

一种基于 ISODATA 聚类和改进相似度的证据推理方法

李新德¹ 王丰羽¹

摘 要 针对智能信息处理中 Dempster 组合规则不能处理高度冲突的问题, 从内、外证据不确定性分析的角度深入揭示了证据冲突产生的原因, 即证据的冲突性不仅仅根源于证据间的矛盾, 也与证据自身的不确定性密切相关, 提出了一种同时考虑证据自冲突和外部冲突的相似性测度, 然后利用新测度计算证据的众信度, 对证据源进行修正; 与此同时, 根据原始证据间的聚类特性, 利用迭代自组织数据分析技术 (Iterative selforganizing data analysis techniques algorithm, ISODATA) 聚类方法进行聚类, 然后利用 Dempster 组合规则合成每一聚类中所有证据为证据代表, 并综合众信度和证据在该聚类的频度计算可靠度, 最后, 利用统一组合规则合成证据代表. 并通过大量的算例, 同其他方法和自身改进前后进行深入比较, 优势比较明显, 有效地解决了冲突证据合成出现的问题.

关键词 证据推理, 冲突, 聚类, 相似测度, 组合规则

引用格式 李新德, 王丰羽. 一种基于 ISODATA 聚类和改进相似度的证据推理方法. 自动化学报, 2015, 41(3): 575–590

DOI 10.16383/j.aas.2015.c140543

A Method of Evidence Reasoning Based on ISODATA Clustering and Improved Similarity Measure

LI Xin-De¹ WANG Feng-Yu¹

Abstract For the problem that the highly conflictive evidences can not be processed by Dempster rule in intelligent information processing, the reason of conflict occurrence is further revealed by analysis of uncertainty among evidences and evidences themselves. A similarity measure by considering self-conflict and external conflict of evidence together is proposed and used to compute the commonality of evidence. All initial evidences can be revised according to the commonality. At the same time, according to the clustering characteristic of initial evidences, they are clustered by applying ISODATA (Iterative selforganizing data analysis techniques algorithm), and then, all evidences in each cluster are combined as a representative of evidences by applying Dempster rule. The reliability of the representative of evidence is computed by considering the credibility and frequency of occurrence of the evidence in a cluster together. Finally, these representatives of evidences are combined by using the unified combination rule. Several numerical examples are given to deeply compare this method with others and itself before improving its similarity measure. The result shows that the new method has a distinct advantage over others and can effectively solve the problem of combination of conflictive evidences.

Key words Evidence reasoning, conflict, clustering, similarity measure, combination rule

Citation Li Xin-De, Wang Feng-Yu. A method of evidence reasoning based on ISODATA clustering and improved similarity measure. *Acta Automatica Sinica*, 2015, 41(3): 575–590

收稿日期 2014-07-18 录用日期 2014-09-27
Manuscript received July 18, 2014; accepted September 27, 2014

国家自然科学基金 (60804063, 61175091), 江苏省“青蓝工程”资助计划, 航空基金 (20140169002), 江苏省“六大高峰人才”资助计划, 江苏高校优势学科建设工程资助项目资助

Supported by National Natural Science Foundation of China (60804063, 61175091), Qing Lan Project of Jiangsu Province, Aeronautical Science Foundation of China (20140169002), Six Major Top-talent Plan of Jiangsu Province, and A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions

本文责任编辑 王聪

Recommended by Associate Editor WANG Cong

1. 东南大学自动化学院复杂工程测量与控制教育部重点实验室 南京 210096

1. Key Laboratory of Measurement and Control of Complex Systems of Engineering of Ministry of Education, School of Automation, Southeast University, Nanjing 210096

证据推理 (Dempster-shafer theory, 又名 D-S 证据理论) 最早是由 Dempster 于 1967 年提出, 后经 Shafer 于 1976 年^[1] 扩展的一种处理多值映射问题的信息处理方法. 由于它具有能够处理不确定性的能力以及在工程上表现出来的实用性能, 近年来在不确定推理、多传感器信息融合、模式识别、图像处理、故障诊断、机器人、多目标决策等领域得到了广泛的应用. 在证据理论中, 概率赋值的更新是通过 Dempster 组合规则实现的, 该规则是证据理论的核心基石之一, 尽管其形式比较简单, 适合机器实现, 但组合的归一化过程中可能会导致推理结果出现反直觉现象. 自从 Zadeh^[2] 发现这个问题以来, 该问题受到领域内专家和学者的广泛关注. Haenni^[3] 从哲学的角度分析出现这一问题的原因: 在模型 Ξ 上使

用方法 Ψ , 从而获得了一个不合理的结论 Σ . 于是很多专家学者分别从模型和方法进行分析与改进证据推理. 也有少数人在模型和方法上同步进行修改, 但多数研究是基于完备辨识框架的.

证据的融合方式一般来说, 分两种方式, 一种是批量同步融合, 另一种是序贯融合. 批量同步融合的前提是所有的证据源及其约束条件融合前全部已知, 可以实现并行同步一次性融合, 并给出最终融合决策结果, 由于融合过程一步到位, 不存在融合顺序的歧义, 于是最终的结果最优, 且唯一; 而序贯融合的前提是证据源及其约束条件是串行给定, 序贯依次融合, 每次融合得出一个决策结果, 由于融合结果受后给定证据源与前合成结果之间的相似程度 (或者冲突程度), 以及证据源的融合次序的影响比较大, 因此最终结果往往是次优的.

本文在已有研究的基础上, 考虑批量式同步融合情况, 分析了复杂条件下证据产生冲突的原因, 假设辨识框架完备或者非完备, 并试图对模型 Ξ 和方法 Ψ 同时修正, 提出了一种分布式冲突证据组合推理方法. 其创新性在于考虑以往对冲突的描述依赖证据间距离定义的不足, 即截止目前, 全部距离的定义只是根据两证据源对应焦元之间的距离进行描述, 而没有考虑证据源自身各个焦元之间的赋值大小顺序, 这里称为“自冲突或者自矛盾”. 然后, 根据证据源之间的相似性约束, 利用迭代自组织数据分析技术 (Iterative selforganizing data analysis techniques algorithm, ISODATA)^[4] 进行证据聚类, 并综合众信度和证据源类内频率计算证据代表可靠度, 利用不同的组合方法对同类内部及不同类之间的证据进行合成.

1 D-S 证据理论及冲突分析

1.1 相同鉴别框架下

设 Θ 为一个有限、互斥且完备的假设空间, 称为鉴别框架, 其幂集空间为 2^Θ . 在该鉴别框架下的基本概率赋值 (Basic probability assignment, BPA) 是一个函数 $m: 2^\Theta \mapsto [0, 1]$ 满足 $m(\phi) = 0$, $\sum_{\theta \in 2^\Theta} m(\theta) = 1$, 其中 ϕ 为空集. 所有满足基本概率赋值 (BPA) $m(\theta) > 0$ 的 θ 被称为焦元 (或命题), 所有由 $m(\theta) > 0$ 组成命题 θ 的集合称为证据的核 Γ . 特别地, 当仅包含焦元 Θ 的 BPA 不为零时, $m(\Theta) = 1$. 根据文献 [1] 定义的 BPA, 对应的信任函数 (Bel), 似然函数 (Pl) 分别定义为

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (1)$$

$$Pl(A) = \sum_{B \cap A \neq \phi} m(B) \quad (2)$$

若 $m_i(\cdot)$ ($i = 1, 2, \dots, n$) 是同一鉴别框架 Θ 下 n 个相互独立的 BPA, 则 Dempster 合成规则为

$$m(A) = \frac{\sum_{\cap A_i^j = A} \prod_{i=1}^n m_i(A_i^j)}{1 - \sum_{\cap A_i^j = \phi} \prod_{i=1}^n m_i(A_i^j)} \quad (3)$$

其中, $\sum_{\cap A_i^j = \phi} \prod_{i=1}^n m_i(A_i^j)$ 被称为冲突因子 κ .

1.2 不同鉴别框架下

设 $\Theta_1, \Theta_2, \dots, \Theta_n$ 为 n 个不同鉴别框架, 其幂集空间分别是 $2^{\Theta_1}, 2^{\Theta_2}, \dots, 2^{\Theta_n}$, 那么, 证据源可表示为

$$\begin{aligned} S_1 &= \{A_1^j | A_1^j \in 2^{\Theta_1}; m_{\Theta_1}(A_1^j) \geq 0, j=1, 2, \dots, |\Gamma_1|\} \\ S_2 &= \{A_2^k | A_2^k \in 2^{\Theta_2}; m_{\Theta_2}(A_2^k) \geq 0, k=1, 2, \dots, |\Gamma_2|\} \\ &\vdots \\ S_n &= \{A_n^l | A_n^l \in 2^{\Theta_n}; m_{\Theta_n}(A_n^l) \geq 0, l=1, 2, \dots, |\Gamma_n|\} \end{aligned}$$

其中, 各个鉴别框架下的概率赋值具有归一化特性, 即

$$\begin{cases} m(\phi) = 0 \\ \sum_{A_1^j \in \Gamma_1} m_{\Theta_1}(A_1^j) = 1, \forall A_1^j \in 2^{\Theta_1}, j=1, 2, \dots, |\Gamma_1| \\ \sum_{A_2^k \in \Gamma_2} m_{\Theta_2}(A_2^k) = 1, \forall A_2^k \in 2^{\Theta_2}, k=1, 2, \dots, |\Gamma_2| \\ \vdots \\ \sum_{A_n^l \in \Gamma_n} m_{\Theta_n}(A_n^l) = 1, \forall A_n^l \in 2^{\Theta_n}, l=1, 2, \dots, |\Gamma_n| \end{cases}$$

那么 n 个证据源合成后的幂集空间为 $2^\Theta = 2^{\Theta_1 \times \Theta_2 \times \dots \times \Theta_n}$, 其幂集空间中的子集在笛卡尔乘积 $2^{\Theta_1} \times 2^{\Theta_2} \times \dots \times 2^{\Theta_n}$ 上的基本概率赋值函数 $m(\cdot)$ 为 $m_{\Theta_n}(\cdot \cap \cdot \cap \dots \cap \cdot) = m_{\Theta_1}(\cdot) \times m_{\Theta_2}(\cdot) \times \dots \times m_{\Theta_n}(\cdot)$. 于是, n 个不同鉴别框架下的 Dempster 合成规则可表示为

$$m(A) = \frac{\sum_{\cap A_i^j = A} \prod_{i=1}^n m_{\Theta_i}(A_i^j)}{1 - \kappa} \quad (4)$$

其中, 冲突因子 $\kappa = \sum_{\cap A_i^j = \phi} \prod_{i=1}^n m_{\Theta_i}(A_i^j)$.

下面通过几个例子分析 Dempster 组合规则的反直觉 (冲突) 现象.

例 1 (早期 Zadeh 给出的一个冲突例子^[2]). 两个医生针对同一病人进行诊断, 认为病症可能是脑膜炎 (M)、脑震荡 (C)、脑肿瘤 (T) 中的一种. 辨识框架 Θ 可设为 $\{M, C, T\}$.

两位医生的诊断结果分别表示为

医生 1: $m_1(M) = 0.99, m_1(T) = 0.01$.

医生 2: $m_2(C) = 0.99, m_2(T) = 0.01$.

从上面的两位医生的诊断结果看, 两位都认为病人患脑肿瘤的可能性最低, 但分别认为是其他两种病症之一的可能性很高. 于是根据 Dempster 组合规则式 (3) 可以得到如下结果:

$$m_{1,2}(M) = 0, m_{1,2}(C) = 0, m_{1,2}(T) = 1$$

据此得出的结论是病人患脑肿瘤的机率是 100%. 这显然是反直觉 (冲突) 的结果. 因为两个医生都认为病人患脑肿瘤的可能性极低. 需要特别注意的是两个医生对其他两种病症可能性的判断几乎完全冲突.

本例就是证据冲突的极端例子, 它能很好地反映出在对高冲突证据之间利用 Dempster 规则进行组合时可能出现的冲突问题.

例 2. 两个医生在 Shafer 模型和鉴别框架 $\Theta = \{M, T\}$ 下:

1) 给定两个相同证据

医生 1: $m_1(M) = 0.99, m_1(T) = 0.01$.

医生 2: $m_2(M) = 0.99, m_2(T) = 0.01$.

由于两个医生对患者的诊断结果不完全肯定, 根据冲突因子 κ 的定义, $\kappa = 0.0198$, 尽管两个证据源是相同的, 但仍然存在冲突因子, 即 $\kappa \neq 0$, 意味着仍然存在冲突.

2) 给定如下两个相同证据

医生 1: $m_1(M) = 1, m_1(T) = 0$.

医生 2: $m_2(M) = 1, m_2(T) = 0$.

两个医生同时认为患者 100% 患了脑膜炎. 此时 $\kappa = 0$, 意味着两个证据源不存在冲突.

3) 给定如下两个证据

医生 1: $m_1(M) = 0, m_1(T) = 1$.

医生 2: $m_2(M) = 1, m_2(T) = 0$.

其中, 一个医生断定患者 100% 患上脑肿瘤, 另一则断定患者 100% 患上脑膜炎, 两者没有任何交集, 此时 $\kappa = 1$, 意味着两个证据源完全冲突.

对于 2) 和 3), 两者完全符合直觉, 但对于 1), 两个完全相同的证据源存在冲突, 不符合直觉.

下面分析一下冲突产生的原因: Martin 等在文献 [5] 给出了 3 种产生冲突的主要因素, 即信息源 (传感器) 不可靠、证据不可靠 (如 BPA 方法存在问题) 和鉴别框架的不完备性. 然而, 对于例 2 中 1) 产生冲突的原因, 文献 [5] 不能很好地揭示, 而这种冲突产生的本质根源在于证据自身对观点的持有具有不确定性, 从而导致存在矛盾和冲突.

于是, 在利用 D-S 组合规则对高冲突证据融合之前, 有必要关注冲突的成因、程度和减小冲突的方法. 自从 Haenni^[3] 给出分析和改进 Dempster 组合

规则的指导思想以来, 很多专家学者分别从模型和方法上进行分析改进证据推理. 也有少数人在模型和方法上同步进行修改, 但多数研究是基于完备辨识框架的.

1) 在规则方法改进

Lefevre 等^[6] 认为: 冲突也是一种信息, 冲突信息的完全遗弃必然造成信息的损失, 可以将冲突信息根据提出的统一信度函数组合规则, 将冲突加权嵌入组合规则中. 然而, Yager^[7] 的思想同 Lefevre 的正好相反, 他认为所有冲突均不能提供有用信息, 因此把证据冲突全部赋给未知项 $m(\Theta)$. Martin 等^[5] 考虑到证据冲突的全局分配和局部分配, 提出一种混合的证据合成方法, 称之为折扣比例冲突再分配 (Discounting proportional conflict redistribution, DPCR), 即在比例冲突再分配的基础上, 应用折扣处理 (Discounting procedure) 方法, 将部分证据冲突分配给部分未知项. Liu^[8] 认为导致证据冲突的原因不仅和冲突系数 k 有关, 还和 pignistic 概率距离有关, 两者一起度量证据体之间的冲突, 并根据导致冲突的两个因素取值大小, 分类讨论了证据冲突的情况. 于是王栋等^[9] 在文献 [8] 的基础上, 提出了一种新的冲突证据合成方法, 当证据之间的冲突小时, 合成结果近似于 D-S 证据合成公式, 当证据之间的冲突较大时, 冲突证据的合成结果更加合理. 邓勇等^[10] 对表征证据冲突的变量进行了研究, 提出了一种新的表征证据之间冲突的变量-关联系数, 基于偏熵和混合熵, 定义了关联系数, 该系数可以定量的表示证据之间的冲突. 当关联系数接近 1 时证据之间冲突很小, 当关联系数接近 0 时证据之间高度冲突. 权文等^[11] 从选择有效的冲突衡量参数的角度出发, 分析了悖论与证据冲突的关系, 给出了新的冲突衡量参数, 并基于此参数提出了一种基于局部冲突分配策略的证据推理组合规则. 何兵等^[12] 提出了一种修正的 DS 证据融合策略, 以解决证据冲突情况下的融合问题, 该方法综合利用了传统 D-S 证据理论方法所具有的收敛性好及加性融合方法所具有的可靠性高两方面的优点. Yang 等^[13] 通过权重和可靠性表示, 建立一个统一的证据推理规则 (Evidential reasoning (ER) rule) 来处理多个独立证据源, 并证明了当每个证据源是完全可靠时, Dempster 规则是 ER 规则的一种特例.

2) 在模型上进行改进

Smets^[14] 剖析了证据合成的本质 (合取与析取, 修改与更新, 静态与动态数据融合), 提出了基于非概率模型的转移信任模型 (Transferable belief model, TBM), 并同基于概率模型的 Dempster 模型进行深入的比较, 分析各自归一化及合取运算的适用范围, 指出 TBM 中冲突的存在并不是真正的

问题,并通过比较消除冲突的不同方法,进一步明确信任函数中冲突证据合成的若干基本问题.韩德强等^[15]针对证据组合时的某些冲突问题提出一种基于证据方差的序贯式加权证据组合方法.即首先基于 Jousselme 证据距离定义了证据方差,每一步证据组合时,依据当前既有证据组合结果序列的方差及当前步新证据加入后的序列方差来生成权重.基于所获权重修正新到证据及前一步组合结果,最后利用 Dempster 规则完成当前步证据组合,该方法应用的前提是需满足序贯融合条件.胡丽芳等^[16]针对经典证据理论中的冲突系数无法合理度量证据之间的冲突程度,提出一种证据冲突系数表示模型,在冲突证据原因分析中总结出冲突的成因.Dezert 和 Smarandache^[17]于 2003 年在 D-S 证据理论和贝叶斯推理的基础上提出 Dezert Smarandache 理论(DSmT),在其框架中, Θ 被认为只是一个包含 n 个完备命题(假设封闭的) $\{\theta_1, \dots, \theta_n\}$ 的集合,其超幂集空间是建立在 Dedekind 格子模型 D^Θ 上,命题之间不涉及到其他的约束(相互排斥或者非存在性约束),即命题之间并不能被精确定义与分离.熊彦铭等^[18]在给出证据之间焦元距离的基础上,计算各个证据的支持度,对所有证据进行支持度加权平均,获得一个“参考证据”,利用该参考证据对各个原始证据进行不确定性判定及修正,并利用 Dempster 组合规则进行组合.

3) 对模型和规则同时修改

卢正才等^[19]提出了同时修正证据源和合成规则的原则,并给出了证据合成的一般框架,在此框架下,首先对证据源进行两级可靠性度量处理,然后再按照合成规则进行合成,最后采用基于 Pignistic 概率距离计算的权重把冲突系数分配给各焦元,实现对合成结果的合理修正.李文立等^[20]提出一种新的基于证据可信度的合成方法,在分配冲突的同时应对证据模型进行修正,实现两者紧密结合,在对一致性证据的“与”运算合成问题上,以及证据冲突在各合成命题的比例分配上,均有效利用了各证据的可信度等全局信息.王连锋等^[21]提出一种基于模糊聚类分析的混合证据组合方法,该方法利用 Pignistic 概率距离构建证据间的模糊相似矩阵及聚类有效性指标,通过传递闭包法对证据进行最优聚类及可信度求取,对证据进行修正后,采用 D-S 规则对同一类中的证据组合,利用统一信度函数组合方法对不同类中的证据进行最终合成.权文等^[22]借鉴证据分类的思想,定义证据一属性支持度、分类门限、证据可信度,将证据分类,保证同一类别的证据具有较好的一致性.在同一类别内采用 D-S 证据理论进行证据合成,并采用加权合成方法给出最终的合成结果.王进花等^[23]在三角模算子和折扣因子分析的基础

上,采用基于三角模算子定义的平均证据距离与冲突因子将证据分成可信任证据、不冲突证据和冲突证据三类,并赋予可信任证据和不冲突证据折扣因子 1,极大程度上保留了证据对正确假设的支持;然后基于证据距离定义了改进的证据权重,基于加权原则对冲突证据进行合成得到修正的证据体,从而消除证据间的冲突,最后利用 Dempster 规则完成证据组合.

综上所述,不管是单纯从模型还是从规则上进行改进,尽管在某些特殊场合取得了一定的效果,总的来说,还是很难从根本上解决 Dempster 的组合冲突问题;对模型和规则同时修改的策略在解决该类问题更具有优势,但目前的方法在冲突证据的划分与聚类,以及组合规则改进方面都存在或多或少的问题,比如,证据之间距离的衡量,聚类分析方法与评价指标,权重智能化调节、影响可靠度计算的因素等.

2 证据相似性表征

证据间距离的衡量不是一个新课题,已有很多专家学者从事这方面的研究^[24-29],目前应用较广的有 Pignistic 距离^[8,27],余弦距离以及 Jousselme^[25]距离等.但这些距离只是根据两证据源对应焦元(或者相关焦元集)之间的距离来进行描述和表征,而没有考虑证据源自身各个焦元之间的赋值大小顺序(可能导致“自冲突或者自矛盾”).

为了考虑证据自身产生的冲突,这里根据焦元之间赋值大小顺序定义一个新的证据相似性测度.在此之前,为了方便给出这个相似性测度,先定义一个两组数据顺序相关系数 μ .

定义 1^[30]. 设有两组数据 $\{x_1, x_2, \dots, x_n\}$, $\{y_1, y_2, \dots, y_n\}$, 对 $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n$ 作递增排序,排序后的数据记为 $x_{p_1}, x_{p_2}, \dots, x_{p_n}$ 及 $y_{q_1}, y_{q_2}, \dots, y_{q_n}$, 满足 $x_{p_1} \leq x_{p_2} \leq \dots \leq x_{p_n}$ 及 $y_{q_1} \leq y_{q_2} \leq \dots \leq y_{q_n}$, 对每个 p_i , 在 $q_1, q_2, \dots, q_n$ 中检索到 p_i 的位置, 设为 q_j , 即 $q_j = p_i$. 记为 $j = f(i)$, 则相关系数

$$\mu = \frac{\sum_{i=1}^n (i - j)^2}{\sum_{i=1}^n [n - (i - 1) - i]^2} \quad (5)$$

相关系数具有性质: $0 \leq \mu \leq 1$. $\mu = 0$ 时, 两组数据趋同性最大, $\mu = 1$ 时, 两组数据反向相关程度最大.

例 3. 假设两组数据: $[x_1, x_2, x_3] = [0, 0.1, 0.9]$, $[y_1, y_2, y_3] = [0.9, 0.1, 0]$, 先对两组数据分别进行从小到大排序, 排序结果为 $[x_1, x_2, x_3]$ 和 $[y_3, y_2, y_1]$, 对于每一个 p_i , 在第二个序列中检索 p_i 的位置 q_j ,

记为 $j = f(i)$, 比如对于第一组数下标 $i = 1$, 检索到下标 1 在第二个序列的第三个位置, 即 $j = 3$, 于是得到如下结果: $f(1) = 3, f(2) = 2, f(3) = 1$, 根据式 (5):

$$\mu = \frac{(1-3)^2 + (2-2)^2 + (3-1)^2}{[3-(1-1)-1]^2 + [3-(2-1)-2]^2 + [3-(3-1)-3]^2} = \frac{(1-3)^2 + (2-2)^2 + (3-1)^2}{[3-1]^2 + [3-1-2]^2 + [3-2-3]^2} = 1$$

于是, 表明两组数据不相关, 其相似性为 0.

2.1 两证据间焦元一致情况

定义 2. 对于任意两个证据源 S_1, S_2 , $m_1(\cdot)$ 和 $m_2(\cdot)$ 分别是对辨识框架 Θ 中的 n 个焦元的概率赋值 (且各个子命题的概率赋值不等), 其中 $X_i, Y_i \in Z^+$ 是根据命题子集的概率赋值大小排序后, 检索对应位置的值, 即 $f(X_i) = Y_i$, 则根据式 (5), 具有表征子命题赋值大小次序的证据顺序相似度函数如下:

$$Sim_{seq}(m_1, m_2) = 1 - \frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (n+1-2i)^2} \quad (6)$$

众所周知, 若存在一个表征任意两个证据源的相似性函数 $Sim(m_i, m_j)$, 那么必须满足下列三个基本条件:

- 1) 对称性: $\forall m_i(\cdot), m_j(\cdot), Sim(m_i, m_j) = Sim(m_j, m_i)$;
- 2) 一致性: $\forall m(\cdot), Sim(m, m) = 1$;
- 3) 非负性: $\forall m_i(\cdot), m_j(\cdot), 1 \geq Sim(m_i, m_j) \geq 0$.

于是下面可以证明 $Sim_{seq}(m_i, m_j)$ 是一个相似性测度函数.

证明.

- 1) 对称性: $\forall m_i(\cdot), m_j(\cdot), Sim_{seq}(m_i, m_j) = 1 - \frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (n+1-2i)^2} = 1 - \frac{\sum_{i=1}^n (Y_i - X_i)^2}{\sum_{i=1}^n (n+1-2i)^2} = Sim_{seq}(m_j, m_i)$, 因此结论得证. $\square$

- 2) 一致性: $\forall m(\cdot), Sim_{seq}(m, m) = 1 - \frac{\sum_{i=1}^n (X_i - X_i)^2}{\sum_{i=1}^n (n+1-2i)^2} = 1 - 0 = 1$, 因此, 结论得证. $\square$

- 3) 非负性: $\forall m_i(\cdot), m_j(\cdot), Sim_{seq}(m_i, m_j) = 1 - \frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (n+1-2i)^2}$, 要证明 $0 \leq Sim_{seq}(m_i, m_j) \leq 1$, 那么只需证明 $0 \leq \frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (n+1-2i)^2} \leq 1$. 因为 $\sum_{i=1}^n (n+1-2i)^2 = \sum_{i=1}^n (X_i - Y_{n-i+1})^2$, 只需证明 $\sum_{i=1}^n (X_i - Y_i)^2 \leq \sum_{i=1}^n (X_i - Y_{n-i+1})^2$. 又由于 $\sum_{i=1}^n (X_i - Y_i)^2 = \sum_{i=1}^n X_i^2 - 2X_i Y_i + Y_i^2$, $\sum_{i=1}^n (X_i - Y_{n-i+1})^2 = \sum_{i=1}^n X_i^2 + Y_{n-i+1}^2 - 2X_i Y_{n-i+1}$, $\sum_{i=1}^n X_i^2 + Y_i^2 = \sum_{i=1}^n X_i^2 + Y_{n-i+1}^2$, 因此, 只需证明 $\sum_{i=1}^n X_i Y_i \geq \sum_{i=1}^n X_i Y_{n-i+1}$, 这

一结论已经在文献 [30] 得证, 这里不再重复. 因此, $1 \geq Sim(m_i, m_j) \geq 0$. $\square$

例 4. 假设两个证据源 $m_1 = [x_1, x_2, x_3] = [0, 0.1, 0.9]$, $m_2 = [y_1, y_2, y_3] = [0, 0.1, 0.9]$, 对于 m_1 和 m_2 , 各自按大小排序后的顺序保持不变, 即 $[X_1, X_2, X_3] = [1, 2, 3]$, $[Y_1, Y_2, Y_3] = [1, 2, 3]$, 于是根据式 (6), $Sim_{seq}(m_i, m_j) = 1$, 表明两个证据完全相似.

定义 3. 对于任意两个证据源 S_1, S_2 , $m_1(\cdot)$ 和 $m_2(\cdot)$ 分别是对辨识框架 Θ 中的 n 个焦元的 BPA (且各个子命题的概率赋值可能相等, 假设 $m_1(\cdot)$ 中存在 s_1 个子命题的 BPA 相等, $m_2(\cdot)$ 中存在 s_2 个子命题 BPA 相等), 其中 X_i, Y_i 是根据命题子集的 BPA 大小排序后的序号, 下标 i 表示第 i 个命题子集, 由于存在子命题 BPA 相等情况, 对于证据源 S_1 , 存在 s_1 种排序可能; 对于证据源 S_2 , 存在 s_2 种排序可能, 因此两个证据源的对应排序存在 $s_1 \times s_2$ 种可能, 则具有表征子命题赋值大小次序的证据相似性函数如下:

$$Sim'_{seq}(m_1, m_2) = 1 - \frac{\sum_{t=1}^{s_1 s_2} \sum_{i=1}^n (X_i^t - Y_i^t)^2}{s_1 s_2 \left(\sum_{i=1}^n (n+1-2i)^2 \right)} \quad (7)$$

同理, 很容易证明 $Sim'_{seq}(m_X, m_Y)$ 也是一个证据相似性测度.

为了综合考虑影响证据距离的因素, 这里在余弦距离 (或相似测度) 的基础上 (当然, 也可以采用其他距离, 如 Jousselme 距离, Pignistic 概率距离等), 结合本文提出的顺序相似测度进行改进.

$$Sim(m_i, m_j) = \alpha Cos(m_i, m_j) + \beta Sim_{seq}(m_i, m_j) \quad (8)$$

其中, $Cos(m_i, m_j)$ 为两个证据向量 m_i, m_j 之间的余弦相似度, 即 $Cos(m_i, m_j) = \frac{m_i \cdot m_j}{\|m_i\| \cdot \|m_j\|}$. $0 \leq \alpha \leq 1, 0 \leq \beta \leq 1, \alpha + \beta = 1$.

注 1. 本文后面的算例中 $\alpha = \beta = 0.5$. 这里很容易证明, 改进后的测度函数 $Sim(m_i, m_j)$ 仍然是一个相似性测度函数, 这是因为两个相似测度函数之间的比例线性和仍然满足相似性测度的定义.

2.2 两证据间焦元不一致情况

假设存在两个鉴别框分别是 Θ_1 和 Θ_2 , 且 $\Theta_1 \neq \Theta_2$, 两者的幂集空间分别是 2^{Θ_1} 和 2^{Θ_2} , 且 $2^{\Theta_1} \neq 2^{\Theta_2}$, Γ 称为证据的核.

下面分单子焦元和析取焦元两种情况进行讨论, 在讨论之前, 先介绍一下两者的定义: 对于幂集中的一个子集 A , 只要有 $m(A) \geq 0$, 则称 A 为焦元. 焦

元中包含的元素个数称为焦元的基, 当 A 中只含一个元素时, 称为单子焦元. 当 A 中含有多个元素时, 称为析取焦元, 例如 $\Gamma = \{A, B, \{B, C\}\}$ 中, A, B 都是单子焦元, $\{B, C\}$ 就是析取焦元.

2.2.1 仅单子焦元情况

假设存在两个独立证据 $\{2^{\Theta_i}, \Gamma_i, m_i\}$, $i = 1, 2$, 其中, m_i 为 $2^{\Theta_i}, i = 1, 2$ 上的基本概率赋值 $m(A_i^j)$, $A_i^j \in \Gamma_i, j = 1, 2, \dots, |\Gamma_i|$, 且 $\Gamma_1 \neq \Gamma_2$, 即 $\Gamma_1 = \{A_1^1, A_1^2, A_1^3, \dots, A_1^{|\Gamma_1|}\}$, $\Gamma_2 = \{A_2^1, A_2^2, A_2^3, \dots, A_2^{|\Gamma_2|}\}$, 两者的核空间不一致, 那么两者合成核空间 $\Gamma = \Gamma_1 \cup \Gamma_2$, 而两者的重叠核空间为 $\Gamma_1 \cap \Gamma_2$. 设 $m_i(A_i^{\bar{j}}) = 0$, $A_i^{\bar{j}} \in \Gamma_i - \Gamma_1 \cap \Gamma_2$, 其中, $\bar{j}$ 表示 j 的补集, 利用式 (8), 即可求两个证据源之间的相似度.

2.2.2 析取焦元情况

假设存在两个独立证据 $\{2^{\Theta_i}, \Gamma_i, m_i\}$, $i = 1, 2$, m_i 为 $2^{\Theta_i}, i = 1, 2$ 上的基本概率赋值 $m(A_i^j)$, $A_i^j \in \Gamma_i, j = 1, 2, \dots, |\Gamma_i|$, 且 $\Gamma_1 \neq \Gamma_2$.

定义 4^[8]. 设 $m(\cdot)$ 为 2^{Θ} 的基本概率赋值 BPA, 命题子集 $A, B \in 2^{\Theta}$, 且满足

$$\text{Bet } P_m(B) = \sum_{A \in 2^{\Theta}} \frac{|A \cap B|}{|A|} m(A) \quad (9)$$

其中, $|B|$ 表示 B 的子集的基, $\text{Bet } P_m(B)$ 描述了基本概率赋值 $m(\cdot)$ 支持幂集 2^{Θ} 上命题子集 B 为真的全部概率值.

根据式 (9), 可以得到上述假设中不同 2^{Θ} 下所有子命题的 Pignistic 概率 $\text{Bet } P_m(A_i^j)$, 设 $\text{Bet } P_i(A_i^{\bar{j}}) = 0$, $A_i^{\bar{j}} \in \Gamma_i - \Gamma_1 \cap \Gamma_2$, 其中 $\bar{j}$ 表示 j 的补集, 将 $m(\cdot)$ 替换为 $\text{Bet } P_m(\cdot)$, 然后利用式 (8), 即可求两个证据源之间的相似度.

3 证据源修正

假设存在 n 个独立证据源, 根据相似性测度函数式 (8) 计算得到证据间相似测度矩阵 $C_{n \times n}$

$$C_{n \times n} = \begin{bmatrix} 1 & \text{Sim}_{12} & \text{Sim}_{13} & \cdots & \text{Sim}_{1n} \\ \text{Sim}_{21} & 1 & \text{Sim}_{23} & \cdots & \text{Sim}_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{Sim}_{n1} & \text{Sim}_{n2} & \text{Sim}_{n3} & \cdots & 1 \end{bmatrix} \quad (10)$$

注意, 矩阵 (10) 中的元素 $\text{Sim}_{ij} = \text{Sim}(m_i, m_j)$, 满足 $\text{Sim}_{ij} = \text{Sim}_{ji}$, 可见相似测度矩阵 $C_{n \times n}$ 是一个对称阵.

对于每一个证据源, 都存在被其他证据源支持的程度. 可以认为, 两证据源间相似度高的表明证据源之间相互支持程度高. 因此, 根据相似测度矩阵 $C_{n \times n}$, 即可获得证据体的众信度为 $\text{Crd}(m_i)$.

$$\text{Crd}(m_i) = \frac{\text{Sup}(m_i)}{\max_{1 \leq j \leq n} [\text{Sup}(m_j)]} \quad (11)$$

其中, $\text{Sup}(m_i) = \sum_{j=1, j \neq i}^n \text{Sim}(m_i, m_j)$, $i = 1, \dots, n$.

根据证据的众信度, 对初始证据源进行修正如下:

$$\begin{cases} m'_i(A) = \text{Crd}(m_i) \cdot m_i(A), \forall A \subseteq \Theta \\ m'_i(\Theta) = \text{Crd}(m_i) \cdot m_i(\Theta) + 1 - \text{Crd}(m_i) \end{cases} \quad (12)$$

4 证据聚类与合成

在批量证据融合中, 证据往往具有群聚性, 也就是说, 表达相似观点的证据比较相近, 它们之间的距离 (冲突) 较小, 容易聚成一类; 而观点不一致的证据之间的距离 (冲突) 往往较大, 不易聚成一类. 若能将其聚成一类的证据, 直接采用 Dempster 规则合成, 因为证据间冲突较小, Dempster 规则的缺陷表现不是很突出. 当聚类之间的冲突较大时, 不易直接采用 Dempster 规则合成, 需对它进行修正.

4.1 ISODATA 证据聚类

由于证据的聚类数预先未知, 于是本文采用经典的非监督学习动态聚类算法 ISODATA^[4], 其步骤如下:

假设每个证据包含 v 个焦元, 每类中心证据源为 M_k , 其聚类证据由类中所有证据源的均值确定, 每类中证据源的数目为 N_k . 设定初始参数: 预期分类数目 K , 同一类证据源标准差上限 θ_s , 各类中心证据源最小距离下限 θ_c .

注 2. 这里的距离采用欧氏距离, 若核空间不同, 且涉及析取焦元情况, 采用证据的 Pignistic 概率 $\text{Bet } P$ 替代基本概率赋值 $m(\cdot)$.

1) 将 n 个证据源 m_i 分给最近的聚类中心 M_k , 并修正聚类中心证据源为

$$M_k(f) = \frac{1}{N_k} \sum_{i=1}^{N_k} m_i(f), f = f_1, f_2, \dots, f_v \quad (13)$$

2) 若迭代未完成, 则通过已有聚类中心数目与迭代次数判断进行分裂处理 3) 还是合并处理 4), 若迭代完成, 直接进行 4).

3) 计算每个聚类中证据源距离的标准差向量

$$\sigma_k = (\sigma_{1k}, \sigma_{2k}, \sigma_{3k}, \dots, \sigma_{vk})^T \quad (14)$$

其中向量的各个分量为

$$\sigma_{fk} = \sqrt{\frac{1}{N_k} \sum_{m_i \in C_k} (m_i(f) - M_k(f))^2} \quad (15)$$

其中, k 为以 M_k 为中心证据源的聚类序号, C_k 表示以证据源 M_k 为中心的证据集合, C_k 中证据源数目为 N_k .

若第 l 类中最大分量大于同一类证据源标准差上限 θ_s , 并且满足分裂条件则对第 l 类进行分裂处理, 并返回 1). 若分裂条件不满足, 则进行合并处理 4).

4) 计算全部聚类中心证据源的距离

$$D_{ij} = \sqrt{\sum_{z=1}^v (M_i(f_z) - M_j(f_z))^2} \quad (16)$$

将小于中心证据源之间最小距离下限 θ_c 的 D_{ij} 进行递增排列, 即

$$D = \{D_{i_1 j_1}, D_{i_2 j_2}, \dots, D_{i_L j_L}\} \quad (17)$$

并对 D 中每一个 D_{i_k, j_k} 所表示的第 i_k 类和第 j_k 类证据集合, 即 C_{i_k} 和 C_{j_k} 进行合并.

5) 判断迭代是否完成, 若未完成返回 1), 否则聚类结束. 其流程图如图 1 所示.

为保证聚类的有效性, 给出有效的评价准则函数来评判聚类质量非常重要.

设 n 个初始证据被分成 c 类, 其中第 j 类中有 n_j 个证据. 误差平方和 J_c 聚类准则函数定义为

$$J_c = \sum_{j=1}^c \sum_{i=1}^{n_j} \|m_{ij} - \bar{m}_j\|^2 \quad (18)$$

类间距离和准则函数定义为

$$J_b = \sum_{j=1}^c (\bar{m}_j - \bar{m})^T (\bar{m}_j - \bar{m}) \quad (19)$$

其中, m_{ij} 为第 j 类中的第 i 个证据, $\bar{m}_j$ 为类型 j 中包含所有证据的均值, 即

$$\bar{m}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} m_{ij}, \quad j = 1, 2, \dots, c \quad (20)$$

$\bar{m}$ 为全部样本的均值, 即

$$\bar{m} = \frac{1}{n} \sum_{k=1}^n m_k \quad (21)$$

J_c 表示描述 n 个证据分成 c 类时, 所产生的总误差平方和, J_c 越小越好. J_b 表示所产生的类间距离和, J_b 越大越好.

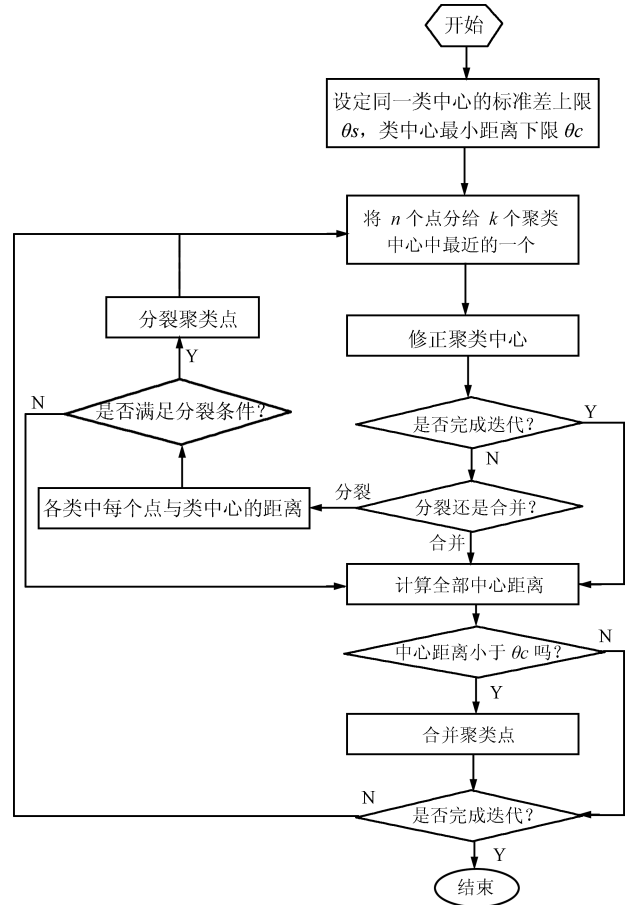


图 1 ISODATA 流程图

Fig. 1 The flowchart of ISODATA

这个指标的好坏, 直接影响到聚类结果, 下面通过例 5 中一个数字例子与文献 [21] 提到的指标进行聚类效果对比.

例 5. 通过采用文献 [21] 的传递闭包聚类方法和评价指标, 将表 1 中的 12 类证据源聚成三类, 即 $\{m_1, m_2, m_5, m_6, m_9, m_{10}\}$, $\{m_3, m_7, m_{11}\}$, $\{m_4, m_8, m_{12}\}$, 根据式 (18) 和式 (19), 分别计算得到 $J_c = 0.0290$, $J_b = 0.3170$. 而采用本文 ISODATA 方法和评价指标, 聚成四类: $\{m_1, m_5, m_9\}$, $\{m_2, m_6, m_{10}\}$, $\{m_3, m_7, m_{11}\}$, $\{m_4, m_8, m_{12}\}$, 计算得到的 $J_c = 0.0027$, $J_b = 0.3608$.

按照 J_c 越小越好 J_b 越大越好的原则, 可见, 通过本文提出的聚类指标进行的聚类效果比文献 [21] 的好.

4.2 聚类的可靠度

把 n 个初始证据分成 c 类之后, 为求出每一类证据的可靠度, 本文提出的聚类可靠度计算方法, 综合考虑了每一类中各个证据的众信度 Crd 以及每一类中包含所有的证据频率 (即该类中证据个数与总证据个数的比值). 其中, 每一类中各个证据的众信度越大, 或者包含的证据个数越多, 则此类的可靠度

表 1 待聚类的随机证据
Table 1 The random evidences to be clustered

	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}
A	0.512	0.588	0.221	0.299	0.488	0.612	0.189	0.301	0.513	0.587	0.212	0.298
B	0.287	0.191	0.198	0.198	0.288	0.192	0.713	0.202	0.288	0.192	0.702	0.199
C	0.201	0.221	0.503	0.503	0.224	0.196	0.098	0.497	0.199	0.221	0.086	0.503

就越大, 否则, 可靠度就越小. 其聚类加权可靠度的公式如下:

$$\gamma_j = \bar{h} \cdot \frac{\sum_{j=1}^{n_j} Crd_j}{n_j} + \lambda \cdot \frac{n_j}{n} \quad (22)$$

其中, $0 \leq \bar{h} \leq 1$, $0 \leq \lambda \leq 1$, $\bar{h} + \lambda = 1$.

注 3. 本文后面的算例中 $\bar{h} = \lambda = 0.5$.

由于 γ_j 是非归一化的, 因此, 必须对 γ_j 进行归一化得到聚类 j 的可靠度为

$$\omega_j = \frac{\gamma_j}{\sum_{j=1}^c \gamma_j} \quad (23)$$

4.3 证据合成

假设存在 n 个鉴别框架 $\Theta_1, \Theta_2, \dots, \Theta_n$, 与之相关的 n 组证据源的基本概率赋值 $m_i, i = 1, 2, \dots, n$, 首先对证据进行预处理, 得到修正后的证据模型 $m'_i, i = 1, 2, \dots, n$, 然后, 对证据进行聚类 (这里假设聚成 c 类), 其中, 每一聚类中的证据比较相近, 且证据间冲突较小, 但两簇不同聚类中的证据之间的冲突比较大. 因此, 对 c 个聚类中的证据分别按照 Dempster 规则 3) 或 4) 进行合成 (称为类内合成), 得到 c 个聚类中心的证据 $\hat{m}'_j, j = 1, 2, \dots, c$, 这里称为证据代表. 由于各个聚类合成后的证据代表之间一般都具有较大的冲突, 若再次直接用 Dempster 规则进行合成, 势必导致该规则对高度冲突处理的不足表现比较突出, 为了进一步解决这个问题, 于是对于聚类后的证据代表 $\hat{m}'_j, j = 1, 2, \dots, c$, 采用文献 [6] 的统一合成规则进行合成, 即将冲突作为有用信息, 根据不同的权重重新嵌入规则中:

$$\begin{cases} m'(A) = \sum_{\cap A_i = A} \prod_{1 \leq j \leq c} \hat{m}'_j(A_i) + K' \cdot \omega(A), \\ A \subseteq \Theta, A \neq \phi \\ m'(A) = 0, A = \phi \end{cases} \quad (24)$$

其中, $K' = \sum_{\cap A_i = \emptyset} \prod_{1 \leq j \leq c} \hat{m}'_j(A_i)$, 表示修正后证据模型的总冲突. $\omega(A) = \sum_{i=1}^c \omega_i \hat{m}'_i(A)$, $\omega(A)$ 表示冲突分配给该命题 A 的权重.

归纳起来, 本文方法的证据合成流程图如图 2 所示, 其主要步骤如下:

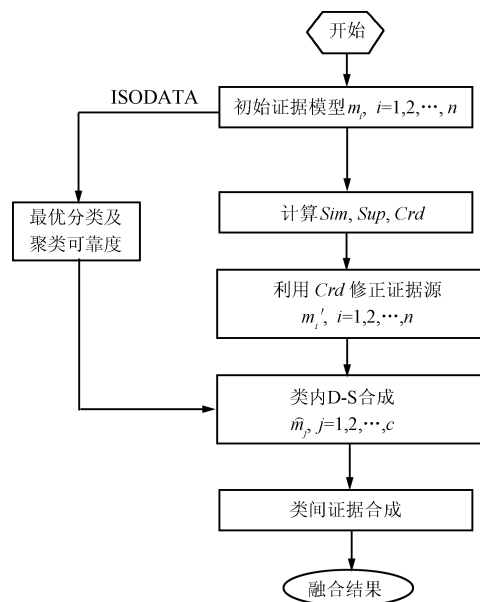


图 2 证据合成流程图

Fig. 2 The flowchart of evidence combination

步骤 1. 对收集到的 n 个证据 $m_i, i = 1, 2, \dots, n$, 计算两两证据之间的基于相关系数的相似度 Sim , 以及每个证据的众信度 Crd ;

步骤 2. 根据众信度 Crd 对初始证据进行修正, 得到证据模型 $m'_i, i = 1, 2, \dots, n$;

步骤 3. 利用 ISODATA 方法对初始证据源进行分类, 得到最优聚类并利用众信度 Crd 计算每一类的可靠度 ω_j ;

步骤 4. 对同一聚类内的证据进行 Dempster 规则融合, 生成新的证据代表 $\hat{m}'_j, j = 1, 2, \dots, c$;

步骤 5. 对所有证据代表 $\hat{m}'_j, j = 1, 2, \dots, c$ 按照式 (24) 所示的统一证据合成方法进行证据合成.

下面通过一个简单数字例子来解释这个算法.

例 6. 设在辨识框架 $\Theta = \{A, B, C\}$ 下, 某系统有 6 条独立证据 $\{\Theta, \Gamma, m_i\}, i = 1, 2, \dots, 6$:

$$\begin{aligned} \Gamma_1 &= \{\{A\}, \{B\}, \{C\}\}, m_1 = \{0.60, 0.10, 0.30\} \\ \Gamma_2 &= \{\{A\}, \{B\}, \{C\}\}, m_2 = \{0.55, 0.1, 0.35\} \\ \Gamma_3 &= \{\{A\}, \{B\}, \{C\}\}, m_3 = \{0.00, 0.90, 0.10\} \\ \Gamma_4 &= \{\{A\}, \{B\}, \{C\}\}, m_4 = \{0.55, 0.10, 0.35\} \end{aligned}$$

$$\Gamma_5 = \{\{A\}, \{B\}, \{C\}\}, m_5 = \{0.55, 0.10, 0.35\}$$

$$\Gamma_6 = \{\{A\}, \{B\}, \{C\}\}, m_6 = \{0.55, 0.1, 0.35\}$$

步骤 1. 对于上述证据集, 利用本文的相似性测度式 (8), 根据式 (10) 计算证据间的相似测度矩阵 $C_{n \times n}$:

$$C_{n \times n} = \begin{bmatrix} 1.000 & 0.997 & 0.098 & 0.997 & 0.997 & 0.997 \\ 0.997 & 1.000 & 0.105 & 1.000 & 1.000 & 1.000 \\ 0.098 & 0.105 & 1.000 & 0.105 & 0.105 & 0.105 \\ 0.997 & 1.000 & 0.105 & 1.000 & 1.000 & 1.000 \\ 0.997 & 1.000 & 0.105 & 1.000 & 1.000 & 1.000 \\ 0.997 & 1.000 & 0.105 & 1.000 & 1.000 & 1.000 \end{bmatrix}$$

然后, 根据式 (11) 计算证据的众信度 Crd : $Crd = [0.996, 1.000, 0.126, 1.000, 1.000, 1.000]$.

步骤 2. 根据得到的众信度 Crd 对初始证据利用式 (12) 进行修正, 得到修正后的证据模型 m'_i 为

$$\begin{bmatrix} m'_1 \\ m'_2 \\ m'_3 \\ m'_4 \\ m'_5 \\ m'_6 \end{bmatrix} = \begin{bmatrix} 0.598 & 0.1 & 0.299 & 0.004 \\ 0.55 & 0.1 & 0.35 & 0 \\ 0 & 0.113 & 0.013 & 0.874 \\ 0.55 & 0.1 & 0.35 & 0 \\ 0.55 & 0.1 & 0.35 & 0 \\ 0.55 & 0.1 & 0.35 & 0 \end{bmatrix}$$

步骤 3. 利用 ISODATA 方法对初始证据源进行聚类, 参数的选取为: 标准差上限 $\theta_s = 0.05$, 最小距离下限 $\theta_c = 0.05$. 证据源被聚成了两类 $\{m_1, m_2, m_4, m_5\}, \{m_3\}$, 然后利用聚类的可靠度式 (22) 与式 (23) 计算出每一类的可靠度:

$$\omega_1 = 0.862, \omega_2 = 0.138$$

步骤 4. 对同一聚类内的所有证据, 利用式 (3) 进行 Dempster 融合, 生成新的证据代表:

$$\hat{m}'_1 = \begin{bmatrix} 0.9236 & 0.0002 & 0.0762 & 0 \end{bmatrix}$$

$$\hat{m}'_2 = \begin{bmatrix} 0 & 0.1132 & 0.0126 & 0.8741 \end{bmatrix}$$

步骤 5. 对两个证据代表按照式 (24) 所示的统一证据合成方法进行证据合成, 得到最后融合结果:

$$m = \begin{bmatrix} 0.9069 & 0.0021 & 0.0760 & 0.0150 \end{bmatrix}$$

5 证据合成实验

为了充分验证本文方法的好处, 这里给出三个数字算例, 分别由单冲突、单冲突与中立、双冲突参

与合成时, 同其他现存的主要合成方法进行深入对比.

尽管目前尚无统一评价融合结果的标准, 常见的衡量方法是在处理冲突证据或者在证据源干扰的时候, 根据融合方法能否得到正确的融合结果, 以及融合的准确性, 收敛速度, 以及抗干扰能力作为判断依据.

5.1 单冲突证据参与合成

例 7. 为了验证本文算法对于存在单独一组冲突证据的有效性, 针对例 6 给出的证据集, 分别以 2 个证据依次递增到 6 个证据的方式进行合成实验, 把本文的方法与 D-S 方法、Yager 方法、王栋方法、王连锋方法等进行比较. 为了方便, 这里重新给出这组证据集: 设在辨识框架 $\Theta = \{A, B, C\}$ 下, 某系统有 6 条独立证据 $\{\Theta, \Gamma, m_i\}, i = 1, 2, \dots, 6$:

$$\Gamma_1 = \{\{A\}, \{B\}, \{C\}\}, m_1 = \{0.60, 0.10, 0.30\}$$

$$\Gamma_2 = \{\{A\}, \{B\}, \{C\}\}, m_2 = \{0.55, 0.1, 0.35\}$$

$$\Gamma_3 = \{\{A\}, \{B\}, \{C\}\}, m_3 = \{0.00, 0.90, 0.10\}$$

$$\Gamma_4 = \{\{A\}, \{B\}, \{C\}\}, m_4 = \{0.55, 0.10, 0.35\}$$

$$\Gamma_5 = \{\{A\}, \{B\}, \{C\}\}, m_5 = \{0.55, 0.10, 0.35\}$$

$$\Gamma_6 = \{\{A\}, \{B\}, \{C\}\}, m_6 = \{0.55, 0.1, 0.35\}$$

显而易见, m_3 与其他证据源冲突, 当采用 ISODATA 算法对证据源 $\{\Theta, \Gamma, m_i\}, i = 1, 2, \dots, 6$ 进行聚类时, 参数设置如下: 以初始的证据个数作为预期分类数, 标准差上限 θ_s 与最小距离下限 θ_c 分别为 0.05 和 0.05. 所有证据源可以聚为 2 类, 即 $\{m_3\}$ 单独作为一类, 其他作为一类.

从表 2 可看出, 在证据 m_3 出现后, D-S 法、Yager 法的 $m(A)$ 值突然变为 0, 而且之后始终为 0, 反应明显过度, 出现“一票否决”现象. D-S 方法和 Yager 的方法都不适合融合高冲突证据. 孙全方法当冲突出现时, 反应过度, 且具有跟 D-S 和 Yager 规则相同的弊病, 即使后面的证据源都是真的 (持肯定态度) 的, 其 6 个证据源的最终合成结果也不尽人意, 且收敛速度极慢. 卢正才方法虽然在冲突出现时, 反应比较正常, 但随着证据的继续合成, 其最终结果出现了不收敛现象, 这显然是不正常的. 熊彦铭方法在冲突证据出现的情况下, 不至于反应过度, 可以得到较为合理的结果, 但是 6 个证据源最终合成结果中真命题的 Mass 值的准确度略低; 王栋方法当冲突出现时, 反应剧烈, 虽然融合结果向真命题 A 聚焦, 但是其收敛速度较慢. 李文立, 韩德强, 王连锋的方法都对证据进行了预处理, 当冲突出现时, 反应正常, 且融合结果都向真命题 A 聚焦, 最终的结果比较接近. 其中, 李文立方法基于焦元距离函数计算证据权重, 收敛速度一般. 韩德强方法基于证据方差的修正方法有效控制了 Dempster 规则

的收敛速度,使得结果更可信,但准确度仍然略低于本文方法,且易受融合次序影响较大,比较适合序贯融合方式. 王连峰方法提出基于传递闭包的可靠度计算及相应的证据组合方法,能有效避开干扰证据 m_3 的影响,融合结果向真命题 聚焦. 但是其聚焦速度和最终合成结果准确度不如本文方法. 本文方法提出的可靠度计算方法,充分考虑了证据自身和外部矛盾对证据合成的影响,能很大程度的减少不利证据对合成结果的影响,抗干扰能力较强. 这一点从表 2 最后两行中余弦相似度改进前后的合成结果对比可以看出,比如当冲突出现时,改进前的合成结果 $m(A) = 0.6574$,改进后 $m(A) = 0.7021$. 最终 6 个证据合成结果:改进前 $m(A) = 0.8801$,改进后 $m(A) = 0.9069$,改进后合成结果总是高于改进前的.

因此,通过以上相关文献和表 2 中数据分析可以发现,本文方法收敛速度快,抗干扰能力较强,准确度较高.

5.2 单冲突和中立证据同时参与

例 8. 对于同时存在冲突证据和较中立的证据集,验证这些证据对本文提出的合成方法以及其他方法的影响. 假设鉴别框架为 $\Theta = \{A, B, C\}$,某系统有 5 条独立证据 $\{\Theta, \Gamma, m_i\}, i = 1, 2, \dots, 5$:

$$\begin{aligned}\Gamma_1 &= \{\{A\}, \{B\}, \{C\}\}, m_1 = \{0.90, 0.00, 0.10\} \\ \Gamma_2 &= \{\{A\}, \{B\}, \{C\}\}, m_2 = \{0.00, 0.01, 0.99\} \\ \Gamma_3 &= \{\{A\}, \{B\}, \{C\}\}, m_3 = \{0.3334, 0.3333, 0.3333\} \\ \Gamma_4 &= \{\{A\}, \{B\}, \{C\}\}, m_4 = \{0.98, 0.01, 0.01\} \\ \Gamma_5 &= \{\{A\}, \{B\}, \{C\}\}, m_5 = \{0.90, 0.05, 0.05\}\end{aligned}$$

其中, m_2 与其他证据明显冲突,而 m_3 可以看作是一个中立的证据,但由于 $m_3(A) = 0.3334 > m_3(B) = m_3(C) = 0.3333$,可以看出还是略微有点支持 A .

通过选取不同的参数,有三种不同的分类情况,即 1) 当 $\theta_s = 0.05, \theta_c = 0.05$ 时, m_2 和 m_3 分别单独作为一类,其他作为一类; 2) 当 $\theta_s = 0.05, \theta_c = 0.75$ 时, m_2 单独作为一类,其他作为一类; 3) 当 $\theta_s = 0.35, \theta_c = 0.05$ 时, m_2 和 m_3 作为一类,其他作为一类.

通过表 3 可以看出,对于 Dempster 规则,当遇到高冲突时,后面再给定多少支持正确结论的证据源都无济于事,这显然是不合理的. 而 Yager 的结论 $m(A) = m(B) = 0$,同理,随着给定支持正确结论的证据源数目的增加, $m(A), m(B)$ 始终为 0, $m(C)$ 的值一直在减少, $m(\Theta)$ 的值却在一直增加,可见,决策结果越来越不确定,出现了熵

增现象,这样的组合结果是无效的. 当冲突出现时,孙全、邓勇和熊彦铭的合成结果,将赋值很大一部分给了 $m(\Theta)$,导致 $m(A)$ 很小,如孙全和邓勇的 $m(A) = 0.164679, m(C) = 0.298444$,熊彦铭的 $m(A) = 0.246347, m(C) = 0.278064$,而我们的方法 $m(A) = 0.405450, m(C) = 0.590045$,显然我们的结论更符合直观反应. 李弼程、王栋和韩德强的方法,当冲突证据融合后,再融合中立证据源时,由于 Γ_3 中 $m_3(A) = 0.3334 > m_3(B) = m_3(C) = 0.3333$,融合的结果都出现了 $m(A)$ 下降的现象,这显然不合理,这是因为中立证据的参与,合成结果至少不会降低. 而只有卢正才、李文立和本文方法能够正确解释,即冲突融合后,中立证据参与合成,出现了 $m(A)$ 略有上升的现象. 但卢正才和李文立方法没有本文方法 5 个证据源在分两类 $\{m_2\}\{m_1, m_3, m_4, m_5\}$ 时,最终融合的结果 $m(A) = 0.935899$ 准确度高. 但当分三类为 $\{m_2\}\{m_3\}\{m_1, m_4, m_5\}$ 时,本文方法 $m(A) = 0.706287$. 由该实验可知,本文方法优势比较明显,但与此同时,当聚类特性比较模糊时,也受聚类参数门限设置影响,最终合成结果有些许差异.

5.3 两冲突证据参与合成

例 9. 当两个冲突证据参与合成时,验证一下各个方法的性能,利用如下算例,某系统有 5 条独立证据 $\{\Theta, \Gamma, m_i\}, i = 1, 2, \dots, 5$:

$$\begin{aligned}\Gamma_1 &= \{\{A\}, \{B\}, \{C\}\}, m_1 = \{0.60, 0.10, 0.30\} \\ \Gamma_2 &= \{\{A\}, \{B\}, \{C\}\}, m_2 = \{0.10, 0.10, 0.80\} \\ \Gamma_3 &= \{\{A\}, \{B\}, \{C\}\}, m_3 = \{0.30, 0.10, 0.60\} \\ \Gamma_4 &= \{\{A\}, \{B\}, \{C\}\}, m_4 = \{0.70, 0.20, 0.10\} \\ \Gamma_5 &= \{\{A\}, \{B\}, \{C\}\}, m_5 = \{0.55, 0.10, 0.35\}\end{aligned}$$

其中, m_2 和 m_3 与 m_1, m_4, m_5 明显冲突,其分类参数的选取如下,以初始的证据个数作为预期类数,标准差上限 θ_s 与最小距离下限 θ_c 分别为 0.1 和 0.3. 可将上述证据源分成两类,即 $\{m_2, m_3\}$ 与 $\{m_1, m_4, m_5\}$ 各为一类. 采用各种不同的合成方法,其合成结果如表 4 所示. 结果分析:当从直观看 $m_1 = \{0.60, 0.10, 0.30\}$ 和 $m_2 = \{0.10, 0.10, 0.80\}$ 两个证据融合时,显然命题 C 是真的,此时只有韩德强的结果 $m(C) = 0.695402$ 高于本文方法 $m(C) = 0.619500$,而熊彦铭的方法只有 $m(C) = 0.436950$. 当 m_1, m_2, m_3 三个证据融合时,从原始的三个证据源可以直观看出,最终的正确的结论应当支持 C ,而高于本文方法 $m(C) = 0.670163$ 的只有熊彦铭的 $m(C) = 0.720752$ 和韩德强的 $m(C) = 0.825216$. 随后,又有两个支持 A 的证据

表 2 单冲突证据参与合成的结果对比

Table 2 The comparison of results when one conflict takes part in combination

融合规则	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	$m_1, m_2, m_3, m_4, m_5, m_6$	
D-S ^[1]	$m(A) = 0.7416$	$m(A) = 0.0000$	$m(A) = 0.0000$	$m(A) = 0.0000$	$m(A) = 0.0000$	
	$m(B) = 0.0224$	$m(B) = 0.4615$	$m(B) = 0.1967$	$m(B) = 0.0654$	$m(B) = 0.0196$	
	$m(C) = 0.2360$	$m(C) = 0.5385$	$m(C) = 0.8033$	$m(C) = 0.9346$	$m(C) = 0.9804$	
Yager ^[7]	$m(A) = 0.3300$	$m(A) = 0.0000$	$m(A) = 0.0000$	$m(A) = 0.0000$	$m(A) = 0.0000$	
	$m(B) = 0.0100$	$m(B) = 0.0009$	$m(B) = 0.0009$	$m(B) = 0.0001$	$m(B) = 0.0000$	
	$m(C) = 0.1050$	$m(C) = 0.0105$	$m(C) = 0.0037$	$m(C) = 0.0013$	$m(C) = 0.0005$	
	$m(\Theta) = 0.5550$	$m(\Theta) = 0.9805$	$m(\Theta) = 0.9954$	$m(\Theta) = 0.9986$	$m(\Theta) = 0.9995$	
孙全等 ^[31]	$m(A) = 0.5132$	$m(A) = 0.1740$	$m(A) = 0.2064$	$m(A) = 0.2262$	$m(A) = 0.2396$	
	$m(B) = 0.0419$	$m(B) = 0.1755$	$m(B) = 0.1466$	$m(B) = 0.1308$	$m(B) = 0.1198$	
	$m(C) = 0.2085$	$m(C) = 0.1240$	$m(C) = 0.1373$	$m(C) = 0.1471$	$m(C) = 0.1545$	
	$m(\Theta) = 0.2364$	$m(\Theta) = 0.5265$	$m(\Theta) = 0.5097$	$m(\Theta) = 0.4960$	$m(\Theta) = 0.4860$	
卢正才等 ^[19]	$m(A) = 0.6491$	$m(A) = 0.5255$	$m(A) = 0.5177$	$m(A) = 0.5123$	$m(A) = 0.5104$	
	$m(B) = 0.0655$	$m(B) = 0.0886$	$m(B) = 0.0966$	$m(B) = 0.1023$	$m(B) = 0.1048$	
	$m(C) = 0.2854$	$m(C) = 0.2376$	$m(C) = 0.2528$	$m(C) = 0.2731$	$m(C) = 0.2896$	
	$m(\Theta) = 0.0000$	$m(\Theta) = 0.1483$	$m(\Theta) = 0.1329$	$m(\Theta) = 0.1123$	$m(\Theta) = 0.0952$	
熊彦铭等 ^[18]	$m(A) = 0.5036$	$m(A) = 0.4928$	$m(A) = 0.6511$	$m(A) = 0.7561$	$m(A) = 0.8287$	
	$m(B) = 0.1384$	$m(B) = 0.1531$	$m(B) = 0.0576$	$m(B) = 0.0183$	$m(B) = 0.0053$	
	$m(C) = 0.3140$	$m(C) = 0.3110$	$m(C) = 0.2861$	$m(C) = 0.2250$	$m(C) = 0.1659$	
	$m(\Theta) = 0.0440$	$m(\Theta) = 0.0430$	$m(\Theta) = 0.0051$	$m(\Theta) = 0.0006$	$m(\Theta) = 0.0001$	
王栋等 ^[9]	$m(A) = 0.4225$	$m(A) = 0.1953$	$m(A) = 0.3427$	$m(A) = 0.4772$	$m(A) = 0.4936$	
	$m(B) = 0.0285$	$m(B) = 0.4773$	$m(B) = 0.2029$	$m(B) = 0.1129$	$m(B) = 0.0638$	
	$m(C) = 0.1614$	$m(C) = 0.1279$	$m(C) = 0.2044$	$m(C) = 0.2638$	$m(C) = 0.2425$	
	$m(\Theta) = 0.3876$	$m(\Theta) = 0.1995$	$m(\Theta) = 0.2500$	$m(\Theta) = 0.1460$	$m(\Theta) = 0.2001$	
李文立等 ^[20]	$m(A) = 0.6391$	$m(A) = 0.4980$	$m(A) = 0.6383$	$m(A) = 0.6944$	$m(A) = 0.7352$	
	$m(B) = 0.0658$	$m(B) = 0.1456$	$m(B) = 0.0704$	$m(B) = 0.0528$	$m(B) = 0.0362$	
	$m(C) = 0.2851$	$m(C) = 0.2961$	$m(C) = 0.2755$	$m(C) = 0.2476$	$m(C) = 0.2275$	
	$m(\Theta) = 0.0100$	$m(\Theta) = 0.0703$	$m(\Theta) = 0.0158$	$m(\Theta) = 0.0042$	$m(\Theta) = 0.0011$	
韩德强等 ^[15]	$m(A) = 0.7409$	$m(A) = 0.6129$	$m(A) = 0.7936$	$m(A) = 0.8776$	$m(A) = 0.9034$	
	$m(B) = 0.0224$	$m(B) = 0.2925$	$m(B) = 0.0903$	$m(B) = 0.0176$	$m(B) = 0.0060$	
	$m(C) = 0.2367$	$m(C) = 0.0946$	$m(C) = 0.1161$	$m(C) = 0.1048$	$m(C) = 0.0905$	
王连锋等 ^[21]	$m(A) = 0.6428$	$m(A) = 0.6021$	$m(A) = 0.6576$	$m(A) = 0.6846$	未分类	分类后
	$m(B) = 0.0690$	$m(B) = 0.1095$	$m(B) = 0.0650$	$m(B) = 0.0506$	$m(A) = 0.71$	$m(A) = 0.81$
	$m(C) = 0.2882$	$m(C) = 0.2884$	$m(C) = 0.2773$	$m(C) = 0.2648$	$m(B) = 0.03$	$m(B) = 0.01$
本文方法 (余弦相似 度改进后)					$m(C) = 0.26$	$m(C) = 0.18$
	$m(A) = 0.7416$	$m(A) = 0.7021$	$m(A) = 0.8023$	$m(A) = 0.8649$	$m(A) = 0.9069$	
	$m(B) = 0.0224$	$m(B) = 0.0332$	$m(B) = 0.0087$	$m(B) = 0.0034$	$m(B) = 0.0021$	
	$m(C) = 0.2360$	$m(C) = 0.2308$	$m(C) = 0.1665$	$m(C) = 0.1140$	$m(C) = 0.0760$	
本文方法 (余弦相似 度改进前)	$m(\Theta) = 0.0000$	$m(\Theta) = 0.0339$	$m(\Theta) = 0.0225$	$m(\Theta) = 0.0177$	$m(\Theta) = 0.0150$	
	$m(A) = 0.7416$	$m(A) = 0.6574$	$m(A) = 0.7684$	$m(A) = 0.8355$	$m(A) = 0.8801$	
	$m(B) = 0.0224$	$m(B) = 0.0571$	$m(B) = 0.0216$	$m(B) = 0.0130$	$m(B) = 0.0101$	
	$m(C) = 0.2360$	$m(C) = 0.2243$	$m(C) = 0.1643$	$m(C) = 0.1133$	$m(C) = 0.0760$	
	$m(\Theta) = 0.0000$	$m(\Theta) = 0.0612$	$m(\Theta) = 0.0457$	$m(\Theta) = 0.0382$	$m(\Theta) = 0.0338$	

表 3 冲突和中立证据同时参与合成的结果对比

Table 3 The comparison of results when one conflictive and neutral evidences simultaneously take part in combination

融合规则	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 ^[1]	$m(A) = 0$	$m(A) = 0$	$m(A) = 0$	$m(A) = 0$
	$m(B) = 0$	$m(B) = 0$	$m(B) = 0$	$m(B) = 0$
	$m(C) = 1$	$m(C) = 1$	$m(C) = 1$	$m(C) = 1$
	$m(\Theta) = 0$	$m(\Theta) = 0$	$m(\Theta) = 0$	$m(\Theta) = 0$
Yager ^[7]	$m(A) = 0$	$m(A) = 0$	$m(A) = 0$	$m(A) = 0$
	$m(B) = 0$	$m(B) = 0$	$m(B) = 0$	$m(B) = 0$
	$m(C) = 0.099000$	$m(C) = 0.032997$	$m(C) = 0.000330$	$m(C) = 0.000016$
	$m(\Theta) = 0.901000$	$m(\Theta) = 0.967003$	$m(\Theta) = 0.999670$	$m(\Theta) = 0.999984$
孙全等 ^[31]	$m(A) = 0.164679$	$m(A) = 0.188782$	$m(A) = 0.283631$	$m(A) = 0.345473$
	$m(B) = 0.001830$	$m(B) = 0.052545$	$m(B) = 0.045273$	$m(B) = 0.044291$
	$m(C) = 0.298444$	$m(C) = 0.250845$	$m(C) = 0.183997$	$m(C) = 0.163894$
	$m(\Theta) = 0.535047$	$m(\Theta) = 0.507828$	$m(\Theta) = 0.487099$	$m(\Theta) = 0.446341$
邓勇等 ^[32]	$m(A) = 0.164679$	$m(A) = 0.191374$	$m(A) = 0.293611$	$m(A) = 0.364009$
	$m(B) = 0.001830$	$m(B) = 0.053266$	$m(B) = 0.046866$	$m(B) = 0.047153$
	$m(C) = 0.298444$	$m(C) = 0.253835$	$m(C) = 0.190460$	$m(C) = 0.173439$
	$m(\Theta) = 0.535047$	$m(\Theta) = 0.501525$	$m(\Theta) = 0.469063$	$m(\Theta) = 0.415399$
李弼程等 ^[33]	$m(A) = 0.405450$	$m(A) = 0.397567$	$m(A) = 0.553167$	$m(A) = 0.622670$
	$m(B) = 0.004505$	$m(B) = 0.110657$	$m(B) = 0.088296$	$m(B) = 0.080659$
	$m(C) = 0.590045$	$m(C) = 0.491775$	$m(C) = 0.358537$	$m(C) = 0.296671$
王栋等 ^[9]	$m(A) = 0.409247$	$m(A) = 0.371308$	$m(A) = 0.542287$	$m(A) = 0.610227$
	$m(B) = 0.004547$	$m(B) = 0.103349$	$m(B) = 0.086559$	$m(B) = 0.079047$
	$m(C) = 0.505543$	$m(C) = 0.431777$	$m(C) = 0.351167$	$m(C) = 0.290727$
	$m(\Theta) = 0.080663$	$m(\Theta) = 0.093566$	$m(\Theta) = 0.019987$	$m(\Theta) = 0.019999$
熊彦铭等 ^[18]	$m(A) = 0.246347$	$m(A) = 0.304055$	$m(A) = 0.730667$	$m(A) = 0.967549$
	$m(B) = 0.002394$	$m(B) = 0.153369$	$m(B) = 0.038400$	$m(B) = 0.005317$
	$m(C) = 0.278064$	$m(C) = 0.298275$	$m(C) = 0.101213$	$m(C) = 0.015495$
	$m(\Theta) = 0.473196$	$m(\Theta) = 0.244301$	$m(\Theta) = 0.129719$	$m(\Theta) = 0.011638$
韩德强等 ^[15]	$m(A) = 0.405365$	$m(A) = 0.359450$	$m(A) = 0.825805$	$m(A) = 0.982867$
	$m(B) = 0.000050$	$m(B) = 0.072132$	$m(B) = 0.003393$	$m(B) = 0.000931$
	$m(C) = 0.594585$	$m(C) = 0.568418$	$m(C) = 0.170802$	$m(C) = 0.016202$
卢正才等 ^[19]	$m(A) = 0.405450$	$m(A) = 0.407566$	$m(A) = 0.663919$	$m(A) = 0.741408$
	$m(B) = 0.004505$	$m(B) = 0.145214$	$m(B) = 0.067720$	$m(B) = 0.057437$
	$m(C) = 0.590045$	$m(C) = 0.447219$	$m(C) = 0.123781$	$m(C) = 0.100032$
	$m(\Theta) = 0.000000$	$m(\Theta) = 0.000000$	$m(\Theta) = 0.144580$	$m(\Theta) = 0.101123$
李文立等 ^[20]	$m(A) = 0.405450$	$m(A) = 0.414626$	$m(A) = 0.731444$	$m(A) = 0.823203$
	$m(B) = 0.004505$	$m(B) = 0.153884$	$m(B) = 0.076423$	$m(B) = 0.056724$
	$m(C) = 0.590045$	$m(C) = 0.431490$	$m(C) = 0.192132$	$m(C) = 0.120073$
本文方法 (中立, 冲突, 其他各为一类, 共三类)	$m(A) = 0.405450$	$m(A) = 0.436692$	$m(A) = 0.653930$	$m(A) = 0.706287$
	$m(B) = 0.004505$	$m(B) = 0.131068$	$m(B) = 0.099187$	$m(B) = 0.084080$
	$m(C) = 0.590045$	$m(C) = 0.292524$	$m(C) = 0.154386$	$m(C) = 0.116864$
	$m(\Theta) = 0.000000$	$m(\Theta) = 0.139715$	$m(\Theta) = 0.092497$	$m(\Theta) = 0.092769$
本文方法 (中立与冲突一 类, 其他单独一 类, 共两类)	$m(A) = 0.405450$	$m(A) = 0.506036$	$m(A) = 0.763036$	$m(A) = 0.806116$
	$m(B) = 0.004505$	$m(B) = 0.131809$	$m(B) = 0.090757$	$m(B) = 0.073923$
	$m(C) = 0.590045$	$m(C) = 0.312116$	$m(C) = 0.145817$	$m(C) = 0.108716$
	$m(\Theta) = 0.000000$	$m(\Theta) = 0.050039$	$m(\Theta) = 0.000390$	$m(\Theta) = 0.011245$
本文方法 (中立与其他一 类, 冲突单独一 类, 共两类)	$m(A) = 0.405450$	$m(A) = 0.513555$	$m(A) = 0.896306$	$m(A) = 0.935899$
	$m(B) = 0.004505$	$m(B) = 0.108166$	$m(B) = 0.001380$	$m(B) = 0.000188$
	$m(C) = 0.590045$	$m(C) = 0.309644$	$m(C) = 0.042928$	$m(C) = 0.018861$
	$m(\Theta) = 0.000000$	$m(\Theta) = 0.068635$	$m(\Theta) = 0.059386$	$m(\Theta) = 0.045052$

表 4 两冲突证据参与合成

Table 4 The comparison of results when two conflictive evidences take part in combination

融合规则	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 ^[1]	$m(A) = 0.193548$	$m(A) = 0.110429$	$m(A) = 0.463235$	$m(A) = 0.577982$
	$m(B) = 0.032258$	$m(B) = 0.006135$	$m(B) = 0.007353$	$m(B) = 0.001668$
	$m(C) = 0.774194$	$m(C) = 0.883436$	$m(C) = 0.529412$	$m(C) = 0.420350$
	$m(\Theta) = 0$	$m(\Theta) = 0$	$m(\Theta) = 0$	$m(\Theta) = 0$
Yager ^[7]	$m(A) = 0.060000$	$m(A) = 0.018000$	$m(A) = 0.012600$	$m(A) = 0.006930$
	$m(B) = 0.010000$	$m(B) = 0.001000$	$m(B) = 0.000200$	$m(B) = 0.000020$
	$m(C) = 0.240000$	$m(C) = 0.144000$	$m(C) = 0.014400$	$m(C) = 0.005040$
	$m(\Theta) = 0.690000$	$m(\Theta) = 0.837000$	$m(\Theta) = 0.972800$	$m(\Theta) = 0.988010$
孙全等 ^[31]	$m(A) = 0.181130$	$m(A) = 0.171118$	$m(A) = 0.229516$	$m(A) = 0.244791$
	$m(B) = 0.044609$	$m(B) = 0.046936$	$m(B) = 0.063999$	$m(B) = 0.063450$
	$m(C) = 0.430348$	$m(C) = 0.404301$	$m(C) = 0.244076$	$m(C) = 0.232329$
	$m(\Theta) = 0.343913$	$m(\Theta) = 0.377645$	$m(\Theta) = 0.462408$	$m(\Theta) = 0.459431$
邓勇等 ^[32]	$m(A) = 0.181131$	$m(A) = 0.171694$	$m(A) = 0.230189$	$m(A) = 0.245566$
	$m(B) = 0.044609$	$m(B) = 0.047108$	$m(B) = 0.064197$	$m(B) = 0.063656$
	$m(C) = 0.430348$	$m(C) = 0.405279$	$m(C) = 0.244789$	$m(C) = 0.233069$
	$m(\Theta) = 0.343912$	$m(\Theta) = 0.375919$	$m(\Theta) = 0.460825$	$m(\Theta) = 0.457709$
李弼程等 ^[33]	$m(A) = 0.301500$	$m(A) = 0.297000$	$m(A) = 0.426040$	$m(A) = 0.451535$
	$m(B) = 0.079000$	$m(B) = 0.084700$	$m(B) = 0.121800$	$m(B) = 0.118581$
	$m(C) = 0.619500$	$m(C) = 0.618300$	$m(C) = 0.452160$	$m(C) = 0.429884$
	$m(\Theta) = 0$	$m(\Theta) = 0$	$m(\Theta) = 0$	$m(\Theta) = 0$
王栋等 ^[9]	$m(A) = 0.246949$	$m(A) = 0.195602$	$m(A) = 0.304180$	$m(A) = 0.318429$
	$m(B) = 0.066985$	$m(B) = 0.056481$	$m(B) = 0.088413$	$m(B) = 0.084366$
	$m(C) = 0.460920$	$m(C) = 0.389224$	$m(C) = 0.322391$	$m(C) = 0.303802$
	$m(\Theta) = 0.225146$	$m(\Theta) = 0.358694$	$m(\Theta) = 0.285016$	$m(\Theta) = 0.293403$
熊彦铭等 ^[18]	$m(A) = 0.281988$	$m(A) = 0.218618$	$m(A) = 0.438538$	$m(A) = 0.622042$
	$m(B) = 0.068915$	$m(B) = 0.047878$	$m(B) = 0.062424$	$m(B) = 0.032843$
	$m(C) = 0.436950$	$m(C) = 0.720752$	$m(C) = 0.442207$	$m(C) = 0.333968$
	$m(\Theta) = 0.212147$	$m(\Theta) = 0.012752$	$m(\Theta) = 0.056831$	$m(\Theta) = 0.011147$
韩德强等 ^[15]	$m(A) = 0.281609$	$m(A) = 0.167116$	$m(A) = 0.417912$	$m(A) = 0.526216$
	$m(B) = 0.022989$	$m(B) = 0.007668$	$m(B) = 0.023082$	$m(B) = 0.008480$
	$m(C) = 0.695402$	$m(C) = 0.825216$	$m(C) = 0.559006$	$m(C) = 0.465304$
卢正才等 ^[19]	$m(A) = 0.301500$	$m(A) = 0.290025$	$m(A) = 0.428742$	$m(A) = 0.463885$
	$m(B) = 0.079000$	$m(B) = 0.084700$	$m(B) = 0.119098$	$m(B) = 0.116411$
	$m(C) = 0.619500$	$m(C) = 0.625275$	$m(C) = 0.452160$	$m(C) = 0.419705$
李文立等 ^[20]	$m(A) = 0.301500$	$m(A) = 0.282364$	$m(A) = 0.434511$	$m(A) = 0.473329$
	$m(B) = 0.079000$	$m(B) = 0.079232$	$m(B) = 0.114697$	$m(B) = 0.113191$
	$m(C) = 0.619500$	$m(C) = 0.638404$	$m(C) = 0.450792$	$m(C) = 0.413480$
本文方法	$m(A) = 0.301500$	$m(A) = 0.231301$	$m(A) = 0.463037$	$m(A) = 0.592673$
	$m(B) = 0.079000$	$m(B) = 0.045052$	$m(B) = 0.050928$	$m(B) = 0.028049$
	$m(C) = 0.619500$	$m(C) = 0.670163$	$m(C) = 0.485165$	$m(C) = 0.366394$
	$m(\Theta) = 0.000000$	$m(\Theta) = 0.053484$	$m(\Theta) = 0.000870$	$m(\Theta) = 0.012884$

表 5 韩德强方法的进一步验证

Table 5 Further validation on Han deqiang's method

合成次序	$m(A)$	$m(B)$	$m(C)$
$\overrightarrow{m_1, m_2, m_3, m_4, m_5}$	0.526216	0.008480	0.465304
$\overrightarrow{m_1, m_3, m_4, m_2, m_5}$	0.607857	0.006467	0.385676
$\overrightarrow{m_1, m_3, m_4, m_5, m_2}$	0.574171	0.005169	0.420660

参与合成,从证据赋值情况看,命题 A 应当是真的,于是 5 个证据源最终的合成结果中只有熊彦铭的 $m(A) = 0.622042$,比本文的 $m(A) = 0.592673$ 略高,而韩德强的只有 $m(A) = 0.526216$,明显比本文的低.当然,韩德强的方法比较适合序贯融合,不同的融合次序,最后的结果不同,比如,仍然采用例 9 的五个证据源,但采用不同的合成次序 1) $\overrightarrow{m_1, m_2, m_3, m_4, m_5}$; 2) $\overrightarrow{m_1, m_3, m_4, m_2, m_5}$; 3) $\overrightarrow{m_1, m_3, m_4, m_5, m_2}$,通过考虑韩德强的方法,实验结果如表 5 所示.命题 A 的合成的概率赋值分别是 0.526216, 0.607857, 0.574171,可见结果不一致.

6 总结与展望

本文针对 Dempster 合成规则在处理高度冲突时,出现的反直觉现象.深入分析了证据冲突产生的原因和当前衡量证据冲突的测度存在的不足,提出了同时考虑证据自冲突或自矛盾和外部冲突的衡量测度,对证据源进行修正,并利用 ISODATA 聚类方法进行聚类,将类内、类外证据分别进行前后组合,并通过大量的算例,同其他方法进行深入比较,取得了较好的合成效果,较好地解决了冲突证据合成出现的反直觉现象,较其他方法优势比较明显.当然,目前该方法仍存在一些不足可以继续完善: 1) 在对证据可靠度计算的时候,可以考虑更多因素,以便更准确地衡量证据聚类的可靠度; 2) 本文采用的聚类方法,当聚类特性模糊时,参数设置的不同,会对聚类,以及最终合成结果产生一定的影响,希望将来有更好的聚类办法,减少这方面的影响; 3) 当面向海量证据融合时,在本文方法的基础上,如何进一步考虑近似推理组合算法^[34-35]; 4) 尝试在某些领域^[36-37]应用该方法.

References

- 1 Shafer G. *A Mathematical Theory of Evidence*. Princeton: Princeton University Press, 1976.
- 2 Zadeh L A. Review of books: a mathematical theory of evidence. *AI Magazine*, 1984, **5**(3): 81-83
- 3 Haenni R. Are alternatives to Dempster's rule of combination real alternatives. *Information Fusion*, 2002, **3**(4):

237-239

- 4 Bian Zhao-Qi, Zhang Xue-Gong. *Pattern Recognition*. Beijing: Tsinghua University Press, 2000.
(边肇祺, 张学工. 模式识别. 北京: 清华大学出版社, 2000.)
- 5 Martin A, Osswald C. Toward a combination rule to deal with partial conflict and specificity in belief functions theory. In: *Proceedings of the 10th International Conference on Information Fusion*. Quebec, Canada: IEEE, 2007. 1-8
- 6 Lefevre E, Colot O, Vannoorenberghe P. Belief function combination and conflict management. *Information Fusion*, 2002, **3**(2): 149-162
- 7 Yager R R. On the D-S framework and new combination rules. *Information Sciences*, 1987, **41**(2): 93-138
- 8 Liu W R. Analyzing the degree of conflict among belief functions. *Artificial Intelligence*, 2006, **170**(11): 909-924
- 9 Wang Dong, Li Qi, Jiang Wen, Xu Xiao-Bin, Deng Yong. New method to combine conflict evidence based on pignistic probability distance. *Infrared and Laser Engineering*, 2009, **38**(1): 149-154
(王栋, 李齐, 蒋雯, 徐晓滨, 邓勇. 基于 pignistic 概率距离的冲突证据合成方法. 红外与激光工程, 2009, **38**(1): 149-154)
- 10 Deng Yong, Wang Dong, Li Qi, Zhang Ya-Juan. A new method to analyze evidence conflict. *Control Theory & Applications*, 2011, **28**(6): 839-844
(邓勇, 王栋, 李齐, 章雅娟. 一种新的证据冲突分析方法. 控制理论与应用, 2011, **28**(6): 839-844)
- 11 Quan Wen, Wang Xiao-Dan, Wang Jian, Zhang Yu-Xi. New combination rule of DST based on local conflict distribution strategy. *Acta Electronica Sinica*, 2012, **40**(9): 1880-1884
(权文, 王晓丹, 王坚, 张玉玺. 一种基于局部冲突分配的 DST 组合规则. 电子学报, 2012, **40**(9): 1880-1884)
- 12 He Bing, Hu Hong-Li. A modified DS evidence combination strategy. *Acta Aeronautica et Astronautica Sinica*, 2003, **24**(6): 559-562
(何兵, 胡红丽. 一种修正的 DS 证据融合策略. 航空学报, 2003, **24**(6): 559-562)
- 13 Yang J B, Xu D L. Evidential reasoning rule for evidence combination. *Artificial Intelligence*, 2013, **205**: 1-29
- 14 Smets P. Analyzing the combination of conflicting belief functions. *Information Fusion*, 2007, **8**(4): 387-412

- 15 Han De-Qiang, Han Chong-Zhao, Deng Yong, Yang Yi. Weighted combination of conflicting evidence based on evidence variance. *Acta Electronica Sinica*, 2011, **39**(3A): 153–157
(韩德强, 韩崇昭, 邓勇, 杨艺. 基于证据方差的加权证据组合. 电子学报, 2011, **39**(3A): 153–157)
- 16 Hu Li-Fang, Guan Xin, Deng Yong, He You. Cause-analysis for conflicting evidences in the generalized power space. *Control Theory and Applications*, 2011, **28**(12): 1717–1722
(胡丽芳, 关欣, 邓勇, 何友. 广义幂集空间中证据冲突的原因分析. 控制理论与应用, 2011, **28**(12): 1717–1722)
- 17 Smarandache F, Dezert J. *Advances and Applications of DSMT for Information Fusion*. Rehoboth, USA: American Research Press, Vol.2, 2006
- 18 Xiong Yan-Ming, Yang Zhan-Ping, Qu Xin-Fen. Novel combination method of conflict evidence based on evidential model modification. *Control and Decision*, 2011, **26**(6): 883–887
(熊彦铭, 杨战平, 屈新芬. 基于模型修正的冲突证据组合新方法. 控制与决策, 2011, **26**(6): 883–887)
- 19 Lu Zheng-Cai, Qin Zheng. General framework for evidence combination and its approach to highly conflicting evidence fusion. *Journal of Tsinghua University (Science and Technology)*, 2011, **51**(11): 1701–1705, 1716
(卢正才, 覃征. 证据合成的一般框架及高度冲突证据合成方法. 清华大学学报 (自然科学版), 2011, **51**(11): 1701–1705, 1716)
- 20 Li Wen-Li, Guo Kai-Hong. Combination rules of D-S evidence theory and conflict problem. *Systems Engineering-Theory and Practice*, 2010, **30**(8): 1422–1432
(李文立, 郭凯红. D-S 证据理论合成规则及冲突问题. 系统工程理论与实践, 2010, **30**(8): 1422–1432)
- 21 Wang Lian-Feng, Song Jian-She, Zhu Yu, Cao Ji-Ping. Evidence combination based on fuzzy clustering analysis. *Systems Engineering and Electronics*, 2013, **35**(1): 113–119
(王连锋, 宋建社, 朱昱, 曹继平. 基于模糊聚类分析的证据组合. 系统工程与电子技术, 2013, **35**(1): 113–119)
- 22 Quan Wen, Wang Xiao-Dan, Shi Zhao-Hui. Conflict evidence combination strategy based on evidence classification. *Journal of Air Force Engineering University (Natural Science)*, 2012, **13**(3): 80–84
(权文, 王晓丹, 史朝辉. 基于证据分类的冲突证据合成方法. 空军工程大学学报 (自然科学版), 2012, **13**(3): 80–84)
- 23 Wang Jin-Hua, Wu Di, Cao Jie, Li Jun. Weighted combination of conflicting evidence based on evidence classification. *Computer Science*, 2013, **40**(1): 247–250
(王进花, 吴迪, 曹洁, 李军. 基于证据分类的加权冲突证据组合. 计算机科学, 2013, **40**(1): 247–250)
- 24 Jousselme A L, Maupin P. Distances in evidence theory: comprehensive survey and generalizations. *International Journal of Approximate Reasoning*, 2012, **53**(2): 118–145
- 25 Jousselme A L, Grenier D, Bossé É. A new distance between two bodies of evidence. *Information Fusion*, 2001, **2**(2): 91–101
- 26 Li X, Dezert J, Smarandache F. Evidence supporting measure of similarity for reducing the complexity in information fusion. *Information Sciences*, 2011, **181**(10): 1818–1835
- 27 Smets P. Data fusion in the transferable belief model. In: *Proceedings of the 3rd International Conference on Information Fusion*. Paris, France: IEEE, 2000. 21–33
- 28 Wen C L, Wang Y C, Xu X B. Fuzzy information fusion algorithm of fault diagnosis based on similarity measure of evidence. In: *Proceedings of the 5th international symposium on Neural Networks: Advances in Neural Networks, Part II*. Berlin, Heidelberg: Springer, 2008. 506–515
- 29 Liu Z, Dezert J, Pan Q, Mercier G. Combination of sources of evidence with different discounting factors based on a new dissimilarity measure. *Decision Support Systems*, 2011, **52**(1): 133–141
- 30 Gu Zhao-Sheng. *The Research and Application of Compositive Mathematical Methods to Water Quality Analysis, Simulation and Forecast in Reservoirs and Lakes* [Ph. D. dissertation], Jilin University, China, 2006.
(谷照升. 水库湖泊水质分析、模拟与预测的综合数学方法及其应用 [博士学位论文]. 吉林大学, 中国, 2006.)
- 31 Sun Quan, Ye Xiu-Qing, Gu Wei-Kang. A new combination rules of evidence theory. *Acta Electronica Sinica*, 2000, **28**(8): 117–119
(孙全, 叶秀清, 顾伟康. 一种新的基于证据理论的合成公式. 电子学报, 2000, **28**(8): 117–119)
- 32 Deng Yong, Shi Wen-Kang. A modified combination rule of evidence theory. *Journal of Shanghai Jiaotong University*, 2003, **37**(8): 1275–1278
(邓勇, 施文康. 一种改进的证据推理组合规则. 上海交通大学学报, 2003, **37**(8): 1275–1278)
- 33 Li Bi-Cheng, Wang Bo, Wei Jun, Qian Ceng-Bo, Huang Yu-Qi. An efficient combination rule of evidence theory. *Journal of Data Acquisition and Processing*, 2002, **17**(1): 33–36
(李弼程, 王波, 魏俊, 钱曾波, 黄玉琪. 一种有效的证据理论合成公式. 数据采集与处理, 2002, **17**(1): 33–36)
- 34 Li Xin-De, Dezert J, Huang Xin-Han, Meng Zheng-Da, Wu Xue-Jian. A fast approximate reasoning method in hierarchical DSMT (A). *Acta Electronica Sinica*, 2010, **38**(11):

2566–2572

(李新德, Dezert J, 黄心汉, 孟正大, 吴雪建. 一种快速分层递阶 DSmT 近似推理融合方法 (A). 电子学报, 2010, **38**(11): 2566–2572)

- 35 Li Xin-De, Yang Wei-Dong, Wu Xue-Jian, Dezert J. A fast approximate reasoning method in hierarchical DSmT (B). *Acta Electronica Sinica*, 2011, **39**(3A): 31–36

(李新德, 杨伟东, 吴雪建, Dezert J. 一种快速分层递阶 DSmT 近似推理融合方法 (B). 电子学报, 2011, **39**(3A): 31–36)

- 36 Li Xin-De, Yang Wei-Dong, Dezert J. An airplane image target's Multi-feature fusion recognition method. *Acta Automatica Sinica*, 2012, **38**(8): 1298–1307

(李新德, 杨伟东, Dezert J. 一种飞机图像目标多特征信息融合识别方法. 自动化学报, 2012, **38**(8): 1298–1307)

- 37 Li Xin-De, Pan Jin-Dong, Dezert J. A target recognition algorithm of sequential aircrafts based on DSmT and HMM. *Acta Automatica Sinica*, 2014, **40**(12): 2862–2876

(李新德, 潘锦东, Dezert J. 一种基于 DSmT 和 HMM 的序列飞机目标识别算法. 自动化学报, 2014, **40**(12): 2862–2876)



李新德 东南大学自动化学院副教授. 主要研究方向为智能机器人, 人机交互, 机器感知, 信息融合, 不确定推理和机器视觉. 本文通信作者.

E-mail: xindeli@seu.edu.cn

(**LI Xin-De** Associate professor at the School of Automation, Southeast University. His research interest covers intelligent robot, human-robot interaction, robot perception, information fusion, reasoning under uncertainty, and machine vision. Corresponding author of this paper.)



王丰羽 东南大学自动化学院硕士研究生. 主要研究方向为信息融合和不确定推理.

E-mail: treexiaohei@163.com

(**WANG Feng-Yu** Master student at the School of Automation, Southeast University. His research interest covers information fusion and reasoning under uncertainty.)