

维修性多源冲突证据数据融合的先验分布确定方法

周 诚¹, 徐 达^{2*}, 曹振地³

(1. 中国人民解放军陆军兵种大学, 江苏 南京 211132; 2. 中国人民解放军陆军兵种大学, 北京 100072;
3. 中国人民解放军 75560 部队, 海南 屯昌 571600)

摘 要:为充分利用维修性多源先验数据,提高先验分布的融合精度以确保维修性验证结果的准确性,针对多源数据存在冲突的问题,提出维修性多源冲突证据数据融合的先验分布确定方法。充分挖掘、提取多源数据的特征信息,分别构造基于样本量、分布特征和数据重要度的证据 mass 函数,综合考虑证据间的关联性和证据本身的不确定性,引入夹角余弦度量证据间的冲突程度,引入信息熵度量证据的不确定度,通过结合证据的支持度和不确定度共同修正证据,建立多源冲突证据数据融合模型以实现多源数据的有效融合,确定综合先验分布。结合两个案例进行分析,验证了所提方法有效可行。

关键词:维修性;数据融合;冲突证据;夹角余弦;信息熵

中图分类号:TP301.6 文献标志码:A 文章编号:1001-2486(2025)05-225-11



论
文
拓
展

Prior distribution determination method for maintainability multi-source conflict evidence data fusion

ZHOU Cheng¹, XU Da^{2*}, CAO Zhendi³

(1. Army Arms University of PLA, Nanjing 211132, China; 2. Army Arms University of PLA, Beijing 100072, China;
3. The PLA Unit 75560, Tunchang 571600, China)

Abstract: To make full use of maintainability multi-source prior data and improve the fusion accuracy of prior distributions to ensure the accuracy of maintainability verification results, a prior distribution determination method for maintainability multi-source conflict evidence data fusion was proposed to solve the problem of conflicts in multi-source data. Fully mining and extracting the feature information of multi-source data, constructing evidence mass function based on sample size, distribution characteristics and data importance respectively, comprehensively considering the correlation between evidence and the uncertainty of evidence itself, introducing angle cosine to measure the degree of conflict between evidence. By combining the support degree and uncertainty of the evidence, the multi-source conflict evidence data fusion model was established to achieve the effective fusion of multi-source data and determine the comprehensive prior distribution. Combined with the analysis of two cases, the proposed method is proved to be effective and feasible.

Keywords: maintainability; data fusion; conflict evidence; angle cosine; information entropy

复杂装备系统在开展维修性试验与评价工作时往往受限于试验周期和试验费用等条件,导致现场试验样本量常达不到经典数理统计法^[1-3]验证装备维修性时对样本数量的要求,而 Bayes 法^[4-5]与经典数理统计法相比,在利用总体信息和样本信息的基础上,进一步引入了先验信息,更适用于小样本条件下维修性定量指标的验证问

题。采用 Bayes 法准确验证装备维修性的前提和关键是获取准确的先验分布^[6],尤其是在具有多源先验数据的情况下,需要先进行多源数据的有效融合,考虑到多源数据具有不确定性,且不同源数据之间可能相互冲突,所以引入 DS(Dempster-Shafer)证据理论^[7-8]对维修性多源数据进行融合能充分利用多源数据信息。但当维修性多源数据

收稿日期:2023-05-12

基金项目:国家部委基金资助项目(2020ZB20)

第一作者:周诚(1996—),男,安徽太和人,讲师,博士,E-mail:zhou_cheng3@163.com

*通信作者:徐达(1969—),男,辽宁丹东人,教授,博士,博士生导师,E-mail:ljzjbxy2022@163.com

引用格式:周诚,徐达,曹振地.维修性多源冲突证据数据融合的先验分布确定方法[J].国防科技大学学报,2025,47(5):225-235.

Citation: ZHOU C, XU D, CAO Z D. Prior distribution determination method for maintainability multi-source conflict evidence data fusion[J]. Journal of National University of Defense Technology, 2025, 47(5): 225-235.

的证据高度冲突时,已知 DS 理论的 Dempster 组合规则此时不能有效地处理高冲突证据,融合易出现与事实相悖的现象,也无法得到准确的先验分布^[9]。

当前,研究人员针对 DS 理论不能有效处理高冲突证据的原因主要持两种观点:一是认为 Dempster 组合规则有问题,该规则不能合理地分配冲突;二是认为组合规则没问题,缺陷是冲突证据本身造成的。相应地,对 DS 理论进行改进也主要从以下两个方向展开:一是修改组合规则^[10-15],这类方法将冲突悖论的产生归因于组合规则的不合理,通过修改组合规则虽能在一定程度上解决高冲突证据的融合问题,消除冲突悖论,但却破坏了 Dempster 组合规则的交换律、结合律等优良性质,融合大量证据时效果较差,体现了这类改进方法的局限性。二是修正证据模型^[16-23],这类方法将冲突悖论的产生归因于冲突证据本身,通过赋予证据权重修正证据,以减弱证据间的冲突程度,可以灵活高效地处理高冲突证据。比如, Murphy^[16]率先提出了平均加权融合证据的方法,首先赋予每个证据以相同权重,然后采用 Dempster 组合规则对加权平均证据融合 $n-1$ 次,融合结果表明该方法具有较强的收敛性,但简单地将证据等权处理则忽略了证据间的关联性和证据本身的不确定性,无法确保融合结果的准确度。Deng 等^[17]在借鉴 Murphy 思想的基础上,引入了 Jousselme 距离度量证据间的冲突程度,并以此确定证据的权重,一定程度上弥补了平均加权法的不足。Chen 等^[18]提出了一种借助夹角余弦度量证据间冲突程度的加权融合方法,也能较好地解决高冲突证据的融合问题。Ye 等^[19]利用 Lance 距离修正证据的性能优于 Minkowski 距离和 Mahalanobis 距离的特点^[20],改进了 DS 证据融合方法,实现了多传感器冲突信息的可靠融合。应文健等^[20]通过引入 Lance 距离、信度熵和专家打分法融合了小子样条件下舰炮制导弹药的测试性冲突数据,提高了舰炮制导弹药测试性指标评估的可靠性。Yuan 等^[23]提出了一种结合 Jousselme 距离和信息熵确定证据加权系数的新方法,既考虑了证据间的关联性,又考虑了证据本身的不确定性,与只考虑单一方面的证据修正方法相比,结果表明这种组合方法融合性能更好。

综上,采用修正证据模型的方式处理高冲突证据,同时考虑证据间的关联性和证据本身的不确定性两方面赋予证据权重修正证据的性能优于只考虑单一方面的证据修正方法,且以证据冲突

度量方法和信息熵加权组合的方式应用最为广泛。但当前研究更多的是一种普适性方法,虽能解决维修性多源冲突证据数据融合的问题,但其融合性能仍有待进一步提高,尚缺少针对维修性数据融合方法的研究。

为此,针对维修性多源先验数据这种只包含单子集焦元证据的融合问题,充分挖掘、提取多源数据的特征信息,分别构造基于样本量、分布特征和数据重要度的证据 mass 函数;引入夹角余弦度量证据间的冲突程度,引入信息熵度量证据的不确定度,并综合证据的支持度和不确定度共同修正证据,建立多源冲突证据数据融合模型以实现多源数据的有效融合,提出一种维修性多源冲突证据数据融合的先验分布确定方法。为验证所提方法的可行性和有效性,最后结合数值案例和实装试验应用案例进行分析。

1 DS 证据理论

DS 理论主要包含以下几个重要定义:

定义 1 识别框架: 设 $\Theta = \{A_1, A_2, \dots, A_n\}$ 是关于命题 A 的集合,若集合 Θ 有限、完备且所包含的 n 个子集互不相容,则将 Θ 定义为识别框架。幂集是由 Θ 中命题所有可能组合的集合,记为 $2^\Theta, 2^\Theta = \{\phi, \{A_1\}, \dots, \{A_n\}, \{A_1, A_2\}, \dots, \{A_1, A_n\}, \dots, \Theta\}$ 。

定义 2 mass 函数: 设 A 为幂集 2^Θ 中的任一命题, $A \in 2^\Theta$, 函数 m 为集合 $2^\Theta \rightarrow [0, 1]$ 的映射,且满足

$$\begin{cases} m(\phi) = 0 \\ \sum_{A \in 2^\Theta} m(A) = 1 \\ m(A) \geq 0 \quad \forall A \in 2^\Theta \text{ 且 } A \neq \phi \end{cases} \quad (1)$$

则称 m 为 Θ 上的 mass 函数或基本概率赋值 (basic probability assignment, BPA) 函数, $m(A)$ 为命题 A 的 BPA, 表示证据对命题 A 的支持度。如果 $m(A) > 0$, 则称命题 A 为焦元。使用向量 $\mathbf{m}_i$ 表示第 i 个证据的 BPA, $\mathbf{m}_i = [m_i(A_1), m_i(A_2), \dots, m_i(A_n)]$ ($i = 1, 2, \dots, M$), $\Psi = \{A_1, A_2, \dots, A_n\}$ 为证据 $\mathbf{m}_i$ 中所有焦元的集合 ($A_1, A_2, \dots, A_n \subseteq 2^\Theta$), 每一个命题 A_j ($j = 1, 2, \dots, n$) 都有确定的 $m(A_j)$ 。

定义 3 组合规则: 设 $m_1, m_2, \dots, m_N$ 分别为识别框架 Θ 上 N 个证据的 mass 函数, 则 DS 组合规则定义为

$$m(A) = \begin{cases} 0 & A = \phi \\ \frac{1}{1-k} \sum_{\cap A_i = A} \prod_{i=1}^N m_i(A_i) & A \neq \phi \end{cases} \quad (2)$$

式中, k 为冲突系数, $k = \sum_{\cap A_i = \phi} \prod_{i=1}^N m_i(A_i)$, $k \in [0, 1)$, 表示两个证据间的冲突程度。

但实际上, DS 证据理论往往无法正确融合高冲突证据, 融合结果易与事实相悖。例如, 假设采用 DS 证据理论融合三种维修性先验数据 A_1 、 A_2 和 A_3 , 识别框架上三个证据的 BPA 分别为 m_1 : $m_1(A_1) = 0.5, m_1(A_2) = 0.2, m_1(A_3) = 0.3$; m_2 : $m_2(A_1) = 0, m_2(A_2) = 0.9, m_2(A_3) = 0.1$; m_3 : $m_3(A_1) = 0.55, m_3(A_2) = 0.1, m_3(A_3) = 0.35$ 。

此时冲突系数 $k = 0.9731$, 表明该组证据为高冲突证据, 其中, 证据 1 和证据 3 均支持数据 A_1 , 说明这两个证据均认为数据 A_1 在多源数据融合时的权重应最大, 证据 2 支持数据 A_2 , 说明证据 2 认为数据 A_2 的融合权重应最大。根据融合结果应与证据保持相对一致的原则, 正确的融合结果应赋予数据 A_1 最大的融合权重, 但实际的融合结果为 $m(A_1) = 0, m(A_2) = 0.6316, m(A_3) = 0.3684$, 数据 A_2 的融合权重最大, 为 0.6316, 而数据 A_2 的融合权重却为 0, 显然与事实相悖, 表明此时 DS 证据理论无法正确融合维修性多源冲突数据, 因此, 需要对 DS 证据理论进行针对性改进。

2 构造 mass 函数

已知进行维修性多源数据有效融合的关键是确定各先验分布合理的融合权重, 因此基于 DS 理论融合多源数据时就需要充分挖掘并提取影响融合权重确定的数据信息。通常情况下, 可以从数据内部和外部两方面挖掘、提取数据的多种特征信息, 其中, 常见的内部信息包括样本量、参数分布特征和重要度等; 外部信息包括数据采集方法、采集水平和可信度等。在利用 DS 理论融合多源数据时, 就是将识别框架 Θ 看作是由多源数据构成的, 不同源数据分别表示识别框架上的不同命题, 数据的多种特征信息则看作是多个证据, 但由于不同的数据特征信息具有不同的表示形式, 不同形式的信息无法进行直接融合, 因此需要进一步对数据的特征信息进行预处理, 即构造证据的 mass 函数。此外, 通常认为获取多源数据的外部信息均一致, 即数据采集方法、采集水平和可信度均相同, 所以证据 mass 函数的构造只需考虑影响多源数据融合权重的内部特征信息。

2.1 构造基于样本量的 mass 函数

基于数据的样本量构造 mass 函数, 最直接的方法就是用各源数据的样本量占多源数据总样本

量的比值作为对应数据源数据的 mass 函数值。显然, 数据源数据的样本量越多, 表明该源数据反映装备维修性水平的代表性越强, 则对应数据源的融合权重应越大, 即 mass 函数值应越大。假设有 n 种维修性多源先验数据 $A_1, A_2, \dots, A_n$, 各源数据的样本量分别为 $s_1, s_2, \dots, s_n$, 则证据 E_1 的 mass 函数为:

$$E_1 = \{m_1(A_i) = s_i / \sum_{i=1}^n s_i\} \quad i = 1, 2, \dots, n \tag{3}$$

2.2 构造基于分布特征的 mass 函数

基于分布特征构造 mass 函数, 首先需要获取维修性多源数据的参数分布, 而进行多源数据融合前, 也需要先将各先验数据转换成维修性参数分布的形式, 先验数据的分布函数与现场试验数据的分布函数越相似, 表明该源数据与现场试验数据越接近, 反映装备维修性真实水平的能力越强, 对应数据源的融合权重应越大, 即 mass 函数值应越大。工程实践中普遍采用 Bootstrap 法和 Bayes Bootstrap 法对小样本数据进行重抽样获取新样本, 并以子样估计量统计推断维修性参数的总体分布^[24]。在此不再介绍两种方法的基本原理, 仅讨论其数据重抽样方式对参数分布拟合精度的影响。

设某源维修性数据为 $\mathbf{X} = [x_1, x_2, \dots, x_n]$, $x_i \sim F(x) (i = 1, 2, \dots, n)$ 将 $\mathbf{X}$ 中数据按升序排列, 记为 $x_{(1)}, x_{(2)}, \dots, x_{(n)}$, 构造经验分布函数 $F_n(x)$:

$$F_n(x) = \begin{cases} 0 & x < x_{(1)} \\ \frac{i}{n} & x_{(i)} \leq x < x_{(i+1)} \quad i \in [1, n-1] \\ 1 & x \geq x_{(n)} \end{cases} \tag{4}$$

传统 Bootstrap 法和 Bayes Bootstrap 法均依据式(4)所示的经验分布函数产生新样本, 受经验分布函数 $F_n(x)$ 结构的约束, 易知新样本数据均局限于初始样本区间 $[x_1, x_n]$ 内, 随机性较差, 尤其无法描述总体为连续分布时非初始样本点外的分布特征, 将降低参数分布的估计精度。针对 $F_n(x)$ 重抽样无法获取 $[x_1, x_n]$ 区间外样本点的问题, 本文采用经验分布函数 $F_n(x)$ 近似服从“S”形的 Boltzmann sigmoid 函数在全局 $(-\infty, +\infty)$ 上拟合样本, 并采用式(5)对拟合曲线进一步修正, 以弥补 $F_n(x)$ 上下限拟合效果的不足, 得到修正后的经验分布函数 $F^{\wedge}(x)$ 及其密度函数 $f^{\wedge}(x)$ 。

$$\int_{-\infty}^{+\infty} x f^{\wedge}(x) dx = \text{初始样本均值} \tag{5}$$

生成一组服从 $U(0,1)$ 的随机数 $\beta = [\beta_1, \beta_2, \cdots, \beta_n]$, 则

$$x^\wedge = F^{\wedge -1}(\beta) \tag{6}$$

式中, $x^\wedge$ 为一组包含 n 个数据的新样本, $x^\wedge = (x_1^\wedge, x_2^\wedge, \cdots, x_n^\wedge)$ 。

重复上述方法 m 次, 即可获取 m 组新样本。在已获取维修性多源先验分布和现场分布的基础上, 借助 KL (Kullback-Leibler) 散度^[25] 衡量各先验分布与现场分布的差异大小, 反映各先验分布与现场分布的相似程度, 构造基于分布特征的 mass 函数。

设有 n 个维修性多源先验数据, 已知关于维修性参数 θ 的先验分布分别为 $\pi_1(\theta), \pi_2(\theta), \cdots, \pi_n(\theta)$, 现场分布为 $\pi_0(\theta)$, 则第 i 个先验分布 $\pi_i(\theta) (i = 1, 2, \cdots, n)$ 对现场分布 $\pi_0(\theta)$ 的 KL 散度为:

$$D_{\text{KL}}(\pi_i \parallel \pi_0) = \int_{-\infty}^{+\infty} \pi_i(\theta) \log \frac{\pi_i(\theta)}{\pi_0(\theta)} \mathrm{d}x \tag{7}$$

由于 KL 散度具有不对称性, 即 $KL(\pi_i \parallel \pi_0) \neq KL(\pi_0 \parallel \pi_i)$, 所以当分别以先验分布和现场分布作为基准时, 会得到不同的 KL 散度值, 此时无法根据 KL 散度比较各先验分布与现场分布的相似度。为克服 KL 散度的非对称性, 对 KL 散度进行对称化改进, 令:

$$\begin{aligned} D(\pi_i, \pi_0) &= D_{\text{KL}}(\pi_i \parallel \pi_0) + D_{\text{KL}}(\pi_0 \parallel \pi_i) \\ &= \int_{-\infty}^{+\infty} (\pi_i(\theta) - \pi_0(\theta)) \log \frac{\pi_i(\theta)}{\pi_0(\theta)} \mathrm{d}x \end{aligned} \tag{8}$$

式中: $D(\pi_i, \pi_0)$ 为 $\pi_i(\theta)$ 与 $\pi_0(\theta)$ 的对称性差异度; $D(\pi_i, \pi_0)$ 越大, 表示两个概率分布的差异越大, 相似度越小, 反之亦然。特殊情况, 当 $\pi_i(\theta) = \pi_0(\theta)$, $D(\pi_i, \pi_0) = 0$, 表示两个概率分布完全相同, 此时可将先验数据作为现场试验数据直接使用。

利用式 (9) 将 $\pi_i(\theta)$ 与 $\pi_0(\theta)$ 的差异度 $D(\pi_i, \pi_0)$ 转化为 $[0, 1]$ 区间的相似度, 记为 $\rho(\pi_i, \pi_0)$:

$$\rho(\pi_i, \pi_0) = 1 - \frac{D(\pi_i, \pi_0)}{1 + D(\pi_i, \pi_0)} = \frac{1}{1 + D(\pi_i, \pi_0)} \tag{9}$$

因此, 进行多源先验分布融合时, 先验分布与现场分布越相似, 融合权重应越大, 则证据 E_2 的 mass 函数为:

$$E_2 = \{m_2(A_i) = \frac{\rho(\pi_i, \pi_0)}{\sum_{i=1}^n \rho(\pi_i, \pi_0)}\} \quad i = 1, 2, \cdots, n \tag{10}$$

2.3 构造基于数据重要度的 mass 函数

装备在其全寿命周期各阶段会产生各种各样的维修性试验数据, 根据数据的来源、产生阶段和产生层级的不同, 可以对多源数据进行不同分类, 比如从来源维度上可将维修性数据分为: 专家经验数据、仿真试验数据、历史试验数据、相似装备试验数据和不同环境试验数据。易知数据的来源不同, 其反映装备维修性真实水平的能力不同, 在数据融合时的重要程度显然也是不同的, 数据越重要, 对应数据源的融合权重应越大, 即 mass 函数值应越大。借助专家经验采用 1~9 标度法对各源数据的重要度进行打分, 构造基于数据重要度的 mass 函数, 假设邀请 H 名专家分别对 n 种维修性数据 $A_1, A_2, \cdots, A_n$ 的重要度进行打分, 如表 1 所示。

表 1 数据重要度的专家打分情况
Tab. 1 Expert scoring of data importance

多源数据	专家 1	专家 2	...	专家 H
A_1	y_{11}	y_{12}	$\cdots$	y_{1H}
A_2	y_{21}	y_{22}	$\cdots$	y_{2H}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_n	y_{n1}	y_{n2}	$\cdots$	y_{nH}

根据打分结果, 构造证据 E_3 的 mass 函数为:

$$E_3 = \{m_3(A_i) = \frac{\sum_{h=1}^H y_{ih}}{\sum_{i=1}^n \sum_{h=1}^H y_{ih}}\} \tag{11}$$

式中: $h = 1, 2, \cdots, H; i = 1, 2, \cdots, n$ 。

3 多源冲突证据数据融合模型

在充分挖掘维修性数据信息、构造 mass 函数的基础上, 为有效融合多源冲突证据数据, 综合考虑证据间的关联性和证据本身的不确定性两方面对冲突证据进行修正, 建立多源冲突证据数据融合模型。引入夹角余弦度量证据间的冲突程度, 借助余弦相似度确定证据的支持度加权因子; 引入信息熵确定证据的不确定度加权因子, 而后结合支持度加权因子和不确定度加权因子共同修正证据得到加权平均证据, 并利用 DS 组合规则融合 $n-1$ 次, 得到维修性多源数据的加权系数, 进而实现维修性多源冲突数据的有效融合。维修性多源冲突证据数据融合流程如图 1 所示。

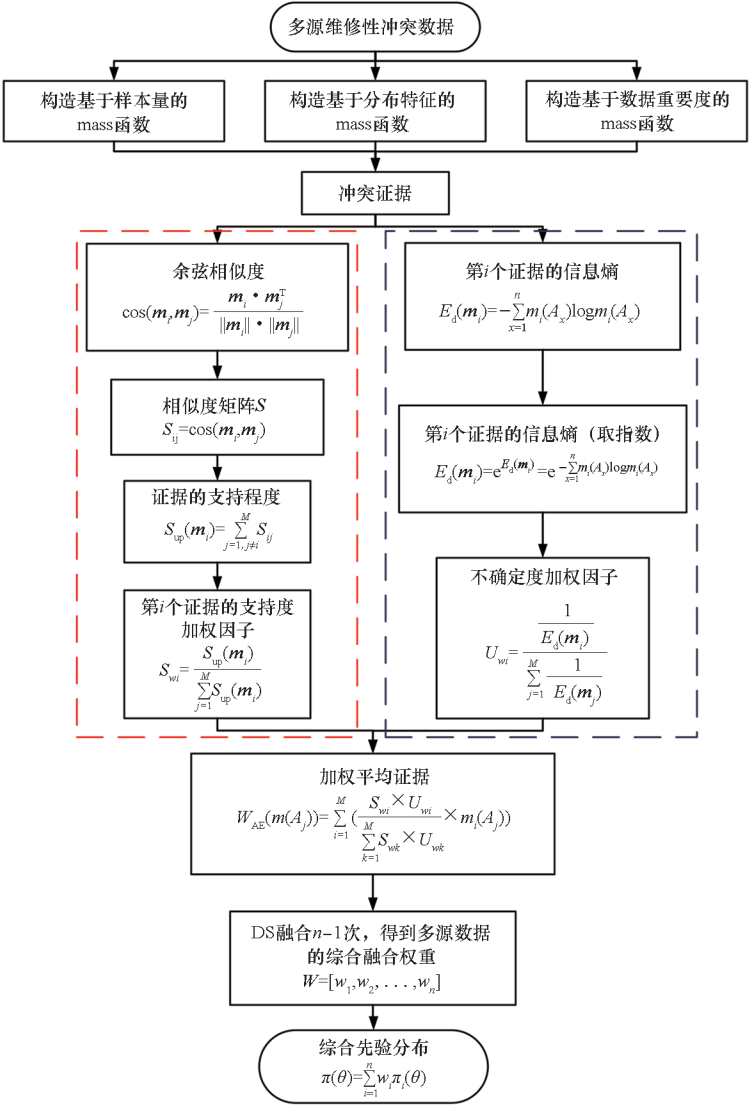


图 1 维修性多源冲突证据数据融合流程图

Fig. 1 Data fusion flow chart of maintainability multi-source conflict evidence

3.1 支持度加权因子

Step 1: 借助夹角余弦表示证据 m_i ($i = 1, 2, \dots, M$) 和 m_j ($j = 1, 2, \dots, M$) 之间的相似度, 则:

$$\cos(\mathbf{m}_i, \mathbf{m}_j) = \frac{\mathbf{m}_i \cdot \mathbf{m}_j^T}{\|\mathbf{m}_i\| \cdot \|\mathbf{m}_j\|} \tag{12}$$

式中, $\|\mathbf{m}_i\| = \sqrt{\mathbf{m}_i \cdot \mathbf{m}_i^T}$, $\|\mathbf{m}_j\| = \sqrt{\mathbf{m}_j \cdot \mathbf{m}_j^T}$ 。

Step 2: 将证据 m_i 和 m_j 间的余弦相似度记为 S_{ij} , 建立相似度矩阵 S :

$$S = \begin{bmatrix} 1 & S_{12} & \cdots & S_{1M} \\ S_{21} & 1 & \cdots & S_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ S_{M1} & S_{M2} & \cdots & 1 \end{bmatrix} \tag{13}$$

式中: S 为对称矩阵, $S_{ij} = S_{ji}$, $0 \leq S_{ij} \leq 1$, S_{ij} 越大, 表示证据 m_i 和 m_j 相互支持的程度越高。

Step 3: 计算证据 m_i 的支持度, 即

$$S_{up}(\mathbf{m}_i) = \sum_{j=1, j \neq i}^M S_{ij} \quad i, j = 1, 2, \dots, M \tag{14}$$

则证据 m_i 的支持度加权因子为:

$$S_{wi} = \frac{S_{up}(\mathbf{m}_i)}{\sum_{j=1}^M S_{up}(\mathbf{m}_j)} \quad i, j = 1, 2, \dots, M \tag{15}$$

3.2 不确定度加权因子

Step 1: 引入信息熵^[26-28] 度量证据的不确定程度, 设 $\mathbf{m}_i = [m_i(A_1), m_i(A_2), \dots, m_i(A_n)]$ 为第 i 个证据 E_i 的 mass 函数值, $i = 1, 2, \dots, M$, A_x ($x = 1, 2, \dots, n$) 为识别框架 Θ 下的第 x 个命题, 则证据 E_i 的 Shannon 熵为:

$$E_d(\mathbf{m}_i) = - \sum_{x=1}^n m_i(A_x) \log m_i(A_x) \tag{16}$$

式中: A_x 为单元素命题; $E_d(\mathbf{m}_i)$ 的值越大, 表示证据的不确定度越大, 包含的信息量越多, 在 DS 组

合中的权重就越小。

Step 2: 为避免由于 $E_d(\boldsymbol{m}_i) = 0$ 的出现而使得 DS 组合中证据 E_i 的权重为 0, 对 $E_d(\boldsymbol{m}_i)$ 取指数, 则:

$$E_d(\boldsymbol{m}_i) = e^{E_d(\boldsymbol{m}_i)} = e^{-\sum_{x=1}^n m_i(A_x) \log m_i(A_x)} \quad (17)$$

Step 3: 计算证据 m_i 的不确定度加权因子

$$U_{wi} = \frac{\frac{1}{E_d(\boldsymbol{m}_i)}}{\sum_{j=1}^M \frac{1}{E_d(\boldsymbol{m}_j)}} \quad i, j = 1, 2, \cdots, M \quad (18)$$

3.3 多源数据的加权融合

Step 1: 结合支持度加权因子 S_{wi} 和不确定度加权因子 U_{wi} 加权融合证据得到加权平均证据 $W_{AE}(m(A_j))$:

$$W_{AE}(m(A_j)) = \sum_{i=1}^M \left(\frac{S_{wi} \times U_{wi}}{\sum_{k=1}^M S_{wk} \times U_{wk}} \times m_i(A_j) \right) \quad (19)$$

Step 2: 通过 DS 组合规则对 $W_{AE}(m(A_j))$ 融合 $n-1$ 次, 得到多源维修性先验数据的综合融合权重 $\boldsymbol{W} = [w_1, w_2, \cdots, w_n]$, 则融合后的综合先验分布 $\pi(\theta)$ 为:

$$\pi(\theta) = \sum_{i=1}^n w_i \pi_i(\theta) \quad (20)$$

4 案例分析

4.1 案例 1

为验证所提改进 DS(improved dempster-shafer, IDS) 证据融合方法的可行性, 以文献[29]中的四种典型高冲突证据悖论为数值案例, 分别应用

IDS 和当前多种 DS 改进方法进行证据融合并对比分析融合结果。由于本文引入了信息熵, 已知当某一命题的 BPA 为 0 时, 取对数计算的结果无意义, 为避免这种情况, 使用 1×10^{-12} 这种很小的数替代 0, 证据融合的结果不变^[30]。四种典型高冲突证据的数值案例分别为:

1) 完全冲突悖论。

$$\begin{aligned} m_1: m_1(A) &= 1, m_1(B) = 0, m_1(C) = 0 \\ m_2: m_2(A) &= 0, m_2(B) = 1, m_2(C) = 0 \\ m_3: m_3(A) &= 0.9, m_3(B) = 0.1, m_3(C) = 0 \end{aligned}$$

2) 0 信任悖论。

$$\begin{aligned} m_1: m_1(A) &= 0.5, m_1(B) = 0.2, m_1(C) = 0.3 \\ m_2: m_2(A) &= 0, m_2(B) = 0.9, m_2(C) = 0.1 \\ m_3: m_3(A) &= 0.55, m_3(B) = 0.1, m_3(C) = 0.35 \end{aligned}$$

3) 1 信任悖论。

$$\begin{aligned} m_1: m_1(A) &= 0.9, m_1(B) = 0.1, m_1(C) = 0 \\ m_2: m_2(A) &= 0, m_2(B) = 0.1, m_2(C) = 0.9 \end{aligned}$$

4) 高冲突悖论。

$$\begin{aligned} m_1: m_1(A) &= 0.3, m_1(B) = 0.2 \\ m_1(C) &= 0.1, m_1(\Theta) = 0.4 \\ m_2: m_2(A) &= 0, m_2(B) = 0.9 \\ m_2(C) &= 0.1, m_2(\Theta) = 0 \\ m_3: m_3(A) &= 0.6, m_3(B) = 0.1 \\ m_3(C) &= 0.1, m_3(\Theta) = 0.2 \\ m_4: m_4(A) &= 0.7, m_4(B) = 0.1 \\ m_4(C) &= 0.1, m_4(\Theta) = 0.1 \\ m_5: m_5(A) &= 0.7, m_5(B) = 0.1 \\ m_5(C) &= 0.1, m_5(\Theta) = 0.1 \end{aligned}$$

四种典型高冲突证据融合结果如表 2 所示, 相对应的直方图如图 2~5 所示。

表 2 四种典型高冲突证据融合结果
Tab. 2 Fusion results of four typical high conflict evidence

悖论类型	融合方法	命题				目标
		A	B	C	Θ	
完全冲突悖论	Yager ^[10]	0	0	0	1	Θ
	Murphy ^[16]	0.837 5	0.162 5	0	0	A
	Chen 等 ^[18]	0.998 6	0.001 4	0	0	A
	应文健等 ^[20]	0.955 4	0.044 6	0	0	A
	Xiao ^[21]	0.998 6	0.001 4	0	0	A
	Deng 等 ^[22]	0.998 6	0.001 4	0	0	A
	IDS	0.998 6	0.001 4	0	0	A

续表

悖论类型	融合方法	命题				目标
		<i>A</i>	<i>B</i>	<i>C</i>	Θ	
0 信任悖论	DS ^[8]	0	0.631 6	0.368 4	0	<i>B</i>
	Yager ^[10]	0	0.018 0	0.010 5	0.989 5	Θ
	Murphy ^[16]	0.350 0	0.522 4	0.127 6	0	<i>B</i>
	Chen 等 ^[18]	0.619 5	0.201 0	0.179 5	0	<i>A</i>
	应文健等 ^[20]	0.731 9	0.067 8	0.200 3	0	<i>A</i>
	Xiao ^[21]	0.731 9	0.079 0	0.189 1	0	<i>A</i>
	Deng 等 ^[22]	0.607 5	0.233 1	0.159 3	0	<i>A</i>
	IDS	0.741 9	0.069 4	0.188 6	0	<i>A</i>
1 信任悖论	DS ^[8]	0	1	0	0	<i>B</i>
	Yager ^[10]	0	0.010 0	0	1	Θ
	Murphy ^[16]	0.488 0	0.024 1	0.488 0	0	<i>A/C</i>
	Chen 等 ^[18]	0.488 0	0.024 1	0.488 0	0	<i>A/C</i>
	应文健等 ^[20]	0.488 0	0.024 1	0.488 0	0	<i>A/C</i>
	Xiao ^[21]	0.488 0	0.024 1	0.488 0	0	<i>A/C</i>
	Deng 等 ^[22]	0.488 0	0.024 1	0.488 0	0	<i>A/C</i>
	IDS	0.488 0	0.024 1	0.488 0	0	<i>A/C</i>
高冲突悖论	DS ^[8]	0	0.915 3	0.084 7	0	<i>B</i>
	Yager ^[10]	0	0.000 2	0	1	Θ
	Murphy ^[16]	0.838 9	0.150 2	0.009 9	0.001 0	<i>A</i>
	Chen 等 ^[18]	0.962 3	0.028 9	0.007 8	0.001 0	<i>A</i>
	应文健等 ^[20]	0.961 0	0.027 0	0.010 3	0.001 7	<i>A</i>
	Xiao ^[21]	0.961 0	0.026 8	0.010 5	0.001 7	<i>A</i>
	Deng 等 ^[22]	0.951 7	0.037 5	0.010 0	0.000 8	<i>A</i>
	IDS	0.962 3	0.025 0	0.010 8	0.001 9	<i>A</i>

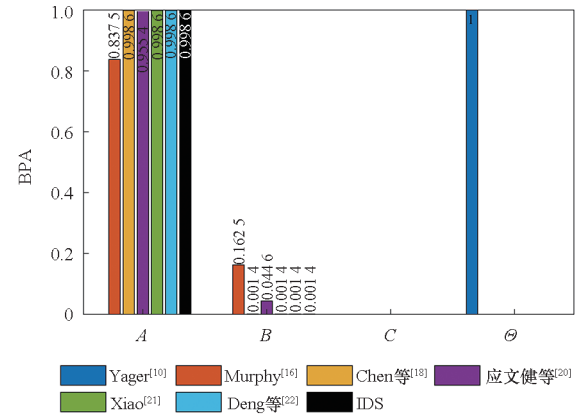


图 2 完全冲突悖论
Fig. 2 Completely conflicting paradox

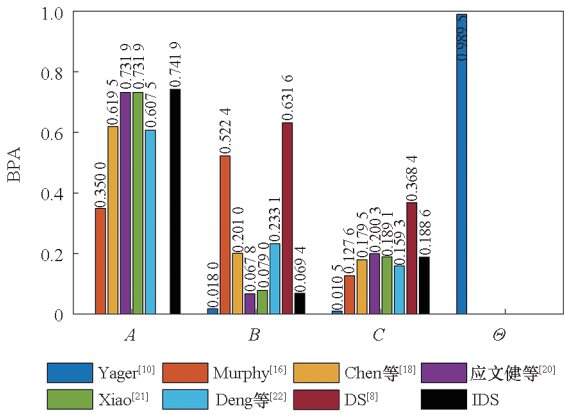


图 3 0 信任悖论
Fig. 3 0 Trust paradox

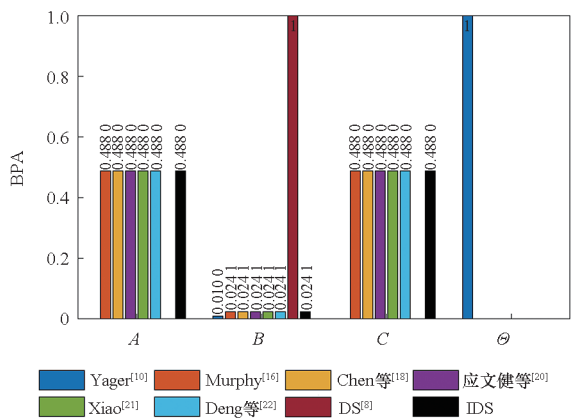


图 4 1 信任悖论

Fig. 4 1 Trust paradox

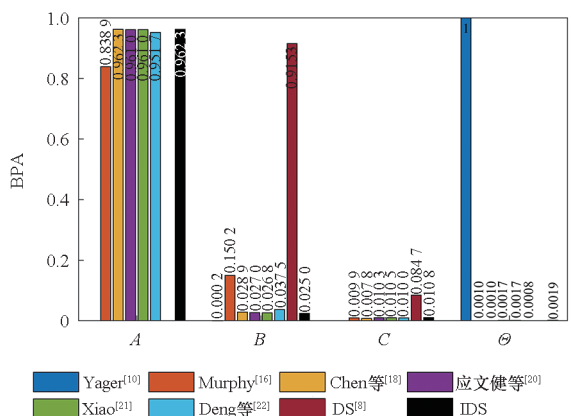


图 5 高冲突悖论

Fig. 5 Highly conflicting paradox

在完全冲突悖论中,已知冲突系数 $k=1$,此时 Dempster 组合规则失效。由表 2 和图 2 可知, Yager^[10] 将 BPA 完全赋予了识别框架 Θ , $m(\Theta)=1$,证据融合后反而增加了命题的不确定性,无法正确识别目标命题 A。其余方法均能正确识别目标命题 A,其中, Murphy^[16] 和 应文健等^[20] 的识别精度略低,分别为 $m(A)=0.8375$ 和 $m(A)=0.9554$, Chen 等^[18]、Xiao^[21]、Deng 等^[22] 和 IDS 的识别精度较高,均为 $m(A)=0.9986$ 。

在 0 信任悖论中,已知证据 m_1 和 m_3 为相对一致的证据,均支持命题 A,证据 m_2 为冲突证据,支持命题 B。由表 2 和图 3 可知, DS^[8] 法和 Murphy^[16] 分别以 $m(B)=0.6316$ 和 $m(B)=0.5224$ 支持命题 B,但此时命题 B 并不是正确的目标。Yager^[10] 赋予了识别框架 $m(\Theta)=0.9895$,表示命题是高度不确定的,无法正确识别目标命题。其余方法均能正确识别目标命题 A,融合结果与相对一致的证据保持一致,且以 IDS 的识别精度最高, $m(A)=0.7419$ 。

在 1 信任悖论中,已知正确的识别目标应为

命题 A 或 C,由表 2 和图 4 可知, DS^[8] 法完全支持命题 B, $m(B)=1$,融合结果显然与事实相悖。Yager^[10] 赋予了识别框架 $m(\Theta)=1$,也无法正确识别目标命题。其余方法均以相同概率 $m(A)=0.4880$ 和 $m(C)=0.4880$ 支持命题 A 和 C,能够正确识别目标命题。

在高冲突悖论中,已知证据 m_1 、 m_3 、 m_4 和 m_5 为相对一致的证据,均支持命题 A,证据 m_2 支持命题 B,由表 2 和图 5 可知, DS^[8] 法以较大概率 $m(B)=0.9153$ 支持命题 B,错误的识别命题 B 为目标命题。Yager^[10] 赋予了识别框架 $m(\Theta)=1$,也无法正确识别目标命题 A。其余方法均能正确识别目标命题 A,其中, Murphy^[16] 的识别精度略低, $m(A)=0.8389$, Chen 等^[18] 和 IDS 的识别精度最高, $m(A)=0.9623$ 。

四种典型高冲突证据下目标命题 A 的融合结果如图 6 所示。

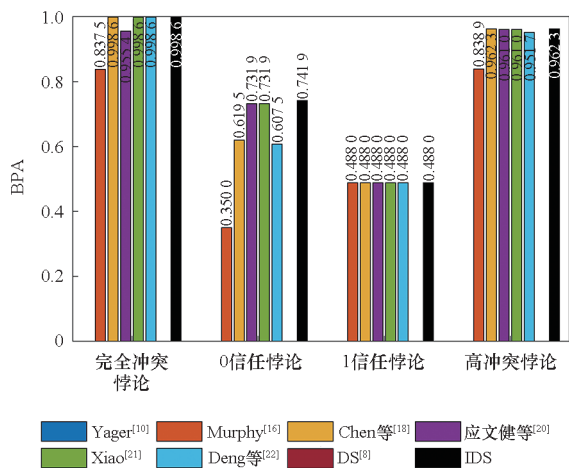


图 6 命题 A 的 BPA

Fig. 6 BPA of proposition A

综合上述分析并结合图 6 可知, DS^[8] 法和 Yager^[16] 均处理不了这四种典型冲突悖论,其中, DS^[8] 法在完全冲突悖论中失效,在其余三种冲突悖论中均错误地选择了命题 B。Yager^[10] 在四种冲突悖论中,基本均将证据冲突的 BPA 赋给了识别框架 Θ ,不仅不能正确识别目标命题,反而进一步增加了命题的不确定性。Murphy^[16] 除在 0 信任悖论中错误地选择了命题 B 外,在其余三种冲突悖论中均能正确识别目标命题,其余 Chen 等^[18]、应文健等^[20]、Xiao^[21]、Deng 等^[22] 和 IDS 在这四种冲突悖论中均能正确识别目标命题。此外,除了由于本文在 1 冲突悖论案例中证据 BPA 设置的特殊性,使得 Murphy^[16] 等六种方法的融合结果相同外,在其余三种冲突悖论中, Murphy^[16] 由于只采用平均加权的方式修正证据,

既没有考虑证据间的相关性,也没有考虑证据本身的不确定性,其融合精度最低,IDS 法的融合精度最高,验证了所提方法的可行性和有效性。

4.2 案例 2

结合装甲车辆小子样维修性试验,选取典型维修性指标平均修复时间 (mean time to repair, MTTR) 进行分析,融合维修性多源冲突数据确定先验分布,验证所提方法的可行性。以历史试验数据 A_1 、相似装备试验数据 A_2 和仿真试验数据 A_3 作为三种先验数据,分别基于样本量、分布特征和数据重要度构造 mass 函数,假定三种先验数据的样本量分别为 10、15 和 30,MTTR 对数均值的分布分别为 $\pi_1(\theta) = N(3.607\ 7, 0.043\ 4^2)$ 、 $\pi_2(\theta) = N(3.551\ 1, 0.046\ 6^2)$ 和 $\pi_3(\theta) = N(3.623\ 5, 0.041\ 3^2)$,现场分布为 $\pi_0(\theta) = N(3.587\ 5, 0.043\ 2^2)$,具体计算过程和专家打分情况不在此详细阐述,数据处理结果如表 3 所示。

表 3 证据的 BPA 值
Tab.3 BPA values of evidence

多源数据	E_1	E_2	E_3
A_1	0.181 8	0.564 2	0.625 0
A_2	0.272 7	0.423 8	0.237 8
A_3	0.545 5	0.012 0	0.137 2

分别采用 IDS 法和多种 DS 改进方法融合证据 E_1 、 E_2 和 E_3 ,得到多源数据的融合权重如表 4 所示,对应直方图如图 7 所示。

表 4 多源数据的融合权重
Tab.4 Fusion weight of multi-source data

方法	多源数据		
	A_1	A_2	A_3
DS ^[8]	0.704 3	0.293 7	0.002 0
Yager ^[10]	0.064 1	0.027 5	0.972 5
Murphy ^[16]	0.662 3	0.311 5	0.026 2
Chen 等 ^[18]	0.697 8	0.284 0	0.018 1
应文健等 ^[20]	0.358 1	0.427 0	0.215 0
Xiao ^[21]	0.673 1	0.299 0	0.027 8
Deng 等 ^[22]	0.688 7	0.291 3	0.020 0
IDS	0.718 7	0.196 0	0.085 3

由表 4 和图 7 可知,IDS 法的融合精度最高,赋予先验数据 A_1 的融合权重最大, $w_1 = 0.718\ 7$,

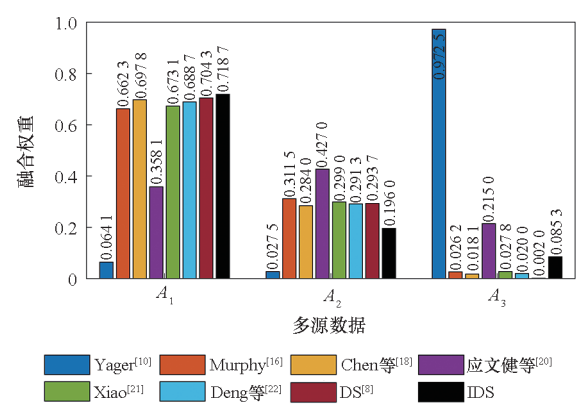


图 7 多源数据的融合权重

Fig.7 Fusion weight of multi-source data

这与证据相对一致支持数据 A_1 的事实相同,且进一步增大了数据 A_1 的融合权重,表明所提方法能更好地凸显与现场分布更接近的数据源数据的重要性,有助于提高先验分布的融合精度。三种先验数据的融合权重 $W = (0.718\ 7, 0.196\ 0, 0.085\ 3)$,则融合后的综合先验分布为 $\pi(\theta) = 0.718\ 7\pi_1(\theta) + 0.196\ 0\pi_2(\theta) + 0.085\ 3\pi_3(\theta) \approx N(3.603\ 9, 0.048\ 6^2)$,为更直观地观察融合效果,分别作 $\pi_1(\theta)$ 、 $\pi_2(\theta)$ 、 $\pi_3(\theta)$ 、 $\pi_0(\theta)$ 和 $\pi(\theta)$ 的曲线图,如图 8 所示, $\pi_1(\theta)$ 、 $\pi_2(\theta)$ 、 $\pi_3(\theta)$ 、 $\pi(\theta)$ 与 $\pi_0(\theta)$ 均值差的绝对值如表 5 所示。

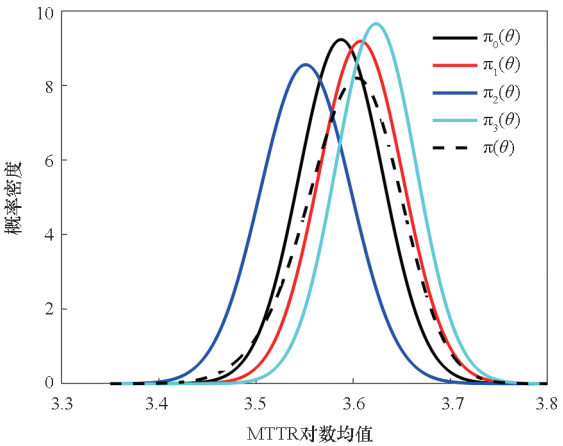


图 8 MTTR 对数均值的分布曲线

Fig.8 MTTR logarithmic mean distribution curve

表 5 均值差的绝对值
Tab.5 Absolute value of mean difference

分布	均值	与 $\pi_0(\theta)$ 均值差的绝对值
$\pi_1(\theta)$	3.607 7	0.020 2
$\pi_2(\theta)$	3.551 1	0.036 4
$\pi_3(\theta)$	3.623 5	0.036 0
$\pi(\theta)$	3.603 9	0.016 4

从图 8 和表 5 可以看出,融合后的先验分布 $\pi(\theta)$ 与各单独数据源的先验分布相比,更贴近于现场分布 $\pi_0(\theta)$,进一步验证了所提方法能较好地融合多源先验分布。

5 结论

为提高维修性多源数据冲突条件下先验分布的融合精度,本文提出了维修性多源冲突证据数据融合的先验分布确定方法。通过充分挖掘、提取多源数据的特征信息,分别构造了基于样本量、分布特征和数据重要度的证据 mass 函数,然后从证据间的关联性和证据本身的不确定性两方面综合考虑,引入了夹角余弦和信息熵分别度量证据间的冲突程度和证据的不确定度,并结合证据的支持度和不确定度修正冲突证据,建立了多源冲突证据数据融合模型,实现了多源数据的有效融合并得到了综合先验分布。两个案例结果表明,所提方法有效可行。

参考文献 (References)

[1] 吴振亚,郝建平,陈汉青,等. 装备平均修复时间验证方法研究综述[J]. 系统工程与电子技术, 2020, 42(12): 2931–2943.
WU Z Y, HAO J P, CHEN H Q, et al. Review on research of equipment mean time to repair demonstration methods[J]. Systems Engineering and Electronics, 2020, 42(12): 2931–2943. (in Chinese)

[2] LI Z J, LI L F. Mathematical statistics algorithm in the bending performance test of corroded reinforced concrete beams under fatigue load [J]. Applied Mathematics and Nonlinear Sciences, 2022, 7(1): 695–702.

[3] MONEIM A F A, GHAZY M, HASSNIEN A. Bayesian estimation of parameters of reliability and maintainability of a component under imperfect repair and maintenance [J]. Arabian Journal for Science and Engineering, 2019, 44(3): 2497–2502.

[4] CHATTHA M I, MATSUO K, SUGIKI N. Risk analysis of pedestrian and bicycle accidents at intersections in Toyohashi city using empirical Bayes method[J]. Journal of the Eastern Asia Society for Transportation Studies, 2022, 14: 2093–2112.

[5] HAN L T, GU L, GONG C, et al. Motion mode recognition in multi-storey buildings based on the naive Bayes method[J]. International Journal of Sensor Networks, 2022, 38(3): 166–176.

[6] KHARIN A Y. The influence of distortions of the prior distribution on the characteristics of the sequential test for composite hypotheses[J]. Journal of Mathematical Sciences, 2022, 267(1): 1–5.

[7] DEMPSTER A P. Upper and lower probabilities induced by a multivalued mapping [J]. The Annals of Mathematical

Statistics, 1967, 38(2): 325–339.

[8] SHAFER G. A mathematical theory of evidence [M]. Princeton: Princeton University Press, 1976.

[9] ZHAO Y X, JIA R F, SHI P. A novel combination method for conflicting evidence based on inconsistent measurements [J]. Information Sciences, 2016, 367/368: 125–142.

[10] YAGER R R. On the Dempster-Shafer framework and new combination rules[J]. Information Sciences, 1987, 41(2): 93–137.

[11] SMETS P. Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem [J]. International Journal of Approximate Reasoning, 1993, 9(1): 1–35.

[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.
SUN Q, YE X Q, GU W K. A new combination rules of evidence theory[J]. Acta Electronica Sinica, 2000, 28(8): 117–119. (in Chinese)

[13] LIN Z, XIE J Y. Research on improved evidence theory based on multi-sensor information fusion [J]. Scientific Reports, 2021, 11: 9267.

[14] 李弼程, 王波, 魏俊, 等. 一种有效的证据理论合成公式[J]. 数据采集与处理, 2002, 17(1): 33–36.
LI B C, WANG B, WEI J, et al. An efficient combination rule of evidence theory [J]. Journal of Data Acquisition & Processing, 2002, 17(1): 33–36. (in Chinese)

[15] LI Y B, CHEN J, YE F, et al. The improvement of DS evidence theory and its application in IR/MMW target recognition [J]. Journal of Sensors, 2016, 2016(1): 1903792.

[16] MURPHY C K. Combining belief functions when evidence conflicts [J]. Decision Support Systems, 2000, 29(1): 1–9.

[17] DENG Y, SHI W K, ZHU Z F, et al. Combining belief functions based on distance of evidence[J]. Decision Support Systems, 2004, 38(3): 489–493.

[18] CHEN B C, TAO X, YANG M R, et al. A saliency map fusion method based on weighted DS evidence theory [J]. IEEE Access, 2018, 6: 27346–27355.

[19] YE F, CHEN J, LI Y B. Improvement of DS evidence theory for multi-sensor conflicting information[J]. Symmetry, 2017, 9(5): 69.

[20] 应文健, 程雨森, 王旋, 等. 基于研制阶段数据融合的舰炮制导弹药测试性评估方法[J/OL]. 系统工程与电子技术, 2024, 46(8): 2730–2737.
YING W J, CHENG Y S, WANG X, et al. Testability evaluation method of naval gun guided ammunition based on data fusion in development stage [J/OL]. Systems Engineering and Electronics, 2024, 46(8): 2730–2737. (in Chinese)

[21] XIAO F Y. Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy [J]. Information Fusion, 2019, 46: 23–32.

[22] DENG Z, WANG J Y. Measuring total uncertainty in evidence theory [J]. International Journal of Intelligent

Systems, 2021, 36(4): 1721 – 1745.

[23] YUAN K J, XIAO F Y, FEI L G, et al. Conflict management based on belief function entropy in sensor fusion [J]. SpringerPlus, 2016, 5(1): 638.

[24] YU L L, ZHAO Y C. A bootstrap method for a multiple-imputation variance estimator in survey sampling[J]. Stats, 2022, 5(4): 1231 – 1241.

[25] LEVADA A L M. On the numerical computation of Fisher-Rao based distances and KL-divergences between Gaussian random fields[J]. Journal of Computational Science, 2022, 62: 101714.

[26] KHAN M N, ANWAR S. Paradox elimination in dempster-shafer combination rule with novel entropy function: application in decision-level multi-sensor fusion[J]. Sensors, 2019, 19(21): 4810.

[27] SHANNON C E. A mathematical theory of communication[J]. ACM SIGMOBILE Mobile Computing and Communications Review, 2001, 5(1): 3 – 55.

[28] ZHOU J, LI J. IE-AK: a novel adaptive sampling strategy based on information entropy for Kriging in metamodel-based reliability analysis [J]. Reliability Engineering & System Safety, 2023, 229: 108824.

[29] FAN X J, GUO Y J, JU Y Y, et al. Multisensor fusion method based on the belief entropy and DS evidence theory[J]. Journal of Sensors, 2020, 2020: 7917512.

[30] 郭效芝, 辛小龙. Fuzzy 集的偏熵与关联熵[J]. 模糊系统与数学, 2005, 19(2): 97 – 102.

GUO X Z, XIN X L. Partial entropy and relative entropy of Fuzzy sets[J]. Fuzzy System & Mathematics, 2005, 19(2): 97 – 102. (in Chinese)