



文章编号:1001-9081(2018)10-2794-07

DOI:10.11772/j.issn.1001-9081.2018030532

降低相似度碰撞的证据融合方法

王 剑^{1*}, 张志勇², 乔阔远²

(1. 郑州大学 信息工程学院, 郑州 450001; 2. 河南科技大学 信息工程学院, 河南 洛阳 471003)

(* 通信作者电子邮箱 iejwang@zzu.edu.cn)

摘要:针对相似度碰撞引发证据融合结果错误的问题,提出一种新的证据融合方法。首先,提取证据的焦元序列特征并将其转化为排序矩阵以弥补证据相似度易碰撞的不足;其次,联合证据的排序矩阵和信息熵完成对证据权重的确定;最后,生成归一化证据(MAE)并使用 Dempster 融合公式将 MAE 融合 $n-1$ 次获得最终的结果。基于在线的鸢尾花数据集对证据平均融合方法、余弦相似度证据融合方法、证据距离融合方法和证据信誉度融合方法进行了花类型识别准确性的 F-Score 对比,上述四种方法的 F-Score 分别为 0.84, 0.88, 0.88 和 0.88, 而所提方法的 F-Score 为 0.91。实验结果表明,所提方法的决策准确率更高,融合结果更加可靠,能为证据决策提供了有效的解决方案。

关键词:D-S 证据理论;证据融合方法;相似度碰撞;基本信任分配;冲突管理

中图分类号:TP181 **文献标志码:**A

Evidence combination rule with similarity collision reduced

WANG Jian^{1*}, ZHANG Zhiyong², QIAO Kuoyuan²

(1. School of Information Engineering, Zhengzhou University, Zhengzhou Henan 450001, China;

2. School of Information Engineering, Henan University of Science and Technology, Luoyang Henan 471003, China)

Abstract: Aiming at the problem of decision error caused by similarity collision in evidence theory, a new combination rule for evidence theory was proposed. Firstly, the features of focal-element sequence in evidence were extracted and converted into a sort matrix to reduce similarity collision. Secondly, the weight of each evidence was determined based on sort matrix and information entropy. Finally, the Modified Average Evidence (MAE) was generated based on the evidence set and evidence weight, and the combination result was obtained by combing MAE for $n-1$ times by using Dempster combination rule. The experimental results on the online dataset Iris show that the F-Score of average-based combination rule, similarity-based combination rule, evidence distance-based combination rule, evidence-credit based combination rule and the proposed method are 0.84, 0.88, 0.88, 0.88 and 0.91. Experimental results show that the proposed method has higher accuracy of decision making and more reliable combination results, which can provide an efficient solution for decision-making based on evidence theory.

Key words: D-S evidence theory; combination rule method; similarity collision; Basic Probability Assignment (BPA); conflict management

0 引言

证据理论是由 Dempster^[1]和 Shafer^[2]提出的一种有效的不确定性推理方法^[3],被广泛用于最优决策^[4-6]、可靠分析^[7-9]、关系度量^[10]、最优计算^[11-13]等领域。然而在现实环境中,由于传感器的故障或信任函数模型的不准确会导致证据之间互相冲突,进而会引起融合结果的反直观现象^[14]。

为了能在冲突证据存在的情况下依然获得正确的融合结果,证据的融合公式必须能适应冲突证据的存在^[15-16]。Dubois 等^[17]提出了一种新的证据融合方法,它将证据的交集部分转化为并集部分,以降低证据冲突程度;但该方法在冲突程度高时表现较差^[14]。Murphy^[18]将所有待融合的证据求平均得出平均证据,然后将平均证据使用 Dempster 融合方法融

合自身 $n-1$ 次得到组合结果。然而,文献[14]指出:不同证据对融合结果的影响不同,应为不同证据分配不同的权重。为了获得证据权重,一些学者提出了证据之间的相似度,并将相似度转化为证据权重^[19]。Cuzzolin^[20]通过将证据体映射成几何空间中特殊线段的方式解释了证据距离的几何意义,为证据距离和证据相似度奠定了理论基础。Jousselme 等^[21]提出了一种基于证据距离的证据相似度计算方法,但没有提出将证据相似度转化为证据权重的方案。Wen 等^[22]提出了一种简化的、基于向量余弦值的证据相似度计算方法,却忽略了证据中焦元个数对相似度计算结果的影响。Wang 等^[23]通过修正证据权重改进了 Murphy^[18]的融合规则,但存在相似度碰撞的缺陷。文献[24]中提出了一种降低相似度碰撞的方法,但是没有提出与之相适应的证据融合方法。

收稿日期:2018-03-15;修回日期:2018-05-05;录用日期:2018-05-08。

基金项目:国家自然科学基金资助项目(61772174, 61370220);河南省自然科学基金资助项目(162300410094)。

作者简介:王剑(1978—),女,河南洛阳人,副教授,博士,CCF 会员,主要研究方向:多媒体社交网络、信息安全、可信计算和使用控制;张志勇(1975—),男,河南新乡人,教授,博士,IEEE 高级会员,主要研究方向:多媒体社交网络、数字版权管理、可信计算和使用控制; 乔阔远(1991—),男,河南濮阳人,硕士研究生,主要研究方向:信任度计算、数据挖掘、信息安全、多源信息融合。



证据的权重除了受到冲突程度的影响之外,还受到证据不确定程度的影响。证据理论中对证据不确定性程度的计算过程被称为不确定性度量(Ambiguity Measurement, AM)^[25]。

陈金广等^[26]提出一种基于经验值 α 和多义度的不确定性度量方法,Wang 等^[27]提出了一种基于置信区间的证据不确定性测度方法,Kilr 等^[28]提出了一种基于证据中基本信任分配函数数值分布的不确定性度量方法,韩德强等^[29]、Deng 等^[30]分别提出了两种基于证据距离和证据不确定性的证据权重确定方法,Wang 等^[31]提出一种基于不确定性度量和软集合的证据决策方法。

以上研究都是在冲突证据被正确识别的前提下,基于证据的不确定性程度完成对证据权重的修正。虽然 AM 在一定程度上能获得更优的融合结果,但其依赖的冲突证据识别本身就存在相似度碰撞的缺陷,即证据的冲突强度无法被证据相似度正确识别。而无法识别的原因是由于相似度计算的缺陷导致两组不同的证据相似度计算结果相同,进而使高冲突证据被误认为是低冲突证据,引起证据支持度的错误。

1 预备知识

1.1 证据体和基本信任分配函数

在证据理论中,用基本概率赋值函数(Basic Probability Assignment, BPA)来表示初始信任度分配,用信任函数来表示每个命题的信任度。基本概率赋值函数也被称为基本信任分配函数(Basic Belief Assignment, BBA)。对于辨识框架 θ ,问题域中任意命题 A 都应属于幂集 2^θ ,即 A 是 2^θ 的子集。若用 $m(A)$ 代表命题 A 的基本概率赋值函数,当 $m(A) > 0$ 时,称 A 为证据的焦元。信任函数 $bel(A)$ 表示证据完全支持 A 的程度,似真函数 $pl(A)$ 表示证据不反对命题 A 的程度,区间 $[bel(A), pl(A)]$ 构成证据的不确定区间,如图 1 所示。

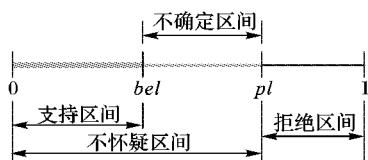


图 1 信任概率范围

Fig. 1 Trust probabilistic scope in evidence

图 1 中,函数 $bel(A)$ 是当前证据对焦元 A 的信任函数,而 $pl(A)$ 被称为焦元 A 的似真函数。一般地,对于任意一个焦元 B , $bel(B)$ 和 $pl(B)$ 之间有以下性质:

$$\begin{cases} bel(\emptyset) = 0 \\ bel(\theta) = 1 \end{cases} \quad (1)$$

$$bel(B) + bel(B') \leq 1; \forall B \subseteq \theta \quad (2)$$

$$bel(B) \leq pl(B) \quad (3)$$

式(2)中: B' 是焦元 B 的补集。需要注意的是,焦元之间可能存在交集,而所有互斥的焦元构成证据的辨识框架 $\theta = \{h_1, h_2, \dots, h_n\}$,其他所有的焦元 f 都是辨识框架的子集 $f \in P = 2^\theta$ 。由于辨识框架中的元素均是单元素焦元,而实际参与融合的证据往往是多焦元证据,定义 $m(A) = \sum_{B \subseteq A} bel(B)$,其中: $m(A)$ 为当前证据对 A 的基本信任分配函数; $bel(A)$ 代表焦元 A 的子焦元的信任函数。将不同焦元的基本信任分配函数的集合记为证据的证据体(Body of Evidence, BOE),并且基于辨识框架 θ ,证据的共性函数的定义为:

$$Q(A) = \sum_{A \subseteq B} m(B); \forall A \subseteq \theta \quad (4)$$

式(4)中的共性函数描述了一个证据中的所有基本信任分配函数对一个指定的焦元的支持程度。基本信任函数、似然函数和共性函数为证据计算的基石,三者具有以下关系:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} bel(B); \forall A \subseteq \theta \quad (5)$$

$$bel(A) = \sum_{B \subseteq A} (-1)^{|B|} Q(B); \forall A \subseteq \theta \quad (6)$$

$$Q(A) = \sum_{B \subseteq A} (-1)^{|B|} bel(\bar{B}); \forall A \subseteq \theta \quad (7)$$

1.2 证据的相似度和支持度

文献[24]基于用于描述证据差别的证据距离,提出了证据相似度函数应满足的三个条件:非负性、无序性和三角性。其中:非负性代表相似度结果永远为正;无序性代表计算结果应与参数顺序无关;三角性代表间接相似度需要大于直接相似度,以满足值域的紧凑性。

基于以上的性质,许多学者提出了证据之间的相似度计算公式,其中由文献[21]和文献[22]的方法应用最为广泛。文献[22]中提出的相似度计算公式是基于两个证据之间的余弦值,具有运算快的特点。文献[21]中提出的相似度计算公式中,将每个焦元中的元素个数进行了计算,相对于文献[22]方法,考虑因素更加全面。本文使用的相似度计算方法是文献[21]中提出的相似度方法,其定义如下:

$$sim(m_i, m_j) = 1 - d_{i,j} = 1 - \sqrt{\frac{1}{2} (\mathbf{v}_i - \mathbf{v}_j)^T \mathbf{D} (\mathbf{v}_i - \mathbf{v}_j)} \quad (8)$$

其中: $\mathbf{v}_i - \mathbf{v}_j$ 是证据 m_i 和 m_j 之间的向量差; \mathbf{D} 为一个基于焦元一阶范数的矩阵,且 $\mathbf{D} = \frac{|A \cap B|}{|A \cup B|}$ 。基于证据的相似度定义证据的支持度如下:

$$sup(m_i) = \sum_{m \in ms; m \neq m_i} sim(m, m_i) \quad (9)$$

其中: ms 为参与融合的证据集合, ms 中的所有证据必须和 m_i 处于同一个辨识框架之下。证据的支持度是证据相似性的整体体现,一个证据与其他证据的相似度之和越高,其支持度越大。

1.3 证据的信息熵

相似度可以完成证据的冲突检测,而冲突检测还不足以确定证据权重,因为每个证据的不确定性程度不同,无论一个证据冲突程度的大小,如果其不确定性程度很大,就应该削弱其在融合过程中的权重。

由于 Deng 等^[30]提出的方法具有计算简便和适应性强的优点,本文采用该方法完成证据的不确定性度量,其定义为:

$$En(m) = - \sum_{A \subseteq \theta} m(A) \ln \left(\frac{m(A)}{2^{|A|} - 1} \right) \quad (10)$$

其中: θ 代表当前证据处于的辨识框架; $|A|$ 代表焦元 A 的一阶范式。

1.4 基本融合公式

假设在辨识框架 θ 中证据的基本信任分配函数分别为 $bel_1, bel_2, \dots, bel_n$,其融合形式为 $bel_{\oplus} = bel_1 \oplus bel_2 \oplus \dots \oplus bel_n$,融合结果 bel_{\oplus} 被称为是其他基本信任分配函数的直



和^[14]。证据的基本融合公式如下：

$$m_{\oplus}(\emptyset) = 0 \quad (11)$$

$$m_{\oplus}(A) = \frac{\sum_{A_{i1} \cap \dots \cap A_{in}} m_1(A_{i1})m_2(A_{i2}) \dots m_n(A_{in})}{1 - C} \quad (12)$$

$$C = \sum_{A_{i1} \cap \dots \cap A_{in} = \emptyset} m_1(A_{i1})m_2(A_{i2}) \dots m_n(A_{in}) \quad (13)$$

其中 C 被称为冲突程度。根据式(11)~(13) 可证明基本的融合公式符合交换律和结合律^[4]。融合结果的共性函数 $Q_{\oplus}(A)$ 为：

$$Q_{\oplus}(A) = \frac{1}{C} \prod_{i=1}^n Q_i(A); \forall A \subseteq \theta \quad (14)$$

2 抗相似度碰撞融合算法

2.1 信任转移模型

在已有的证据融合算法中,融合模型主要分为两部分:信任转移模型和决策模型。在信任转移模型中,为每个证据确定其在融合过程中的权重,每个证据的权重表示其可信程度。而可信程度的度量主要是基于证据之间的相似度和有用性的度量,其架构如图 2 所示。

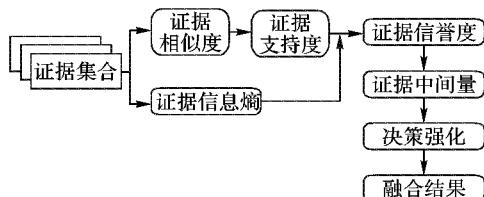


图 2 传统的证据融合架构

Fig. 2 Traditional combination scheme for evidence

在图 2 所示的架构中,证据的权重建立在证据的相似度计算之上,如果相似度不能把证据之间的差异正确地映射为相似度数值的差异,将导致证据支持度的偏差,进而影响最终的融合结果。而在最近关于证据理论的研究^[14]中,已经发现证据相似度存在相似度碰撞的缺陷。为了应对以上碰撞,本文使用排序矩阵修正证据支持度,图 3 为本文算法的架构。

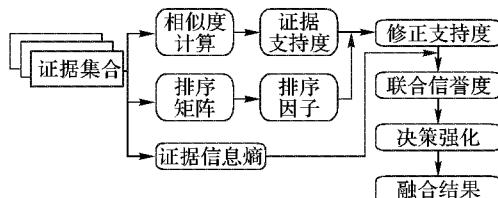


图 3 本文的证据融合架构

Fig. 3 Proposed combination scheme for evidence

在图 3 所示的本文架构中,证据焦元序列被转化为排序矩阵,通过检测每个矩阵焦元序列的差来检测易产生碰撞的证据,同时基于信息熵和排序矩阵的差完成证据权重的确定,并获得融合结果。

2.2 相似度碰撞和支持度修正

假设 $ms = \{m_1, m_2, m_3, m_4\}$ 是处于同一辨识框架下的四个证据,每个证据的证据体如表 1 所示。

根据式(8),证据 m_1 和 m_2 之间的相似度为:

$$sim(m_1, m_2) = 1 - \sqrt{\frac{1}{2}(\mathbf{v}_1 - \mathbf{v}_2)^T D(\mathbf{v}_1 - \mathbf{v}_2)} = 0.7 \quad (15)$$

同样根据式(8),证据 m_3 和 m_2 之间的相似度为:

$$sim(m_3, m_2) = 1 - \sqrt{\frac{1}{2}(\mathbf{v}_3 - \mathbf{v}_2)^T D(\mathbf{v}_3 - \mathbf{v}_2)} = 0.7 \quad (16)$$

通过式(15)~(16) 可以看出, $sim(m_1, m_2) = sim(m_2, m_3)$, 但 $m_1 \neq m_3$ 。在本文中,将这种证据不相同但是相似度相同的情况称为相似度碰撞。基于证据的相似度,证据 m_1, m_2, m_3, m_4 的支持度分别为:

$$sup(m_1) = \sum_{m \neq m_1, m \in ms} sim(m_1, m) = 1.76 \quad (17)$$

$$sup(m_2) = \sum_{m \neq m_2, m \in ms} sim(m_2, m) = 2.24 \quad (18)$$

$$sup(m_3) = \sum_{m \neq m_3, m \in ms} sim(m_3, m) = 1.76 \quad (19)$$

$$sup(m_4) = \sum_{m \neq m_4, m \in ms} sim(m_4, m) = 2.16 \quad (20)$$

根据式(13),算出证据 m_1, m_2, m_3, m_4 两两之间的冲突程度如表 2 所示。

基于表 2 中的数据,统计出一个证据对其他证据冲突程度的和,如表 3 所示。

参照表 1 中证据 m_1 和 m_3 ,由于 m_1 和 m_3 的区别仅仅在于焦元 A 和焦元 C 的 BPA 互换,而 BPA 的大小未发生变化,导致式(15)和式(16)计算获得的相似度相同。由于证据的支持度是相似度的和,相似度未发生改变导致了证据 m_1 和证据 m_3 的支持度相同(式(17)、式(19))。参照表 2,由于证据 m_3 与其他证据的冲突程度之和最大,支持度应该尽可能小,说明相似度的碰撞影响了证据的支持度确定。

表 1 证据的证据体

Tab. 1 Body of each evidence

证据	焦元 A	焦元 B	焦元 C	焦元 AB
m_1	0.650	0.200	0.050	0.1
m_2	0.350	0.200	0.350	0.1
m_3	0.050	0.200	0.650	0.1
m_4	0.517	0.017	0.467	0.0

表 2 证据之间的冲突程度

Tab. 2 Conflict degrees between evidence

证据	与另一个证据的冲突程度			
	m_1	m_2	m_3	m_4
m_1	—	0.6250	0.7750	0.5922
m_2	0.6250	—	0.6250	0.6072
m_3	0.7750	0.6250	—	0.6222
m_4	0.5922	0.6072	0.6222	—

表 3 每个证据对其他证据的冲突程度和

Tab. 3 Summation of conflict degrees of each evidence with other evidence

证据	与其它证据冲突程度和	证据	与其它证据冲突程度和
m_1	1.9922	m_3	2.0222
m_2	1.8572	m_4	1.8216

参考式(8),证据的相似度仅仅计算了证据中 BPA 的大小区别,而不关心证据中 BPA 的大小关系。由于每个证据中所有 BPA 的和一定为 1,则每个证据中的所有 BPA 必定存在一个大小关系且大小关系不同的证据一定不同。以表 1 中的证据 m_1 为例,其大小关系为 $m_1(A) > m_1(B) \geq m_1(AB) >$



$m_1(C)$, 而证据 m_3 中 BPA 的大小关系为 $m_3(C) > m_3(B) \geq m_3(AB) > m_3(A)$ 。将证据的证据体看作是一个单列矩阵 M , 同时指定一个有序矩阵(Ordered Matrix, OM), 可以找到排序矩阵(Sort Matrix, SM), 使得该证据对应的单列矩阵 M 和 SM 的积为 OM 。对于给定的证据 m , 获得其排序矩阵的过程如算法 1 所示。

算法 1 计算证据的排序矩阵。

```

输入 证据  $m$ 。
输出 证据  $m$  的排序矩阵  $SM$ 。
定义证据  $m' = m$ , 定义有序集合  $F = \emptyset, J = \emptyset$ 
for  $\forall f \in P = 2^{\theta}$ :
    if  $f \notin F$ : 将  $f$  添加到  $F$  的最后
end for
for  $\exists m'(\alpha) \neq 0 (\alpha \in P)$ :
    将  $m'$  中基本信任分配函数最大的焦元记为  $f$ 
    将  $m'(f)$  添加到  $J$  的结尾, 并令  $m'(f) = 0$ 
end for
定义矩阵  $SM$ , 其维度为  $|P| \times |P|$ 
for  $\forall f \in P$ :
    将  $f$  在  $F$  中的序号记为  $i$ 
    将  $f$  在  $J$  中的序号记为  $j$ 
    将  $SM_{i,j}$  的值设为 1
end for
将矩阵  $SM$  的其他元素赋值为 0
返回矩阵  $SM$  作为证据  $m$  的排序矩阵

```

通过算法 1, 每个证据被映射为一个排序矩阵, 多个排序矩阵之间的差异可以映射为另一个矩阵差值矩阵(Difference Matrix, DM), 其计算方式为:

$$DM_i = SM_i - \frac{1}{n} \sum_{j=1}^n SM_j \quad (21)$$

证据的差值矩阵 DM 描述了一个证据在所有参与融合的证据中焦元序列的一致程度, 而通过指数曲线可以将差值矩阵映射为一个值域为 $[0, n]$ 的排序因子 F 。排序因子的定义为:

定义 1 假设 m_1, m_2, \dots, m_n 是处于同一辨识框架下的 n 个证据, 并且 DM_1, DM_2, \dots, DM_n 是证据 m_1, m_2, \dots, m_n 的差值矩阵, 则 m_i 的排序因子 F_i 为:

$$F_i = \frac{n \cdot \exp(-1 \cdot |DM_i|)}{\sum_{m \in ms} \exp(-1 \cdot |DM_m|)} \quad (22)$$

基于式(22)中的排序因子, 式(9)处的证据支持度被修正为:

$$msup(m_i) = \frac{F_i \times sup(m_i)}{\sum_{m \in ms} F_m \times sup(m)} \quad (23)$$

2.3 融合结果

假设 m_1, m_2, \dots, m_n 是处于同一辨识框架下的 n 个证据, $En(m_i)$ 是证据 m_i 的信息熵, 则证据 m_i 的综合支持度为:

$$gsup(m_i) = msup(m_i) \cdot \left(\frac{En(m_i)}{\sum_{m \in ms} En(m)} \right)^{uc(m_i)} \quad (24)$$

$$uc(m_i) = -msup(m_i) + \frac{1}{n} \cdot \sum_{m \in ms} msup(m) \quad (25)$$

证据的综合支持度是对每个证据在冲突检测和不确定性度量两个方面的信任度衡量, 冲突证据的权重通过支持度的计算被削弱, 相似度的碰撞被排序矩阵的计算削弱, 最后通过

证据的信息熵完成证据不确定性度量。由于证据的综合支持度的值域并非 $[0, 1]$, 需要将其转化为更适合作为权重的另一个参数综合信誉度 w , 可以通过证据的综合支持度的归一化完成:

$$w(m_i) = \frac{gsup(m_i)}{\sum_{m \in ms} gsup(m)} \quad (26)$$

而基于证据的权重, 定义归一化证据(Modified Average Evidence, MAE), 该证据将代表原始的所有证据和每个证据的冲突强度以及信息量。证据 MAE 的定义如下:

定义 2 假设 m_1, m_2, \dots, m_n 是处于同一辨识框架下的 n 个证据, 并且 $w(m)$ 是证据 m 的权重, 那么证据 MAE 为:

$$MAE = \sum_{m \in ms} m \times w(m) \quad (27)$$

MAE 是原始证据体和冲突检测的综合结果, 最终的融合结果可以通过将 MAE 使用基本公式融合 $n-1$ 次得到, 其流程如图 4 所示。

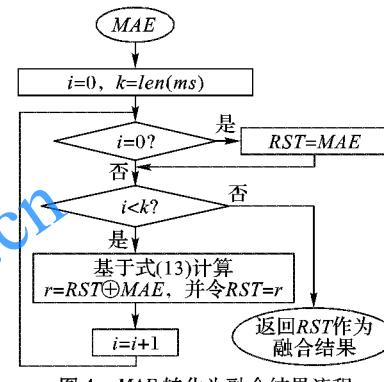


图 4 MAE 转化为融合结果流程

Fig. 4 Flow chart of converting MAE into combination result

在图 4 中, 首先定义变量 RST , 并将其赋值为 MAE , 通过 $n-1$ 循环融合 MAE 自身获得融合结果。

3 实验

在实验部分, 分别模拟了文献[18, 22-23, 29]中提出的融合算法, 并设计了两组实验检测本文方法的有效性。在第一组实验中, 使用文献[18, 22-23, 29]共同的比较方法, 即通过随机选取一组证据并比较融合结果中正确焦元的基本信任分配函数大小。在第二组实验中, 使用在线数据集 Iris (<http://archive.ics.uci.edu/ml/datasets/iris>) 作为输入, 使用文献[32]中提出的方法将每条记录转化为 4 条证据, 分别使用文献[18, 22-23, 29]中提出的融合算法和本文提出的融合算法将四条证据融合为融合结果并对鸢尾花类型作出判断, 通过统计准确率和召回率来判断本文方法的有效性。

3.1 比较融合结果中 BPA 大小

在一个由 6 个传感器组成的目标识别系统中, 已知总共有三种可能的目标类型 A, B, C 。随机抽取一个时刻传感器 $S_1, S_2, S_3, S_4, S_5, S_6$ 对同一目标的识别证据信息 $m_1, m_2, m_3, m_4, m_5, m_6$, 其中: m_1 代表传感器 S_1 返回的证据 1, m_2 代表传感器 S_2 返回的证据 2, 以此类推。每个证据如表 4 所示。

基于表 4 的数据, 证据之间的相似度如表 5 所示, 每个证据的排序因子和信息熵如表 6 所示。

根据表 5 和表 6 中的数据和式(26), 每个证据被不同方法确定的权重如表 7 所示。



表4 传感器返回证据
Tab. 4 List of evidence sent back by sensors

证据	焦元 A	焦元 B	焦元 C	焦元 AB
m_1	0.55	0.25	0.12	0.08
m_2	0.00	0.50	0.20	0.30
m_3	0.49	0.31	0.10	0.10
m_4	0.70	0.21	0.06	0.03
m_5	0.82	0.09	0.05	0.04
m_6	0.65	0.30	0.03	0.02

表5 证据之间的相似度(基于表4)

Tab. 5 Similarity between evidence corresponding to Tab. 4

证据	与另一个证据的相似度					
	m_1	m_2	m_3	m_4	m_5	m_6
m_1	1.000	0.579	0.936	0.888	0.775	0.912
m_2	0.579	1.000	0.635	0.472	0.358	0.528
m_3	0.936	0.635	1.000	0.837	0.720	0.888
m_4	0.888	0.472	0.837	1.000	0.879	0.925
m_5	0.775	0.358	0.720	0.879	1.000	0.809
m_6	0.912	0.528	0.888	0.925	0.809	1.000

表6 证据的排序因子和信息熵(基于表4)

Tab. 6 Sort factor and entropy of each evidence corresponding to Tab. 4

证据	排序因子	信息熵	证据	排序因子	信息熵
m_1	1.1288	1.7597	m_4	1.1288	1.2759
m_2	0.3558	1.9610	m_5	1.1288	1.0127
m_3	1.1288	1.8510	m_6	1.1288	1.2214

表7 不同方法确定的证据权重(基于表4)

Tab. 7 Evidence weight determined by different methods corresponding to Tab. 4

方法	证据权重					
	$w(m_1)$	$w(m_2)$	$w(m_3)$	$w(m_4)$	$w(m_5)$	$w(m_6)$
文献[18]方法	0.1666	0.1666	0.1666	0.1666	0.1666	0.1666
文献[23]方法	0.1872	0.1058	0.1826	0.1827	0.1549	0.1865
文献[29]方法	0.1873	0.1058	0.1827	0.1826	0.1549	0.1864
文献[22]方法	0.1903	0.0782	0.1911	0.1833	0.1704	0.1866
本文方法	0.2023	0.0312	0.1971	0.1983	0.1681	0.2027

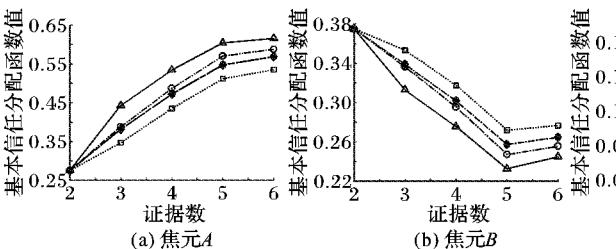
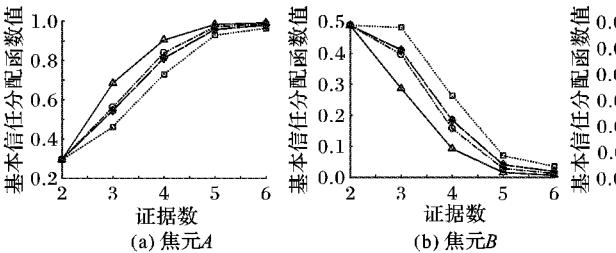
图5 各方法确定的MAE中焦元的基本信任分配函数
Fig. 5 BPA of focal-element in MAE determined by each method图6 各方法的融合结果
Fig. 6 Combination result of each method

表7中 $w(m_i)$ 表示证据 m_i 的权重,为了检测每种方法确定证据权重的合理性,通过式(13)计算出每个证据与其他证据的冲突强度之和如表8所示。

表8 每个证据对其他证据的冲突程度和(基于表4)

Tab. 8 Summation of conflict degrees of each evidence corresponding to Tab. 4

证据 与其它证据的冲突程度和	证据 与其它证据的冲突程度和
m_1 2.6653	m_4 2.3859
m_2 4.0240	m_5 2.3495
m_3 2.7337	m_6 2.3680

从表8可知,由于 m_2 与其他证据的冲突程度最大,为 m_2 分配的权重应该尽可能小。在表7中,证据 m_2 被文献[18, 22–23, 29]和本文方法确定的权重分别为0.1666、0.1058、0.1058、0.0782和0.0312。其中本文方法确定的证据权重是最小的。通过证据的权重和信息熵,可以确定代表着原始证据体和证据权重的归一化证据MAE。每个方法确定的MAE中各个焦元的基本信任分配函数如图5所示。

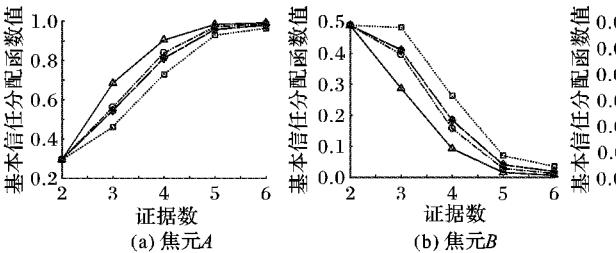
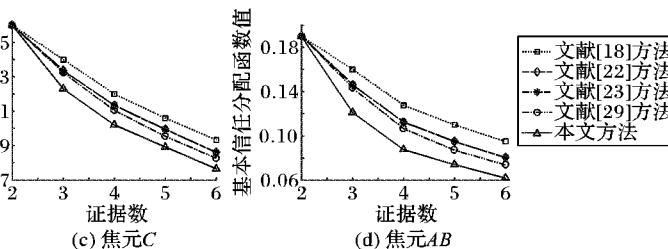
从图5可看出:本文方法确定的MAE中焦元A的基本信任分配函数值最大,同时其他几个焦元的基本信任分配函数最小。说明本文方法获得的MAE更加合理。

在获得MAE之后,融合结果中焦元的基本信任分配函数如图6所示。

由于焦元A在绝大多数证据的基本信任分配函数是最大的,融合结果中焦元A的基本信任分配函数越大越好;并且由于焦点元素B、C、AB,不是最后的融合结果,它们的基本信任分配函数应该越小越好。从图6(a)可以看出,无论证据数据是多少,本文方法中确定焦元A的基本信任分配函数始终是最大的;从图6(b)~(d)可以看出,由本文方法确定的其他几个焦元的基本信任分配函数始终是最小的。由此可得出:本文方法确定的基本信任分配函数更加合理,同时在证据数目较少时能更快地作出决策。

3.2 比较决策准确率

实验中使用鸢尾花植物数据集作为实验数据,数据集中每条记录包含四个特征:萼片长度、萼片宽度、花瓣长度、花瓣宽度和植物类型。通过文献[32]方法将一条记录转化为代表





四个特征的证据,然后通过证据的融合完成植物类型的决策。随机挑选一条在不同融合公式下表现不同的记录,该记录为:

```
type = Vi
features = [6.3, 2.5, 5.0, 1.9]
```

该记录代表着当前植物的真实类型为 V_i ,同时萼片长度、萼片宽度、花瓣长度、花瓣宽度分别为 6.3、2.5、5.0、1.9。将其四个属性转化为四条证据,如表 9 所示。

表 9 转化获得的证据
Tab. 9 Evidence converted by plant features

信任函数	萼片长度	萼片宽度	花瓣长度	花瓣宽度
$m(S_e)$	0.10256	0.08204	0.04892	0.06312
$m(V_e)$	0.17354	0.19404	0.15181	0.15168
$m(V_i)$	0.15490	0.13311	0.15523	0.22528
$m(S_e, V_e)$	0.13181	0.14454	0.06983	0.08379
$m(S_e, V_i)$	0.13181	0.10766	0.08554	0.09944
$m(V_e, V_i)$	0.17354	0.19404	0.35273	0.24513
$m(\theta)$	0.13181	0.14454	0.13591	0.13152

基于表 9 中的证据,使用参与比较的几种方法进行融合,其融合结果如表 10 所示。

表 10 融合结果中的基本信任分配函数
Tab. 10 BPAs in combination result

融合方法	融合结果
文献[18]方法	$m(S_e) = 0.0434835130314171$ $m(V_e) = 0.3593585019731172$ $m(V_i) = 0.3546828671221338$ $m(S_e, V_e) = 0.0244379475519$ $m(S_e, V_i) = 0.02382931434336$ $m(V_e, V_i) = 0.1535809559701$ $m(S_e, V_e, V_i) = 0.0026327020$
文献[23]方法	$m(S_e) = 0.0438271099903598$ $m(V_e) = 0.3596586373270804$ $m(V_i) = 0.35453621921017014$ $m(S_e, V_e) = 0.02459803779527$ $m(S_e, V_i) = 0.02392962979440$ $m(V_e, V_i) = 0.15265313951954$ $m(S_e, V_e, V_i) = 0.00263771659$
文献[29]方法	$m(S_e) = 0.0438270978672930$ $m(V_e) = 0.3596586884665241$ $m(V_i) = 0.3545360876452441$ $m(S_e, V_e) = 0.0245980291471$ $m(S_e, V_i) = 0.0239296261002$ $m(V_e, V_i) = 0.1526532541780$ $m(S_e, V_e, V_i) = 0.00263771625$
文献[22]方法	$m(S_e) = 0.0438270978672930$ $m(V_e) = 0.3596586884665241$ $m(V_i) = 0.3545360876452441$ $m(S_e, V_e) = 0.0245980291471$ $m(S_e, V_i) = 0.0239296261002$ $m(V_e, V_i) = 0.1526532541780$ $m(S_e, V_e, V_i) = 0.0026377162$
本文方法	$m(S_e) = 0.0332519125503896$ $m(V_e) = 0.3367366915522059$ $m(V_i) = 0.38127136576777315$ $m(S_e, V_e) = 0.01954271810636$ $m(S_e, V_i) = 0.02072756520453$ $m(V_e, V_i) = 0.17292136421963$ $m(S_e, V_e, V_i) = 0.00246237507$

从表 10 可以看出:文献[18,22-23,29]中融合算法作出了错误的决策,本文方法作出了正确决策。每种方法对数据集中每条记录的决策统计如表 11 所示。表 11 的中间三列中每个元素由三个数值构成:第一个数值代表由当前方法正确召回的个数,第二数据代表当前方法认为正确的召回个数,第三数据代表当前数据集中的当前类型的当前数据个数。最后一列的数据是统计每种方法对不同类型的 F-Score 的平均值,可以发现本文方法的 F-Score 的最高。

表 11 每种方法的决策表现
Tab. 11 Decision performance of each method

融合方法	对不同类型的决策结果			F-Score
	S_e	V_e	V_i	
文献[18]方法	25/26/25	21/32/25	17/17/25	0.84
文献[23]方法	25/27/25	23/30/25	18/18/25	0.88
文献[29]方法	25/27/25	23/30/25	18/18/25	0.88
文献[22]方法	25/27/25	23/30/25	18/18/25	0.88
本文方法	25/30/25	23/27/25	20/28/25	0.91

4 结语

本文针对相似度易产生相似度碰撞产生证据权重失衡并引发融合结果错误的问题,提出了一种新的抗相似度碰撞的证据融合方法。该方法首先将证据中焦元的序列特征量化为矩阵,然后将证据支持度和焦元序列特征转化为证据权重,最后使用 Dempster 证据融合算法将归一化证据融合 $n-1$ 次获得融合结果。通过两组实验检测本文方法的有效性:与文献[18,22-23,29]方法相比,第一组实验中,正确焦元的基本信任分配函数从 0.964、0.981、0.981、0.986 提高到 0.992;第二组实验中,对在线数据集鸢尾花的类型决策的 F-Score 从 0.84、0.88、0.88、0.88 提升到 0.91。实验结果表明本文提出的证据融合方法在分类决策方面取得了出色的表现,能够为证据融合提供更加准确、有效的融合方法,获得更加可靠的融合结果,为人工智能中分类决策、目标识别等问题提供新的解决方案。

参考文献(References)

- [1] DEMPSTER A P. Upper and lower probabilities induced by a multi-valued mapping [J]. Annals of Mathematical Statistics, 1967, 38(2): 325-339.
- [2] SHAFER G. A mathematical theory of evidence [J]. Technometrics, 1967, 20(1): 1-10.
- [3] YAGER R. On the Dempster-Shafer framework and new combination rules [J]. Information Sciences, 1987, 41(2): 93-137.
- [4] BEYNON M, CURRY B, MORGAN P. The Dempster-Shafer theory of evidence: an alternative approach to multicriteria decision modelling [J]. Omega, 2000, 28(1): 37-50.
- [5] LEUNG Y, JI N N, MA J H. An integrated information fusion approach based on the theory of evidence and group decision-making [J]. Information Fusion, 2013, 14(4): 410-422.
- [6] JIANG W, YANG Y, LUO Y. Determining basic probability assignment based on the improved similarity measures of generalized fuzzy numbers [J]. International Journal of Computers Communications & Control, 2015, 10(3): 333-347.
- [7] LIU H C, YOU J X, FAN X J. Failure mode and effects analysis



- using D numbers and grey relational projection method [J]. Expert Systems with Applications, 2014, 41(10): 4670 – 4679.
- [8] JIANG W, XIE C, WEI B. A modified method for risk evaluation in failure modes and effects analysis of aircraft turbine rotor blades [J]. Advances in Mechanical Engineering, 2016, 8(4): 1 – 16.
- [9] SU X, MAHADEVAN S, XU P. Dependence assessment in human reliability analysis using evidence theory and AHP [J]. Risk Analysis, 2015, 35(7): 1296 – 1316.
- [10] 赵秋月, 左万利, 田中生, 等. 一种基于改进 D-S 证据理论的信任关系强度评估方法研究 [J]. 计算机学报, 2014, 37(4): 873 – 883. (ZHAO Q Y, ZUO W L, TIAN Z S, et al. A method for assessment of trust relationship strength based on the improved D-S evidence theory[J]. Chinese Journal of Computers, 2014, 37 (4): 873 – 883.)
- [11] DENG Y, MAHADEVAN S, ZHOU D. Vulnerability assessment of physical protection systems: a bio-inspired approach [J]. International Journal of Unconventional Computing, 2015, 11(3/4): 227 – 243.
- [12] DENG Y, LIU Y, ZHOU D. An improved genetic algorithm with initial population strategy for symmetric TSP [J]. Mathematical Problems in Engineering, 2015, 2015(3): 1 – 6.
- [13] DU W, GAO Y, LIU C, et al. Limited-information particle swarm optimization [J]. Applied Mathematics and Computation, 2015, 268: 832 – 838.
- [14] 李弼程, 高洁, 高世海, 等. 信息融合技术及其应用 [M]. 北京: 国防工业出版社, 2010: 149 – 165. (LI B C, GAO J, GAO S H, et al. Information Fusion Technology and Application [M]. Beijing: Nation Defense Industry Press, 2010: 149 – 165.)
- [15] ZADEH L A. A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination [J]. AI Magazine, 1986, 7(2): 85 – 90.
- [16] DENG Y. Generalized evidence theory [J]. Applied Intelligence, 2015, 43(3): 530 – 543.
- [17] DUBOIS D, PRADE H. Representation and combination of uncertainty with belief functions and possibility measures [J]. Computational Intelligence, 1988, 4(3): 244 – 264.
- [18] MURPHY C K. Combining belief functions when evidence conflicts [J]. Decision Support Systems, 2000, 29(1): 1 – 9.
- [19] 韩德强, 杨艺, 韩崇昭. DS 证据理论研究进展及相关问题探讨 [J]. 控制与决策, 2014, 29(1): 1 – 11. (HAN D Q, YANG Y, HAN C Z. Advances in DS evidence theory and related discussions[J]. Control and Decision, 2014, 29(1): 1 – 11.)
- [20] CUZZOLIN F. A geometric approach to the theory of evidence [J]. IEEE Transactions on Systems, Man & Cybernetics, Part C, 2008, 38(4): 522 – 534.
- [21] JOUSSELME A L, GRENIER D, BOSSÉ É. A new distance between two bodies of evidence [J]. Information Fusion, 2001, 2 (2): 91 – 101.
- [22] WEN C, WANG Y, XU X. Fuzzy information fusion algorithm of fault diagnosis based on similarity measure of evidence [C]// Proceedings of the 2008 International Symposium on Neural Networks: Advances in Neural Networks. Berlin: Springer, 2008, 506 – 515.
- [23] WANG J, XIAO F, DENG X. Weighted evidence combination based on distance of evidence and entropy function [J]. International Journal of Distributed Sensor Networks, 2016, 12 (7): 3218784.
- [24] WANG J, QIAO K Y, ZHANG Z Z, et al. A new conflict management method in Dempster-Shafer theory [J]. International Journal of Distributed Sensor Networks, 2017, 13 (3): 1550147717696506.
- [25] KLIR G J, LEWIS H. Remarks on “measuring ambiguity in the evidence theory”[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, 2008, 38(4): 995 – 999.
- [26] 陈金广, 张芬, 马丽丽. 一种证据不确定度量方法及其应用 [J]. 计算机工程与应用, 2015, 51(23): 48 – 53. (CHEN J G, ZHANG F, MA L L. Uncertainty measure method of evidence and its application[J]. Computer Engineering & Applications, 2015, 51(23): 48 – 53.)
- [27] WANG X, SONG Y. Uncertainty measure in evidence theory with its applications [J]. Applied Intelligence, 2018, 48(7): 1672 – 1688.
- [28] KLIR G, WIERNAN M J. Uncertainty-based Information: Elements of Generalized Information Theory[M]. Berlin: Springer, 1999: 1 – 20.
- [29] 韩德强, 邓勇, 韩崇昭, 等. 基于证据距离与不确定度的证据组合方法 [J]. 红外与毫米波学报, 2011, 30(5): 396 – 400. (HAN D Q, DENG Y, HAN C Z, et al. Weighted evidence combination based on distance of evidence and uncertainty measure [J]. Journal of Infrared and Millimeter Waves, 2011, 30(5): 396 – 400.)
- [30] DENG Y, SHI W K, ZHU Z F. Combining belief functions based on distance of evidence [J]. Decision Support Systems, 2004, 38 (3): 489 – 493.
- [31] WANG J, HU Y, XIAO F, et al. A novel method to use fuzzy soft sets in decision making based on ambiguity measure and Dempster-Shafer theory of evidence: an application in medical diagnosis [J]. Artificial Intelligence in Medicine, 2016, 69: 1 – 15
- [32] 康兵义, 李娅, 邓勇, 等. 基于区间数的基本概率指派生成方法及应用 [J]. 电子学报, 2012, 40(6): 1092 – 1096. (KANG B Y, LI Y, DENG Y, et al. Determination of basic probability assignment based on interval numbers and its application [J]. Acta Electronica Sinica, 2012, 40(6): 1092 – 1096.)

This work is partially supported by the National Natural Science Foundation of China (61772174, 61370220), the Natural Science Foundation of Henan Province (162300410094).

WANG Jian, born in 1978, Ph. D., associate professor. Her research interests include multimedia social network, information security, trusted computing and usage control.

ZHANG Zhiyong, born in 1975, Ph. D., professor. His research interests include multimedia social networks, digital rights management, trusted computing and usage control.

QIAO Kuoyuan, born in 1991, M. S. candidate. His research interests include trust evaluation, data mining, information security, multi-source information fusion.