

DOI: 10.7652/xjtuxb201806001

证据距离的选取对冲突证据组合影响的研究

周永庆¹, 韩德强¹, 杨艺^{2,3}

(1. 西安交通大学电子与信息工程学院, 710049, 西安; 2. 西安交通大学机械结构强度与振动国家重点实验室, 710049, 西安; 3. 西安交通大学航天航空学院, 710049, 西安)

摘要: 为了研究证据距离的选取对冲突证据组合的影响,提出一种将不同的证据距离度量方式应用在加权证据组合中的对比方法。首先采用不同的证据距离度量方式得到证据相似度,依据证据相似度求得证据可靠度,之后用证据可靠度对证据进行加权求和得到平均证据,最终对平均证据进行证据组合,对比在不同的证据距离度量方式下上述证据组合过程的合理性,得到适合加权证据组合中权重生成的距离度量方式。采用 4 个算例,对不同的证据距离度量方式下求得的证据可靠度、平均证据和证据组合结果进行分析比较,实验结果表明,与其他证据距离相比,利用该方法选出的证据距离能够更有效地体现证据之间的差异性,所求取的证据可靠度与实际情况更为贴合。

关键词: 冲突证据;加权证据组合;证据距离;证据可靠度

中图分类号: TP391 **文献标志码:** A **文章编号:** 0253-987X(2018)06-0001-08

A Research on the Influence of Evidence Distance Selection on Combination of Conflict Evidences

ZHOU Yongqing¹, HAN Deqiang¹, YANG Yi^{2,3}

(1. School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China;
2. State Key Laboratory for Strength and Vibration of Mechanical Structures, Xi'an Jiaotong University, Xi'an 710049, China; 3. School of Aerospace, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: A comparative method is proposed to study the influence of evidence distance selection on combination of conflict evidences and the method applies different evidence distance measurements to combination of weighted evidences. Firstly, different evidence distance measurements are used to get similarities of evidences. Then evidence credibilities are calculated using evidence similarities and the averages of evidences are calculated by summing up weighted evidences using the credibilities as weights. Finally, resulting averages are combined. The rationality of these combinations of evidences in different evidence distance measurements is compared to obtain a suitable weights generation in weighted evidence combination. Numerical examples are given to analyze and to compare evidence credibilities, averages of evidences and combination results of evidences. Experimental results and comparisons with other evidence distance measurements show that the evidence distance measurements selected by the proposed method reflect the difference between the evidences more effectively, and the calculated evidence credibilities are more consistent with the actual situation.

收稿日期: 2017-12-18。 作者简介: 周永庆(1993—),男,硕士生;韩德强(通信作者),男,教授,博士生导师。 基金项目: 国家自然科学基金资助项目(61573275,61671370);国家“973 计划”资助项目(2013CB329405);中国博士后科学基金资助项目(2016M592790);陕西省科技计划资助项目(2013KJXX-46);中央高校基本科研业务费专项资金资助项目(xj2016066)。

网络出版时间: 2018-03-15 网络出版地址: <http://kns.cnki.net/kcms/detail/61.1069.T.20180315.2057.016.html>

Keywords: conflict evidence; weighted evidence combination; evidence distance; evidence credibility

在 Dempster-Shafer (DS) 证据理论^[1]中, 独立证据可以依据证据组合规则进行证据组合。当证据之间存在冲突时, 证据组合后的结果往往与实际直觉相悖^[2], 针对这一情况, 出现了许多改进的证据组合规则^[3-8], 其中有一类是基于证据间距离的冲突证据组合规则^[3-5], 这类算法用证据之间的距离度量证据之间的相似性, 依据相似性得到证据的权值, 对证据进行加权融合。在此类方法中, 证据距离度量方式对证据组合有重要的影响。

随着证据理论的研究深入, 涌现了诸多证据之间距离的度量定义^[9], 如 Jousselme 距离^[10]、基于 Pignistic 概率的距离^[11]、基于模糊隶属度的距离^[12]和基于信任区间的距离^[13]等。在既有的基于证据间距离的冲突证据组合研究中, 上述距离定义方式均有所采用。邓勇等人依据证据间的 Jousselme 距离得到证据可靠度, 用证据可靠度对证据进行加权融合获取平均证据, 再将平均证据进行组合^[3]。肖建于等人则基于 Pignistic 概率距离得到证据的权重, 求得平均证据后再进行组合^[4]。刘志成等人则依据修正的 City Block 距离获取证据的权重^[5]。在这些方法中, 只选取了某一种证据间的距离度量方式, 没有对不同的证据距离度量方式进行对比。

为了探究不同的证据距离度量方式对冲突证据组合的影响, 从而选取适合加权证据组合权重生成的距离定义, 本文将各种距离度量方式应用于基于证据距离修正证据组合的方法中, 对比了在不同的距离度量方式下, 证据可靠度的大小和将平均证据组合后所得组合结果之间的差异, 分析了产生差异的原因。采用 4 个算例的实验结果表明, 在证据组合过程中, 与其他的证据距离相比, 采用本文方法选取的证据距离所求得的证据可靠度更符合实际, 体现证据差异更有效。

1 证据理论概述

设 Θ 为辨识框架, 若集函数 $m: 2^\Theta \rightarrow [0, 1]$ 满足

$$\left. \begin{aligned} m(\emptyset) &= 0 \\ \sum_{A \subseteq \Theta} m(A) &= 1 \end{aligned} \right\} \quad (1)$$

则称 m 为 Θ 上的 Mass 函数。Mass 函数实际上就是对各种假设的评价权值。其中, 使得 $m(A) > 0$ 的 A 称为焦元。

在辨识框架 Θ 上的某个命题 A 的信任函数 (belief function) 定义为

$$\text{bel}(A) = \sum_{B \subseteq A} m(B) \quad (2)$$

似真度函数 (plausibility function) 定义为

$$\text{pl}(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad (3)$$

区间 $[\text{bel}(A), \text{pl}(A)]$ 为信任区间, 用以表示对命题 (焦元) A 的不确定程度。

Dempster 组合规则也称证据组合公式, 其定义为: 设统一辨识框架上的两个独立证据, 其相应的 mass 函数分别为 m_1 和 m_2 , $\forall A \subseteq \Theta, A \neq \emptyset$, 则组合之后证据为

$$(m_1 \oplus m_2)(A) = \frac{1}{1 - K} \sum_{B \cap C = A} m_1(B) m_2(C) \quad (4)$$

式中: $K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$ 是冲突项。Dempster 组合规则满足交换律与结合律。

在将证据进行融合之后, 需要将融合结果的 mass 函数转化为概率以便于进行决策。Mass 函数转化为概率有多种方法, 一种经典的方法为 Smets 提出的 Pignistic 概率转换方法^[14]

$$P(A) = \sum_{B \in \Theta} \frac{|A \cap B|}{|B|} \frac{m(B)}{1 - m(\emptyset)} \quad (5)$$

2 冲突证据组合方法

2.1 证据冲突现象

当证据之间存在较大冲突的时候, 应用 Dempster 组合规则融合后得到的结果往往与直观结果相违背。以下为 Zadeh 所举的经典算例^[2]。

两名医生为同一个病人看病, 该病人可能得的病为脑膜炎 (M)、脑震荡 (C) 和脑肿瘤 (T), 即辨识框架为 $\Theta = \{M, C, T\}$ 。两名医生给出的各种病症的可能结果如下

$$m_1(\{M\}) = 0.99; m_1(\{T\}) = 0.01$$

$$m_2(\{C\}) = 0.99; m_2(\{T\}) = 0.01$$

将以上两个独立证据应用 Dempster 组合规则进行融合, 得到融合后的结果为 $m(\{T\}) = 1$, 此时的冲突系数 $K = 0.9999$ 。

在这个例子中, 两名医生均认为该病人所得的病是脑肿瘤 (T) 这件事的可能性很低, 但是应用 Dempster 组合规则融合后的结果却表明该病人得的病只可能是脑肿瘤 (T), 这与两名医生的判断均

相悖,不符合直观结果。这个例子表明,在证据存在较高冲突的情况下,应用 Dempster 组合规则得到的结果会反直观。此外,Dempster 组合规则并不满足幂等性,即使证据体都相同,组合过程中也会产生信任偏移^[15],在某些情况下还会出现证据吸收的问题^[16]。

针对证据冲突问题,一种观点归因于融合规则本身的缺陷,对融合规则进行了修正,如在 Yager 提出的规则^[6]中,将冲突赋给全集;Dubois 等则将焦元的并集引入到证据组合规则中^[7]。另一种观点认为该问题是由证据本身的不准确、不可靠所造成的,因此应该对证据进行进一步修正,如 Murphy 采用求平均的方式来对证据进行修正^[8],邓勇、肖建于、刘志成等人则基于证据距离来构造证据权值,对证据进行加权融合^[3-5]等。

2.2 基于证据距离的证据修正与组合方法

在基于证据距离的证据修正与组合方法中,邓勇等人所提出的方法具有代表性。在该方法中,核心概念为证据可靠度,它描述了证据的可靠程度。用证据可靠度作为权值来对证据进行加权,得到融合后的证据。关于证据可靠性的评估,学者们做了许多的研究。Elouedi 等人依据可传递信任模型,基于加权后的平均证据应与数据真值距离最小,利用优化方法得到证据可靠度^[17]。邢清华等人把各个焦元信任区间的长度之和作为目标函数,通过极小化目标函数得到证据可靠度^[18]。付耀文等人借鉴冲突处理中的 Dubois & Prade 规则,将由冲突得到的折扣量按局部冲突的大小分配给涉及各局部冲突的集合的并,得到证据可靠度^[19]。

在邓勇等人的方法中,证据可靠度是基于证据距离构建的,其构建过程如下^[3]

$$\text{sim}(m_i, m_j) = 1 - d(m_i, m_j) \quad (6)$$

$$\text{sup}(m_i) = \sum_{j=1, j \neq i}^s \text{sim}(m_i, m_j) \quad (7)$$

$$\text{cred}(m_i) = \frac{\text{sup}(m_i)}{\sum_{j=1}^s \text{sup}(m_j)} \quad (8)$$

式中: $\text{cred}(m_i)$ 为证据 m_i 的可靠度; $d(m_i, m_j)$ 表示证据 m_i 与 m_j 间的证据距离。

得到各个证据的可靠度之后,就可以用证据可靠度作为权值对证据进行加权融合,得到平均证据。

假设有 s 个证据,则平均证据为

$$m_{\text{ave}} = \sum_{i=1}^s \text{cred}(m_i) m_i \quad (9)$$

然后,应用 Dempster 组合规则对平均证据 m_{ave} 自身组合 $s-1$ 次。

在该方法中,证据可靠度的求取与证据距离紧密相关。随着证据理论的研究深入,出现了一系列的证据距离定义。不同的距离应用于证据修正,效果或有不同。本文将针对证据距离的不同,对于证据修正与组合效果的影响开展分析研究。在下一节中,将首先介绍几类常用的证据距离,并将不同的证据距离应用于冲突证据组合的方法中进行比较。

3 基于不同证据距离的加权证据融合

3.1 证据距离度量方式

证据距离用来描述证据之间的差异性或者相似性,两组证据之间的距离越大,表明它们之间的差异性越大,相似性越低。几类常用的证据距离度量方式有 Jousselme 距离^[10]、Tessem 距离^[11]、基于隶属度函数的距离^[12]和基于信任区间的距离^[13]。

3.1.1 Jousselme 距离 Jousselme 距离^[10]的求取过程如下

$$d_J(m_1, m_2) = \left[\frac{1}{2} (m_1 - m_2) \mathbf{D} (m_1 - m_2) \right]^{1/2} \quad (10)$$

式中: $\mathbf{D}$ 是一个 $2^n \times 2^n$ 的矩阵,其中的元素定义为 $\mathbf{D}(i, j) = |A_i^1 \cap A_j^2| / |A_i^1 \cup A_j^2|$ 。这里 A_i^k 表示 m_k 的第 i 个子集, $|A|$ 表示集合 A 中元素的个数。

3.1.2 Tessem 距离 在 Tessem 的研究中^[11],证据距离的度量是基于 pignistic 概率的。Mass 函数转化为 Pignistic 概率的公式如下

$$P_m(A) = \sum_{B \in \Theta} \frac{|A \cap B|}{|B|} m(B) \quad (11)$$

Tessem 距离定义如下

$$d_T(m_1, m_2) = \max_{A \in \Theta} \{ |P_1(A) - P_2(A)| \} \quad (12)$$

3.1.3 基于隶属度函数的距离^[12] 在模糊理论中,隶属度函数描述了元素属于某一模糊集合的程度。若 $\mu_A: \Theta \rightarrow [0, 1]$, $\theta \rightarrow \mu_A(\theta)$ 是在辨识框架 Θ 上的给定映射,则 A 是 Θ 上的模糊集合, $\mu_A(\theta)$ 是模糊集合 A 的隶属度函数, $\mu_A(\theta)$ 在没有混淆时简记为 $\mu(\theta)$ 。 $\mu(\theta_i)$ 为单点焦元 θ_i 的隶属度函数, $\mu(\theta_i)$ 可以取 θ_i 的信任函数或者似真度函数 $\mu(\theta_i) = \text{bel}(\theta_i)$, 或 $\mu(\theta_i) = \text{pl}(\theta_i)$, 则基于隶属度函数的证据距离定义为

$$d_F(m_1, m_2) = 1 - \frac{\sum_{i=1}^n (\mu^{(1)}(\theta_i) \wedge \mu^{(2)}(\theta_i))}{\sum_{i=1}^n (\mu^{(1)}(\theta_i) \vee \mu^{(2)}(\theta_i))} \quad (13)$$

式中:符号 $\wedge$ 为取小运算(合取);符号 $\vee$ 为取大运算(析取)。

在一个有限的辨识框架中,若单点焦元的隶属度函数之和小于 1,则隶属度函数等价于相应焦元的信任函数,若隶属度函数之和大于 1,则隶属度函数等价于相应焦元的似真度函数^[20]。就实际操作而言,在证据函数中,如果混合焦元多,单点焦元少,此时若使用单点焦元的信任函数作为隶属度函数,会出现较多的 0 元素,求取距离时实用性较差,因此更倾向于使用似真度函数作为单点焦元的隶属度函数。本文的实验中选取 $\mu(\theta_i)=\text{pl}(\theta_i)$ 。

3.1.4 基于信任区间的距离 本课题组以往的工作中设计了一种基于信任区间的距离^[13]。若一个焦元在两个证据中的信任区间 $([\text{bel}(A), \text{pl}(A)])$ 分别为 $[a_1, b_1]$ 和 $[a_2, b_2]$,则基于信任区间的距离定义为

$$d_{\text{BI}}([a_1, b_1], [a_2, b_2]) = \left[\left(\frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2} \right)^2 + \frac{1}{3} \left(\frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2} \right)^2 \right]^{1/2} \quad (14)$$

基于所有信任区间的两个 mass 函数之间的距离为

$$d_{\text{BI}}^E(m_1, m_2) = \left[N_c \sum_{i=1}^{2^n-1} [d_{\text{BI}}(B_1(A_i), B_2(A_i))]^2 \right]^{1/2} \quad (15)$$

式中: A_i 为焦元; $B_1(A_i)$ 和 $B_2(A_i)$ 分别表示 m_1 和 m_2 中焦元 A_i 的信任区间; $N_c=1/2^{n-1}$ 是归一化因子。

3.2 算例

为了研究证据距离度量方式对证据可靠度以及证据组合结果的影响,设计如下算例。

算例 1 设辨识框架为 $\Theta=\{A, B, C\}$,利用 6 个不同信息源得到的证据函数如下。

$$\begin{aligned} m_1: & m_1(A)=0.65, m_1(B)=0.2, m_1(C)=0.15; \\ m_2: & m_2(A)=0.55, m_2(B)=0.25, m_2(C)=0.2; \\ m_3: & m_3(A)=0.5, m_3(B)=0.3, m_3(C)=0.2; \\ m_4: & m_4(A)=0.3, m_4(B)=0.35, m_4(C)=0.35; \\ m_5: & m_5(A)=0.1, m_5(B)=0.3, m_5(C)=0.6; \\ m_6: & m_6(B)=0.8, m_6(C)=0.2。 \end{aligned}$$

直观来看,6 组证据中第 1、2、3 组证据相似,其中 $m(A)$ 最高,第 4 组证据中 $m(A)$ 、 $m(B)$ 、 $m(C)$ 大小接近,第 5 组证据中 $m(C)$ 最高,第 6 组证据中 $m(B)$ 最高。

使用邓勇的方法,在不同的证据距离度量方式下求得的各证据可靠度如表 1 所示。

表 1 算例 1 中不同证据距离度量方式下证据可靠度对比

距离	证据可靠度
d_J	$\text{cred}(m_1)=0.168\ 3; \text{cred}(m_2)=0.184\ 6$
	$\text{cred}(m_3)=0.188\ 1; \text{cred}(m_4)=0.184\ 8$
	$\text{cred}(m_5)=0.149\ 9; \text{cred}(m_6)=0.124\ 2$
d_T	$\text{cred}(m_1)=0.166\ 7; \text{cred}(m_2)=0.187\ 5$
	$\text{cred}(m_3)=0.192\ 7; \text{cred}(m_4)=0.182\ 3$
	$\text{cred}(m_5)=0.148\ 4; \text{cred}(m_6)=0.122\ 4$
d_F	$\text{cred}(m_1)=0.170\ 4; \text{cred}(m_2)=0.200\ 7$
	$\text{cred}(m_3)=0.206\ 0; \text{cred}(m_4)=0.182\ 9$
	$\text{cred}(m_5)=0.136\ 2; \text{cred}(m_6)=0.103\ 8$
d_{BI}	$\text{cred}(m_1)=0.168\ 3; \text{cred}(m_2)=0.184\ 6$
	$\text{cred}(m_3)=0.188\ 1; \text{cred}(m_4)=0.184\ 8$
	$\text{cred}(m_5)=0.149\ 9; \text{cred}(m_6)=0.124\ 2$

得到证据可靠度后,运用证据可靠度进行加权得到平均证据如表 2 所示。

表 2 算例 1 中不同证据距离度量方式下的平均证据

距离	组合结果
d_J	$m(A)=0.375\ 4, m(B)=0.345\ 3,$ $m(C)=0.279\ 3$
d_T	$m(A)=0.377\ 3, m(B)=0.344\ 3,$ $m(C)=0.278\ 4$
d_F	$m(A)=0.392\ 6, m(B)=0.334\ 0,$ $m(C)=0.273\ 4$
d_{BI}	$m(A)=0.375\ 4, m(B)=0.345\ 3,$ $m(C)=0.279\ 3$

应用 Dempster 组合规则对平均证据自身组合 5 次,得到融合后的 mass 函数如表 3 所示。

表 3 算例 1 中不同证据距离度量方式下组合结果对比

距离	组合结果
d_J	$m(A)=0.563\ 4, m(B)=0.341\ 1,$ $m(C)=0.095\ 5$
d_T	$m(A)=0.575\ 4, m(B)=0.331\ 8,$ $m(C)=0.092\ 8$
d_F	$m(A)=0.669\ 8, m(B)=0.253\ 8,$ $m(C)=0.076\ 4$
d_{BI}	$m(A)=0.563\ 4, m(B)=0.341\ 1,$ $m(C)=0.095\ 5$

表 1 中证据可靠度对比结果显示,根据 4 种证据距离度量方式得到的前 4 组证据的可靠度均比后两组证据的可靠度高,且最后一组证据的可靠度最小。由于证据可靠度是依据证据之间的相似度得到的,假设多数人给出的信息是可靠的,那么某一条证据与其他证据之间的相似度越高,该条证据的可靠度越大。综合 6 组证据来看,前 3 组证据之间的相似度大,它们之间相互支持,因此得到的前 3 组证据的可靠度高,第 4 组证据中 $m(A)$ 、 $m(B)$ 、 $m(C)$ 大小接近,后两组证据与前 4 组证据以及它们中的另一组证据的差异性都很大,所以后两组证据的可靠度较低。表 1 给出的结果与直观结论相符。

在只有单点焦元有值的情况下,焦元的 BetP 与 mass 函数值相等。表 3 中组合结果对比显示,所有度量方式下组合结果中 $P(A)$ 都是最大,原因在于前 3 组证据中的 $m(A)$ 均为最高,它们之间相互支持,得到的证据可靠度高,加权后平均证据中的 $m(A)$ 也是最大。在此算例中 4 种距离度量方式下的证据组合结果在决策方面并无差异。

算例 2 设辨识框架为 $\Theta=\{A,B,C\}$,利用 6 个不同信息源得到的证据函数如下。

$$m_1:m_1(A)=0.4,m_1(B)=0.3,m_1(A\cup C)=0.3;$$
$$m_2:m_2(A)=0.5,m_2(B)=0.2,m_2(A\cup B\cup C)=0.3;$$
$$m_3:m_3(A)=0.7,m_3(B)=0.2,m_3(A\cup B)=0.1;$$
$$m_4:m_4(A)=0.6,m_4(B)=0.3,m_4(B\cup C)=0.1;$$
$$m_5:m_5(A)=0.2,m_5(B)=0.7,m_5(A\cup C)=0.1;$$
$$m_6:m_6(B)=0.7,m_6(C)=0.1,m_6(A\cup B)=0.2。$$

与算例 1 不同的是,算例 2 中加入了混合焦元,这 6 组证据中前 4 组证据相似,其中 $m(A)$ 大于 $m(B)$,后两组证据中 $m(B)$ 最大。在不同的证据距离度量方式下求得各证据的可靠度,如表 4 所示。

表 4 中证据可靠度对比结果显示:在所有的证据距离度量方式下,前 4 组证据的可靠度均大于后两组证据的可靠度;前 4 组证据彼此之间的相似度高,支持度大,在假设多数人意见可靠的前提下得到的证据可靠度高,这与直观结果相符合;后两组证据与前 4 组证据之间的差异较大,可靠度较低。表 4 给出的结果与直观结论相符。

表 4 算例 2 中不同证据距离度量方式下证据可靠度对比

距离	证据可靠度
d_J	$\text{cred}(m_1)=0.177\ 4;\text{cred}(m_2)=0.177\ 5$
	$\text{cred}(m_3)=0.166\ 1;\text{cred}(m_4)=0.183\ 5$
	$\text{cred}(m_5)=0.160\ 4;\text{cred}(m_6)=0.135\ 2$
d_T	$\text{cred}(m_1)=0.181\ 2;\text{cred}(m_2)=0.186\ 0$
	$\text{cred}(m_3)=0.161\ 8;\text{cred}(m_4)=0.186\ 0$
	$\text{cred}(m_5)=0.154\ 6;\text{cred}(m_6)=0.130\ 4$
d_F	$\text{cred}(m_1)=0.177\ 4;\text{cred}(m_2)=0.184\ 5$
	$\text{cred}(m_3)=0.164\ 2;\text{cred}(m_4)=0.185\ 4$
	$\text{cred}(m_5)=0.157\ 0;\text{cred}(m_6)=0.131\ 5$
d_{BI}	$\text{cred}(m_1)=0.180\ 4;\text{cred}(m_2)=0.179\ 3$
	$\text{cred}(m_3)=0.164\ 8;\text{cred}(m_4)=0.183\ 5$
	$\text{cred}(m_5)=0.157\ 9;\text{cred}(m_6)=0.134\ 1$

运用证据可靠度对证据进行加权,得到平均证据,如表 5 所示。

表 5 算例 2 中不同证据距离度量方式下的平均证据

距离	组合结果
d_J	$m(A)=0.418\ 1,m(B)=0.383\ 9,m(C)=0.013\ 5$
	$m(A\cup B)=0.043\ 7,m(A\cup C)=0.069\ 2$
	$m(B\cup C)=0.018\ 3,m(A\cup B\cup C)=0.053\ 3$
d_T	$m(A)=0.421\ 3,m(B)=0.379\ 2,m(C)=0.013\ 0$
	$m(A\cup B)=0.042\ 3,m(A\cup C)=0.069\ 8$
	$m(B\cup C)=0.018\ 6,m(A\cup B\cup C)=0.055\ 8$
d_F	$m(A)=0.420\ 8,m(B)=0.380\ 5,m(C)=0.013\ 1$
	$m(A\cup B)=0.042\ 7,m(A\cup C)=0.068\ 9$
	$m(B\cup C)=0.018\ 5,m(A\cup B\cup C)=0.055\ 4$
d_{BI}	$m(A)=0.418\ 8,m(B)=0.382\ 4,m(C)=0.013\ 4$
	$m(A\cup B)=0.043\ 3,m(A\cup C)=0.069\ 9$
	$m(B\cup C)=0.018\ 3,m(A\cup B\cup C)=0.053\ 8$

表 6 组合结果中, $m(A\cup C)$ 很小,可以认为 A 、 B 、 C 的 BetP 与 A 、 B 、 C 的 mass 函数值相等。4 种距离度量方式下给出的组合结果相似,其中 $P(A)$ 大于 $P(B)$, $P(C)$ 最小。原因在于前 4 组证据中 $m(A)$ 最大,它们相互支持,证据可靠度高,在加权的过程中对平均证据的贡献较大,因此平均证据中的 $m(A)$ 也是最大,如表 5 所示。表 6 中的组合结果表明,在证据中存在混合焦元的情况下,4 种距离度量方式下的证据组合结果在决策方面也没有差异。

表 6 算例 2 中不同证据距离度量方式下
组合结果对比

距离	组合结果
d_J	$m(A)=0.719\ 9, m(B)=0.279\ 9$
	$m(C)=0.000\ 2, m(A\cup C)=0.000\ 1$
d_T	$m(A)=0.737\ 4, m(B)=0.262\ 3$
	$m(C)=0.000\ 2, m(A\cup C)=0.000\ 1$
d_F	$m(A)=0.731\ 8, m(B)=0.267\ 9$
	$m(C)=0.000\ 2, m(A\cup C)=0.000\ 1$
d_{BI}	$m(A)=0.726\ 2, m(B)=0.273\ 5$
	$m(C)=0.000\ 2, m(A\cup C)=0.000\ 1$

3.2.3 有关证据距离的分析与探讨 严谨的证据距离应满足以下 4 条性质^[13]:

- (1)非负性, $d(x,y)\geq 0$;
- (2)非退化性, $d(x,y)=0\iff x=y$;
- (3)对称性, $d(x,y)=d(y,x)$;
- (4)三角不等式, $d(x,y)+d(y,z)\geq d(x,z)$ 。

在 3.1 节提及的 4 种证据距离中, d_J 、 d_{BI} 是严谨的证据距离^[10,13],满足全部 4 条性质。 d_F 、 d_T 不是严谨的证据距离。算例 1 和算例 2 中 4 种证据距离度量方式下证据可靠度以及证据组合结果均直观、合理,但是在这 4 种证据距离中, d_F 、 d_T 这两种距离在计算的过程中进行了框架的切换。其中, d_T 将证据函数转化到概率框架下计算, d_F 则将证据函数转化到模糊理论的框架下计算,在框架切换的过程中产生了信息损失^[13],由此造成了 d_F 、 d_T 的不严谨性。为了进一步分析证据距离对证据可靠度的影响,举出下面两个算例。

算例 3 设辨识框架为 $\Theta=\{A,B,C\}$,利用 6 个不同信息源得到的证据函数如下。

$m_1:m_1(A)=0.45, m_1(B)=0.35, m_1(C)=0.05, m_1(A\cup B)=0.1, m_1(A\cup B\cup C)=0.05;$

$m_2:m_2(A)=0.4, m_2(B)=0.3, m_2(C)=0.1, m_2(A\cup B)=0.2;$

$m_3:m_3(A)=0.5, m_3(B)=0.4, m_3(A\cup B\cup C)=0.1;$

$m_4:m_4(A)=0.48, m_4(B)=0.38, m_4(C)=0.02, m_4(A\cup B)=0.04, m_4(A\cup B\cup C)=0.08;$

$m_5:m_5(A)=0.5, m_5(B)=0.5;$

$m_6:m_6(B)=0.8, m_6(C)=0.2。$

在这 6 组证据中,前 4 组证据相似,其中 $m(A)$ 略大于 $m(B)$, $m(C)$ 最小。第 5 组证据中 $m(A)$ 与 $m(B)$ 相同,与前 4 组证据有一定的相似度。第 6 组

证据中 $m(B)$ 远大于 $m(C)$,与前 5 组证据的差异较大。在不同证据距离度量方式下得各证据的可靠度,如表 7 所示。

表 7 算例 3 中不同证据距离度量方式下
证据可靠度对比

距离	证据可靠度
d_J	$\text{cred}(m_1)=0.180\ 0; \text{cred}(m_2)=0.174\ 3$
	$\text{cred}(m_3)=0.179\ 0; \text{cred}(m_4)=0.180\ 6$
	$\text{cred}(m_5)=0.173\ 9; \text{cred}(m_6)=0.112\ 1$
d_T	$\text{cred}(m_1)=0.181\ 5; \text{cred}(m_2)=0.178\ 7$
	$\text{cred}(m_3)=0.180\ 4; \text{cred}(m_4)=0.181\ 5$
	$\text{cred}(m_5)=0.175\ 8; \text{cred}(m_6)=0.102\ 0$
d_F	$\text{cred}(m_1)=0.188\ 4; \text{cred}(m_2)=0.188\ 4$
	$\text{cred}(m_3)=0.188\ 4; \text{cred}(m_4)=0.188\ 4$
	$\text{cred}(m_5)=0.164\ 2; \text{cred}(m_6)=0.082\ 1$
d_{BI}	$\text{cred}(m_1)=0.179\ 8; \text{cred}(m_2)=0.176\ 0$
	$\text{cred}(m_3)=0.179\ 0; \text{cred}(m_4)=0.180\ 1$
	$\text{cred}(m_5)=0.174\ 0; \text{cred}(m_6)=0.111\ 1$

从表 7 中的证据可靠度对比结果来看,4 种证据距离度量方式下得到的前 5 组证据的可靠度都比较高,最后一组证据的可靠度很低。在假设多数人的信息可靠的前提下与直观结果相吻合。

在 d_F 度量方式下,前 4 组证据的可靠度相同,且与后两组证据的可靠度相差较大,但是前 4 组证据并不相同。事实上基于隶属度函数的距离与证据之间的差异并不是一一对应的,如果证据中单点焦元的信任函数或者似真度函数相同,那么在 d_F 度量方式下他们之间的距离为零,即 d_F 不满足非退化性。本例中前 4 组证据单点焦元的似真度函数相同,即

$\text{pl}(A)=0.6; \text{pl}(B)=0.5; \text{pl}(C)=0.1$
则依据式(13), m_1 、 m_2 之间的距离为

$$d_F(m_1, m_2) = 1 - \frac{0.6 + 0.5 + 0.1}{0.6 + 0.5 + 0.1} = 0$$

类似地,求得这 4 条证据中任意两条证据之间的距离为零,从而得到这 4 组证据的可靠度相同。这表明 d_F 度量方式并不能体现出某些证据之间的差异。由此得到的可靠度必然存在偏差。可见证据距离的不严谨会对证据可靠度产生影响。

算例 4 设辨识框架为 $\Theta=\{A,B,C\}$,利用 6 个不同信息源得到的证据函数如下。

$m_1:m_1(A)=0.4, m_1(B)=0.4, m_1(C)=0.2;$

$m_2:m_2(A\cup B)=0.8, m_2(C)=0.2;$

$m_3:m_3(A)=0.4,m_3(B)=0.2,m_3(B\cup C)=0.4;$
 $m_4:m_4(A)=0.2,m_4(C)=0.2,m_4(A\cup B\cup C)=0.6;$
 $m_5:m_5(C)=0.65,m_5(A\cup B\cup C)=0.35;$
 $m_6:m_6(A)=0.35,m_6(C)=0.65。$

在这 6 组证据中:第 1 组证据的 $m(A)$ 、 $m(B)$ 相同,且大于 $m(C)$,第 3 组证据中 $m(A)$ 大于 $m(B)$;第 2 组证据中 $m(C)$ 很小, $m(A\cup B)$ 较大;第 4 组证据中 $m(A)$ 、 $m(C)$ 相同,均很小,第 5 组证据中 $m(C)$ 较大,第 6 组证据中 $m(A)$ 小于 $m(C)$;第 1、第 3、第 4 组证据较为相似。在不同的证据距离度量方式下求得各证据的可靠度,如表 8 所示。

表 8 算例 4 中不同证据距离度量方式下
证据可靠度对比

距离	证据可靠度
d_J	$\text{cred}(m_1)=0.176\ 0;\text{cred}(m_2)=0.150\ 9$
	$\text{cred}(m_3)=0.175\ 9;\text{cred}(m_4)=0.177\ 9$
	$\text{cred}(m_5)=0.156\ 5;\text{cred}(m_6)=0.162\ 8$
d_T	$\text{cred}(m_1)=0.180\ 2;\text{cred}(m_2)=0.180\ 2$
	$\text{cred}(m_3)=0.180\ 2;\text{cred}(m_4)=0.180\ 2$
	$\text{cred}(m_5)=0.128\ 6;\text{cred}(m_6)=0.150\ 8$
d_F	$\text{cred}(m_1)=0.167\ 6;\text{cred}(m_2)=0.155\ 3$
	$\text{cred}(m_3)=0.190\ 2;\text{cred}(m_4)=0.182\ 6$
	$\text{cred}(m_5)=0.167\ 9;\text{cred}(m_6)=0.136\ 5$
d_{BI}	$\text{cred}(m_1)=0.173\ 8;\text{cred}(m_2)=0.168\ 5$
	$\text{cred}(m_3)=0.177\ 4;\text{cred}(m_4)=0.176\ 8$
	$\text{cred}(m_5)=0.147\ 0;\text{cred}(m_6)=0.156\ 5$

从表 8 中的证据可靠度对比结果来看,在 d_J 、 d_F 、 d_{BI} 度量方式下得到的第 1、3、4 组的证据可靠度较大,剩余 3 组的证据可靠度较小。

在 d_T 度量方式下得到的前 3 组证据的可靠度相同,但是前 3 组证据并不相同。事实上,Tessem 距离是根据 pignistic 概率求得的,如果证据的 pignistic 概率相同,那么在 Tessem 距离的度量方式下,证据之间的差异就是零。即 d_T 不满足非退化性。本例的前 3 条证据中,各条证据之间单点焦元的 pignistic 概率都相同,即

$P(A)=0.4;\ P(B)=0.4;\ P(C)=0.2$
则依据式(12), m_1 、 m_2 之间的距离为
 $d_T(m_1,m_2)=\max\{|0.4-0.4|,|0.4-0.4|,$
 $\quad |0.2-0.2|\}=0$

类似地,求得这 3 条证据中任意两条证据之间

的距离为零,且其他证据到这 3 条证据的距离都相等,从而得到这 3 组证据的可靠度相同。这表明, d_T 度量方式下距离与证据之间的差异也不是一一对应的。可见,证据距离的不严谨会对证据可靠度产生影响。

证据可靠度以及证据组合的合理性受到诸多因素影响,如果基于证据距离构造可靠度并用于证据组合,距离的选取只是其中的一个因素,使用严谨的证据距离并不能够保证得到的证据可靠度和证据组合结果最优。本文关注证据距离对证据可靠度和证据组合的影响,算例表明严谨的证据距离能够更好地表征证据之间的差异,在基于证据距离生成证据可靠度时结果更符合直观。

4 结 论

本文提出一种将不同的证据距离度量方式应用在加权证据组合中的对比方法,对不同证据距离下的加权证据组合过程进行分析比较。实验结果表明,不同的证据距离度量方式对证据可靠度的求取有影响,其中使用 d_F 、 d_T 度量方式在算例中得到的证据可靠度不能体现出证据之间的差异,这与 d_F 、 d_T 的不严谨性有关。本文建议选择 d_J 、 d_{BI} 这两种严谨的证据距离。

对于证据冲突问题,本文的解决方式是修改证据体。在未来的工作中,我们将更加关注有关修改证据规则的研究,关注当存在证据冲突时,各类证据距离度量方式对修改后的证据规则的影响。

参考文献:

[1] SHAFER G. A mathematical theory of evidence [M]. Princeton, NJ, USA: Princeton University Press, 1976: 1-10.
[2] ZADEH L A. Review of Shafer’s a mathematical theory of evidence [J]. AI Magazine, 1984, 5(3): 81-83.
[3] DENG Yong, SHI Wenkang, ZHU Zhenfu, et al. Combining belief functions based on distance of evidence [J]. Decision Support Systems, 2004, 38(3): 489-493.
[4] 肖建于,童敏明,朱昌杰,等. 基于 pignistic 概率距离的改进证据组合规则 [J]. 上海交通大学学报, 2012, 46(4): 636-641.
XIAO Jianyu, TONG Minming, ZHU Changjie, et al. Improved combination rule of evidence based on pignistic probability distance [J]. Journal of Shanghai Jiaotong University, 2012, 46(4): 636-641.

- [5] 刘志成, 乔慧, 何佳洲. 基于加权证据距离的高度冲突证据组方法 [J]. 计算机工程与应用, 2014, 50(3): 103-107.
LIU Zhicheng, QIAO Hui, HE Jiazhou. Combination approach of highly conflicting evidence based on weighted distance of evidence [J]. Computer Engineering & Applications, 2014, 50(3): 103-107.
- [6] YAGER R R. On the Dempster-Shafer framework and new combination rules [J]. Information Science, 1987, 41(2): 93-137.
- [7] DUBOIS D, PRADE H. Representation and combination of uncertainty with belief functions and possibility measures [J]. Computational Intelligence, 1988(4): 244-264.
- [8] MURPHY C K. Combining belief functions when evidence conflicts [J]. Decision Support Systems, 2000, 29(1): 1-9.
- [9] JOUSSELME A L, PATRICK M. Distance in evidence theory: comprehensive survey and generalization [J]. International Journal of Approximate Reasoning, 2012, 53(2): 118-145.
- [10] JOUSSELME A L, GRENIER D, BOSSÉ E. A new distance between two bodies of evidence [J]. Information Fusion, 2001, 2(2): 91-101.
- [11] TESSEM B. Approximations for efficient computation in the theory of evidence [J]. Artificial Intelligence, 1993, 61(2): 315-329.
- [12] HAN Deqiang, DEZERT J, HAN Chongzhao, et al. New dissimilarity measures in evidence theory [C]// Proceedings of the 14th International Conference on Information Fusion. Piscataway, NJ, USA: IEEE, 2011: 5977681.
- [13] HAN Deqiang, DEZERT J, YANG Yi. Belief interval-based distance measures in the theory of belief functions [J]. IEEE Transactions on Systems Man & Cybernetics Systems, 2016, 99: 1-18.
- [14] SMETS P. Data fusion in the transferable belief model [C]// Proceedings of the 3rd International Conference on Information Fusion. Piscataway, NJ, USA: IEEE Computer Society, 2000: 21-33.
- [15] 杨艺, 韩德强, 韩崇昭. 基于多准则排序融合的证据组方法 [J]. 自动化学报, 2012, 38(5): 823-831.
YANG Yi, HAN Deqiang, HAN Chongzhao. Evidence combination based on multi-criteria rank-level fusion [J]. Acta Automatica Sinica, 2012, 38(5): 823-831.
- [16] 杨风暴, 王肖霞. D-S 证据理论的冲突证据合成方法 [M]. 北京: 国防工业出版社, 2010, 70-71.
- [17] ELOUEDI Z, MELLOULI K, SMETS P. Assessing sensor reliability for multisensor data fusion within the transferable belief model [J]. IEEE Transactions on Systems Man & Cybernetics, Part B Cybernetics, 2004, 34(1): 782-787.
- [18] 邢清华, 刘付显. 基于证据折扣优化的冲突证据组方法 [J]. 系统工程与电子技术, 2009, 31(5): 1158-1161.
XING Qinghua, LIU Fuxian. New combination rule of conflict evidence based on optimized evidence discount [J]. Systems Engineering & Electronics, 2009, 31(5): 1158-1161.
- [19] 付耀文, 贾宇平, 杨威, 等. 传感器动态可靠性评估与证据折扣 [J]. 系统工程与电子技术, 2012, 34(1): 212-216.
FU Yaowen, JIA Yuping, YANG Wei, et al. Sensor dynamic reliability evaluation and evidence discount [J]. Systems Engineering & Electronics, 2012, 34(1): 212-216.
- [20] DE FERIET J K. Interpretation of membership functions of fuzzy sets in terms of plausibility and belief [J]. Fuzzy Information and Decision Processes, 1982(1): 93-98.

[本刊相关文献链接]

- 魏港明, 刘真, 李林峰, 等. 加入用户对项目属性偏好的奇异值分解推荐算法. 2018, 52(5): 101-107. [doi: 10.7652/xjtuxb201805015]
- 蒋逸飞, 张俊, 何勇, 等. 面向裂纹扩展的连结型刚体-弹簧离散元模型. 2018, 52(4): 48-54. [doi: 10.7652/xjtuxb201804007]
- 赵振冲, 王晓丹. 引入拒识的最小风险弹道目标识别. 2018, 52(4): 132-138. [doi: 10.7652/xjtuxb201804019]
- 卫博雅, 寿业航, 蒋雯. 一种可视化的加权平均信息融合方法. 2018, 52(4): 145-149. [doi: 10.7652/xjtuxb201804021]
- 姜洪权, 王岗, 高建民, 等. 一种适用于高维非线性特征数据的聚类算法及应用. 2017, 51(12): 49-55. [doi: 10.7652/xjtuxb201712008]
- 周亚东, 陈凯悦, 冷俊园, 等. 软件定义网络流表溢出脆弱性分析及防御方法. 2017, 51(10): 53-58. [doi: 10.7652/xjtuxb201710009]

(编辑 刘杨)