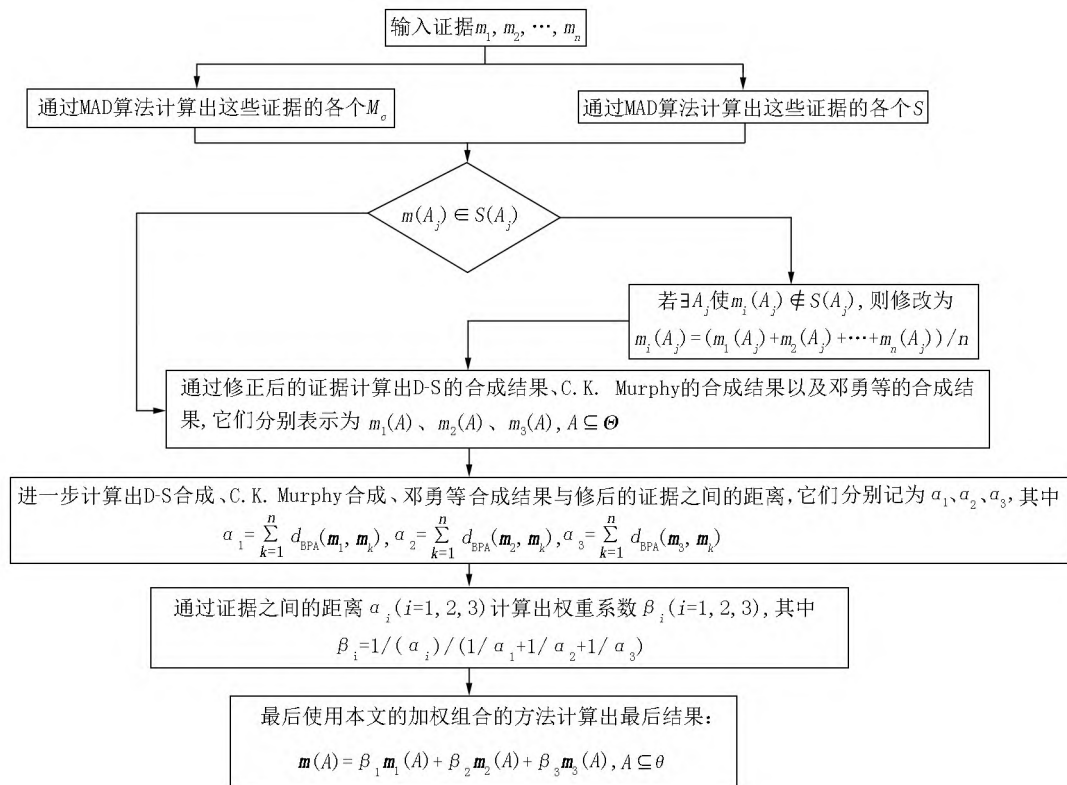


图4 本文方法流程图

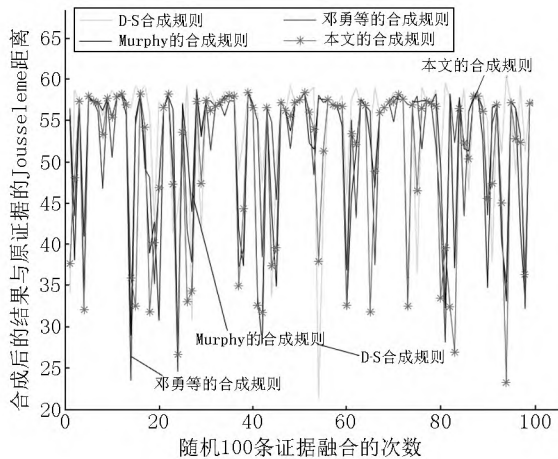


图5 随机模拟实验图

2.6 使用本文方法对例1、例2进行分析

对例1使用本文的加权组合方法进行证据的合成.由于本文的MAD算法无法检测出只有2条证据的异常证据,所以本文的检测方法对例1是失效的,故直接使用本文提出的基于Josselme距离的证据权重组合方法进行数据的融合.通过式(7)~(9)可以计算出 $\beta_1 = 0.1934$, $\beta_2 = 0.4033$, $\beta_3 = 0.4033$;然后通过式(6)可以计算出 $m(a) = 0.40$, $m(b) = 0.20$, $m(c) = 0.40$.其结果与D-S合成规则、C. K. Murphy的合成规则、邓勇等的合成规则的融合结果进行比较(见表5).

表5 例1的4种合成规则结果比较

融合方法	2条证据 m_1, m_2 的合成结果		
D-S 合成规则	$m(a) = 0$	$m(b) = 1.0$	$m(c) = 0$
C. K. Murphy 的合成规则	$m(a) = 0.5$	$m(b) = 0$	$m(b) = 0.5$
邓勇等的合成规则	$m(a) = 0.5$	$m(b) = 0$	$m(c) = 0.5$
本文的合成规则	$m(a) = 0.4$	$m(b) = 0.2$	$m(c) = 0.4$

从表5中的数据可以得出:经典的D-S合成规则在面对冲突性数据融合时产生与直觉不符的结果,而C. K. Murphy的合成规则、邓勇等的合成规

则的融合结果对目标a与目标c的支持度均为0.5,对目标b持否定态度,存在一定的缺陷.从本文的合成规则结果来看,该方法对这3种方法都考虑了,得

到的结果没有绝对地否定目标 b ,对目标 a 和目标 c 的支持度都是一样的(各占 0.4) ,这比较合理.

对例 2 使用本文的加权组合方法进行证据的合成. 由于在使用本文的 MAD 算法时 ,当只有 2 条证据 m_1 、 m_2 时无法检测出异常证据 ,所以计算方法与

例 1 的一样 ,最终融合结果与 D-S 合成规则、C. K. Murphy 的合成规则、邓勇等的合成规则的融合结果进行比较(见表 6) . 当加入第 3 条证据 m_3 时 ,通过 MAD 算法检测异常焦元 ,表 7 为在 m_1 、 m_2 、 m_3 3 条证据合成时各个焦元的正常范围.

表 6 例 2 的 4 种合成规则结果比较

融合方法	2 条证据 m_1 、 m_2 的合成结果		
D-S 合成规则	$m(a) = 0$	$m(b) = 0.20$	$m(c) = 0.80$
C. K. Murphy 的合成规则	$m(a) = 0.42$	$m(b) = 0.06$	$m(b) = 0.52$
邓勇等的合成规则	$m(a) = 0.42$	$m(b) = 0.06$	$m(c) = 0.52$
本文的合成规则	$m(a) = 0.28$	$m(b) = 0.11$	$m(c) = 0.61$

表 7 在 m_1 、 m_2 、 m_3 3 条证据合成时各个焦元的正常范围

焦元	a	b	c
S	[0.329 4 ,1.070 7]	0.2	0.1

由表 7 中的焦元 a 的正常范围可知证据 m_2 需要修改 ,由焦元 b 的正常范围可知证据 m_1 需要修改 ,故将证据 m_1 、 m_2 使用这组证据的平均值进行修正 ,修改后的结果为

$m'_1:m'_1(a) = 0.50 \quad m'_1(b) = 0.17 \quad m'_1(c) = 0.33;$

$m'_2:m'_2(a) = 0.50 \quad m'_2(b) = 0.17 \quad m'_2(c) = 0.33;$
 $m'_3 = m_3:m_3(a) = 0.70 \quad m_3(b) = 0.20 \quad m_3(c) = 0.10.$

将修改后的证据使用本文提出的基于 Jousselme 距离的证据权重组合方法进行数据的融合. 通过式 (7) ~ (9) 可以计算出 $\beta_1 = 0.347 8$, $\beta_2 = 0.332 4$ $\beta_3 = 0.319 8$. 再使用本文的证据权重组合方法与证据修正后的 D-S 合成规则、C. K. Murphy 的合成规则、邓勇等的合成规则的融合结果进行比较(见表 8) .

表 8 例 2 的修正后 4 种合成规则结果比较

融合规则	2 条证据 m_1 、 m_2 的合成结果		
D-S 合成规则	$m(a) = 0.913 0$	$m(b) = 0.030 2$	$m(c) = 0.056 8$
C. K. Murphy 的合成规则	$m(a) = 0.896 2$	$m(b) = 0.028 2$	$m(b) = 0.075 6$
邓勇等的合成规则	$m(a) = 0.882 5$	$m(b) = 0.029 2$	$m(c) = 0.088 3$
本文的合成规则	$m(a) = 0.897 7$	$m(b) = 0.029 2$	$m(c) = 0.073 1$

从表 8 的数据可以得出: 在使用本文的 MAD 算法检测出异常证据并修正后 ,再使用 D-S 合成规则、C. K. Murphy 的合成规则、邓勇等的合成规则、本文的合成规则进行融合 ,得到的结果均比较合理 ,这说明本文的 MAD 算法是有效的. 由于在 2.5 节的随机模拟实验中考考虑合成结果的稳健性 ,所以使用本文的加权组合方法来对修正后的证据源进行合成. 从合成结果来看 ,本文方法可以较好地解决 Zadeh 悖论问题与“一票否决”问题.

下面通过具体仿真实验来说明本文方法在面對高度冲突证据的融合时也能够快速高效地做出决策.

3 仿真实验

为了验证本文加权组合的新方法的效果 ,本文将采用文献 [17] 的完整算例进行仿真实验来验证本文算法的有效性.

例 6 设有 3 个目标的识别框架: $\Theta = \{A, B, C\}$,收集的 5 条证据对目标 A 、 B 、 C 的基本概率分配值如图 6 所示.

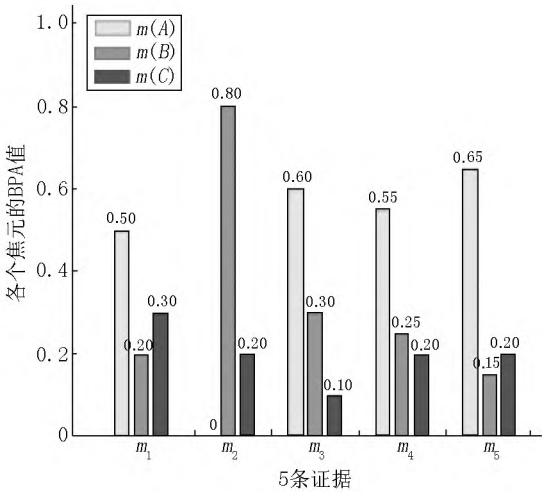


图 6 5 条证据对目标 A、B、C 的 BPA 值

从这 5 条证据来看 m_1 、 m_3 、 m_4 、 m_5 都认为目标 A 的可能性较大 ,而证据 m_2 却否定了目标 A 根据直觉可以认为证据 m_2 的情况应该是因受到某些不确定因素影响而导致的. 根据本文的合成规则 ,首先通过 MAD 算法找出异常焦元 ,然后对含有异常焦元的异常证据进行修改 ,最后在对修改的证据进行证据

的加权平均后使用 D-S 合成规则融合. 表 9 是在证据 m_3 、 m_4 、 m_5 逐步加入后各个焦元的正常范围.

从表 9 中各个焦元的正常范围可以得到: 当证据 m_1 、 m_2 融合时, 无法检测出异常证据; 当加入证据 m_3 时, 由各个焦元的正常范围可知证据 m_2 需要修改; 当再加入证据 m_4 时, 从表 9 中数据可知证据 m_2 需要修改; 当 5 条证据都进行融合时, 由各个焦元的

正常范围可知证据 m_1 、 m_2 、 m_3 需要修改. 将修正后的证据使用本文的合成方法进行合成并与 7 种合成规则做比较(为了更好地体现本文方法的有效性, 文献[25]的合成方法的结果也列于表中), 其结果如表 10 所示.

由表 10 的数据可得 8 种合成规则对 A、B、C 支持度随着融合证据的加入的变化趋势(见图 7 ~ 图 9).

表 9 各焦元的正常范围

焦元	A	B	C
$S_{m_1 m_2}$	$[-0.306\ 0\ 0.806\ 0]$	$[-0.356\ 7\ 1.356\ 7]$	$[0.138\ 8\ 0.361\ 2]$
$S_{m_1 m_2 m_3}$	$[0.214\ 4\ 0.785\ 6]$	$[0.066\ 2\ 0.533\ 8]$	$[-0.011\ 6\ 0.411\ 6]$
$S_{m_1 m_2 m_3 m_4}$	$[0.378\ 6\ 0.671\ 4]$	$[0.161\ 0\ 0.389\ 0]$	$[0.094\ 2\ 0.305\ 8]$
$S_{m_1 m_2 m_3 m_4 m_5}$	$[0.399\ 9\ 0.700\ 1]$	$[0.138\ 8\ 0.361\ 2]$	0.2

表 10 8 种合成规则结果比较

融合方法	$m_1\ m_2$	$m_1\ m_2\ m_3$	$m_1\ m_2\ m_3\ m_4$	$m_1\ m_2\ m_3\ m_4\ m_5$
D-S 合成规则	$m(A) = 0.000\ 0$	$m(A) = 0.000\ 0$	$m(A) = 0.000\ 0$	$m(A) = 0.000\ 0$
	$m(B) = 0.727\ 3$	$m(B) = 0.888\ 9$	$m(B) = 0.909\ 1$	$m(B) = 0.882\ 4$
	$m(C) = 0.272\ 7$	$m(C) = 0.111\ 1$	$m(C) = 0.090\ 9$	$m(C) = 0.117\ 6$
R. R. Yager 的合成规则	$m(A) = 0.000\ 0$	$m(A) = 0.000\ 0$	$m(A) = 0.000\ 0$	$m(A) = 0.000\ 0$
	$m(B) = 0.160\ 0$	$m(B) = 0.048\ 0$	$m(B) = 0.012\ 0$	$m(B) = 0.001\ 8$
	$m(C) = 0.060\ 0$	$m(C) = 0.006\ 0$	$m(C) = 0.001\ 2$	$m(C) = 0.000\ 2$
	$m(\Theta) = 0.780\ 0$	$m(\Theta) = 0.946\ 0$	$m(\Theta) = 0.986\ 8$	$m(\Theta) = 0.998\ 0$
孙权等的合成规则	$m(A) = 0.089\ 4$	$m(A) = 0.170\ 5$	$m(A) = 0.206\ 2$	$m(A) = 0.236\ 9$
	$m(B) = 0.338\ 8$	$m(B) = 0.249\ 5$	$m(B) = 0.205\ 7$	$m(B) = 0.176\ 9$
	$m(C) = 0.149\ 4$	$m(C) = 0.099\ 0$	$m(C) = 0.101\ 2$	$m(C) = 0.103\ 2$
	$m(\Theta) = 0.422\ 4$	$m(\Theta) = 0.481\ 0$	$m(\Theta) = 0.486\ 9$	$m(\Theta) = 0.483\ 0$
C. K. Murphy 的合成规则	$m(A) = 0.166\ 7$	$m(A) = 0.355\ 5$	$m(A) = 0.545\ 3$	$m(A) = 0.809\ 0$
	$m(B) = 0.666\ 7$	$m(B) = 0.586\ 8$	$m(B) = 0.424\ 6$	$m(B) = 0.178\ 5$
	$m(C) = 0.166\ 6$	$m(C) = 0.057\ 7$	$m(C) = 0.030\ 1$	$m(C) = 0.012\ 5$
邓勇等的合成规则	$m(A) = 0.166\ 7$	$m(A) = 0.497\ 8$	$m(A) = 0.741\ 8$	$m(A) = 0.927\ 7$
	$m(B) = 0.666\ 7$	$m(B) = 0.443\ 4$	$m(B) = 0.231\ 2$	$m(B) = 0.063\ 3$
	$m(C) = 0.166\ 6$	$m(C) = 0.058\ 8$	$m(C) = 0.027\ 0$	$m(C) = 0.009\ 0$
刘康等的合成规则	$m(A) = 0.394\ 2$	$m(A) = 0.644\ 8$	$m(A) = 0.802\ 3$	$m(A) = 0.942\ 2$
	$m(B) = 0.294\ 9$	$m(B) = 0.256\ 6$	$m(B) = 0.150\ 9$	$m(B) = 0.040\ 9$
	$m(C) = 0.285\ 6$	$m(C) = 0.096\ 8$	$m(C) = 0.046\ 8$	$m(C) = 0.016\ 9$
	$m(\Theta) = 0.025\ 4$	$m(\Theta) = 0.001\ 8$	$m(\Theta) = 0.000\ 1$	
Zhang Weiquan 等的合成规则	$m(A) = 0.029\ 6$	$m(A) = 0.380\ 8$	$m(A) = 0.616\ 2$	$m(A) = 0.856\ 9$
	$m(B) = 0.800\ 0$	$m(B) = 0.452\ 3$	$m(B) = 0.344\ 6$	$m(B) = 0.125\ 2$
	$m(C) = 0.170\ 4$	$m(C) = 0.166\ 9$	$m(C) = 0.039\ 2$	$m(C) = 0.017\ 5$
本文的合成规则	$m(A) = 0.112\ 3$	$m(A) = 0.766\ 2$	$m(A) = 0.905\ 6$	$m(A) = 0.946\ 5$
	$m(B) = 0.686\ 5$	$m(B) = 0.186\ 9$	$m(B) = 0.078\ 3$	$m(B) = 0.045\ 4$
	$m(C) = 0.201\ 2$	$m(C) = 0.046\ 9$	$m(C) = 0.016\ 1$	$m(C) = 0.008\ 1$

根据表 10 和图 7 信息可以看出: 对于 D-S 合成规则, 尽管在 5 个证据源中有 4 个证据源都支持目标 A, 但是由于冲突证据 $m_2(A) = 0$, 所以融合的结果 $m(A)$ 始终是 0, 出现了“一票否决”的结果, 这明显是不合理的结果. 在面对高度冲突的证据时, R. R. Yager、孙权等的合成规则是将大部分支持度

赋值给未知 Θ , 这也一样没有得到合理的融合结果. 当只有 3 条证据时, C. K. Murphy 的合成规则、邓勇等的合成规则并不能较好地做出决策, 这是因为这 2 种方法对目标 A 与 B 的支持度都接近 0.5; 只有刘康等的合成规则与本文的合成规则对目标 A 的支持度均高于 0.6 且对其他目标的支持度都很低, 并且

本文的合成规则对目标 A 的支持度高达 0.905 6。随着更多的证据源进行融合, 本文的合成规则对目标 A 的支持度越来越合理, 当有 5 条证据源融合时, 本文的合成规则对目标 A 的支持度高达 0.946 5。

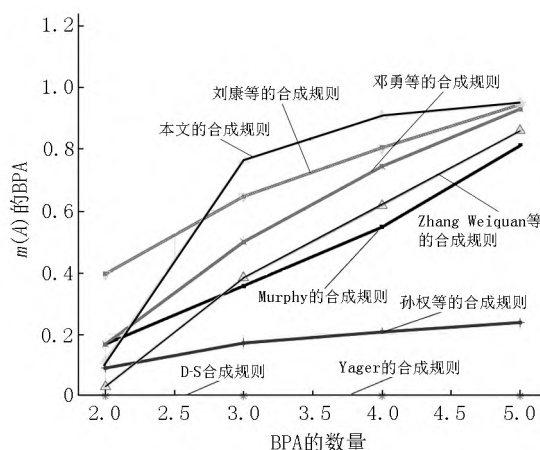


图7 8种合成规则对目标 A 的 BPA 值

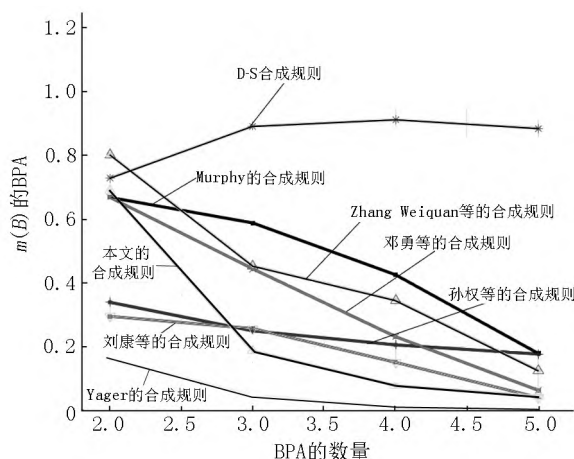


图8 8种合成规则对目标 B 的 BPA 值

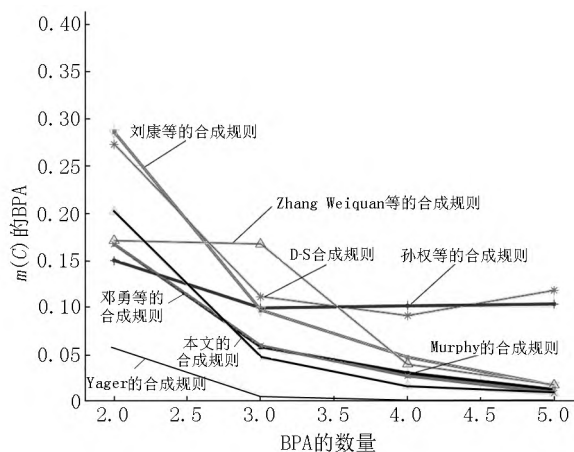


图9 8种合成规则对目标 C 的 BPA 值

本文的合成规则是建立在文献 [1, 14-15] 的基础上的, 它的过程是: 先将接收到的信息源进行预处理, 再通过 MAD 算法检测出异常证据源, 最后将异常证据源用这组证据的平均值进行修正后通过证据加权组合得到最终结果。

由图 7 可以看出: 本文的合成规则可以较快地对目标 A 做出决策; 然而当只有 2 条证据源进行融合时, 该方法无法检测出异常证据; 随着证据的加入, 对错误目标 B 、 C 的支持度能较快地下降。与其他方法比较, 本文方法可以快速高效地做出决策。

4 结束语

针对经典的 D-S 合成规则在面对较高冲突证据融合时会产生有悖常理的结果, 本文提出了一种基于双重中位数绝对偏差的检测方法和证据加权组合修正方法, 该检测方法是在数理统计中经常使用的对数据异常值检测的方法, 本文将其对在证据理论中的异常证据进行检测, 从算例的数据可以得出该检测方法有效。仿真实验数据表明: 本文改进的方法对证据融合具有较快的收敛性, 比在实验中的其他 8 种方法效果更好, 这在一定程度上解决了证据源的高度冲突融合问题。

在对未来的研究工作中, 笔者将重点研究合理分配算法的证据权重、提高算法的计算效率、降低算法的时间复杂度等。

5 参考文献

- [1] DEMPSTER A P. Upper and lower probabilities induced by a multivalued mapping [J]. Annals of Mathematical Statistics, 1967, 38(2): 325-339.
- [2] SHAFER G. A mathematical theory of evidence [M]. Princeton: Princeton University Press, 1976.
- [3] 赵玲. 多传感器信息融合技术及其应用 [J]. 红外, 2021, 42(1): 21-26.
- [4] LEYS C, LEY C, KLEIN O, et al. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median [J]. Journal of Experimental Social Psychology, 2013, 49(4): 764-766.
- [5] FAN Chengli, SONG Yafei, LEI Lei, et al. Evidence reasoning for temporal uncertain information based on relative reliability evaluation [J]. Expert Systems with Applications, 2018, 113: 264-276.
- [6] LI Junwei, XIE Baolin, JIN Yong, et al. Weighted conflict evidence combination method based on Hellinger distance and the belief entropy [J]. IEEE Access, 2020, 8: 225507-225521.
- [7] XIONG Leihui, SU Xiaoyan, QIAN Hong. Conflicting evidence combination from the perspective of networks [J]. Information Sciences, 2021, 580(2): 408-418.

- [8] 周国宪, 伍星, 刘韬. 基于多传感器的神经网络和 D-S 证据理论在故障诊断中的应用 [J]. 测试技术学报, 2017, 31(4): 290-297.
- [9] 姬晓飞, 石宇辰, 王昱, 等. D-S 理论多分类器融合的光学遥感图像多目标识别 [J]. 电子测量与仪器学报, 2020, 34(5): 127-132.
- [10] ZADEH L A. The concept of a linguistic variable and its application to approximate reasoning: Part 1 [J]. Information Sciences, 1975, 8(3): 199-249.
- [11] YAGER R R. On the dempster-shafer framework and new combination rules [J]. Information Sciences, 1987, 41(2): 93-137.
- [12] 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.
- [13] 吴根秀. 冲突证据组合方法 [J]. 计算机工程, 2005, 31(9): 151-154.
- [14] MURPHY C K. Combining belief functions when evidence conflicts [J]. Decision Support Systems, 2000, 29(1): 1-9.
- [15] 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.
- [16] JOUSSELME A L, GRENIER D, BOSSÉ É. A new distance between two bodies of evidence [J]. Information Fusion, 2001, 2(2): 91-101.
- [17] 刘康, 何明浩, 韩俊, 等. 基于本征向量和 Jusselme 距离的高冲突证据融合方法 [J]. 系统工程与电子技术, 2022, 44(7): 2175-2180.
- [18] 严志军, 陶洋. 基于证据距离和不确定度的冲突数据融合算法 [J]. 计算机工程与应用, 2018, 54(10): 54-58, 104.
- [19] 余玉萍. 基于中位数的双 MAD 的离群值检测方法 [J]. 廊坊师范学院学报(自然科学版), 2016, 16(2): 25-27.
- [20] 龚循强, 张方泽, 鲁铁定, 等. 基于中位数绝对偏差的异常训练样本探测方法 [J]. 激光与光电子学进展, 2020, 57(23): 173-178.
- [21] HAMPEL F R. The influence curve and its role in robust estimation [J]. Journal of the American Statistical Association, 1974, 69(346): 383-393.
- [22] HAMPEL F R, ROUSSEAU P J, RONCHETTI E M, et al. Robust statistics: the approach based on influence functions [J]. Journal of the Royal Statistical Society Series D (The Statistician), 1986, 35(5): 1-48.
- [23] LIU Zhunga, LIU Yu, DEZERT J, et al. Evidence combination based on credal belief redistribution for pattern classification [J]. IEEE Transactions on Fuzzy Systems, 2019, 28(4): 618-631.
- [24] LIU Jing, CHANG Chaowen, Zhang Yuchen, et al. A multi-source situation information fusion method based on dynamic evidence combination [J]. Mathematical Problems in Engineering, 2021, 2021: 8834917.
- [25] ZHANG Weiquan, DENG Yong. Combining conflicting evidence using the DEMATEL method [J]. Soft Computing, 2019, 23(17): 8207-8216.

The Improved Evidence Weighted Combination Method Based on Double Median Absolute Detection

YU Xin, WU Genxiu*, CAI Aoli

(College of Mathematics and Statistics, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

Abstract: Dempster-shafer evidence theory has been applicable in all walks of life of digital marketing in a large scale, but when it is applied to the integration of highly conflicting evidence sources, it often produces some counter-intuitive results and has certain limitations. In order to solve this problem, the improved method based on double median absolute deviation detection and a new improved method of evidence weighted combination evidence weighted combination is proposed. Firstly, the abnormal evidence is detected by MAD algorithm, then the average value of this group of evidence is used to correct the abnormal evidence, and the new evidence weighting combination is applied to weight average the evidence to get the final count. Through random simulation experiments and example experiments, and compared with several other classical evidence synthesis rules, the simulation results show that the synthesis rules in this paper can effectively solve the problem of highly conflicting evidence fusion.

Key words: Dempster-Shafer evidence theory; median absolute deviation; high conflict

(责任编辑: 曾剑锋)