

基于可靠度和一致强度的冲突证据组合算法^①

万江文^② 陈斌* 冯仁剑 于宁

(北京航空航天大学仪器科学与光电工程学院 北京 100083)

(* 北京邮电大学计算机学院 北京 100876)

摘要 考虑到利用无线传感器网络对天然气管道进行泄漏监控时,不同传感器节点的诊断结果之间会发生冲突,致使直接利用 Dempster-Shafer (D-S)组合规则或其修正组合规则进行冲突证据组合的方法难以让汇聚节点做出正确决策,为此提出了一种基于可靠度和一致强度的冲突证据组合算法——CECARCI。该算法首先利用各监测节点的可靠度对证据集进行预处理,减小不可靠证据对组合结果的影响,然后通过对所有证据进行整体分析,引入了证据一致强度和基元支持度,调整证据组合时的次序,并以合理的权重来分配证据间的冲突。算例实验结果表明,相比于 D-S 组合法则、Yager 组合法则和 Murphy 组合法则,CECARCI 算法可获得收敛性更好的组合结果。

关键词 证据理论, 冲突证据, 一致强度, 可靠度

0 引言

无线传感器网络 (wireless sensor networks, WSN) 技术可用于实现对地下管网的远程、在线安全监控,具有较好的应用前景^[1]。由于受外界环境干扰以及传感器自身缺陷等因素的影响,单节点诊断结果存在识别不确定性,并且不同节点的诊断结果间可能发生严重冲突,为此需要对多个节点的诊断信息进行综合处理和协同分析,提高识别的准确性。

Dempster-Shafer (D-S) 证据理论能够较好地处理具有模糊、不确定性或低度冲突的信息合成问题,在模式识别、信息融合和决策分析等领域得到了广泛应用^[2,3]。为了避免将非零信度赋给空集,D-S 证据组合公式中采用了归一化因子,当证据间信息高度冲突时,归一化使得组合结果常常违背常理^[4]。针对其不足,许多学者在组合规则改进方面作了大量工作,认为冲突信息不能简单地丢弃,应该加以综合利用。Yager 等^[5]去掉了 D-S 证据组合公式中的归一化因子,将所有的冲突概率分配给了未知命题;Murphy 等^[6]提出一种不改变 D-S 证据组合规则的修改模型方法,将 n 条证据的基本信任分配值进行

平均后,利用 D-S 规则组合 $n - 1$ 次;Lefevre 等人^[7]提出“统一信度函数组合方法”,将冲突按照一定权重分配给各个子集,并采用梯度下降法自动学习获得权值。此外,一些学者还提出了基于冲突概率再分配^[8,9]和证据间距离^[10]等其它改进组合方法。尽管这些改进方法在某些情况下取得了较好的效果,但大多没有考虑实际应用背景以及证据组合次序对组合结果的影响等,难以保证最终结果向正确的命题方向可靠收敛。

结合天然气管道泄漏监控的实际应用背景,有必要给出一种新的冲突证据组合规则,获得收敛性更好的组合结果。本文通过对证据源的可靠性、一致性以及证据集对各基元元素的支持度分析,提出了一种基于可靠度和一致强度的冲突证据组合算法 (conflicting evidence combination algorithm based on reliability and coherence intensity, CECARCI),并实验验证了其有效性。

1 经典证据理论及存在的问题

1.1 D-S 证据理论基础^[7,11]

证据理论中,由一些完备的互不相容的元素组成的集合称为识别框架,表示为 $\Theta = \{\theta_1, \theta_2, \dots\}$,

① 863 计划(2006AA01Z222)和北京市教育委员会共建项目资助。

② 男,1963 年生,教授,博士生导师;研究方向:传感网络与仪器;联系人,E-mail:sensory@buaa.edu.cn
(收稿日期:2008-06-18)

$\theta_n\}$ 。由 Θ 的所有子集构成的一个有限集合称为 Θ 的幂集合,记作 $P^\Theta = \{\emptyset, \{\theta_1\}, \{\theta_2\}, \dots, \{\theta_1, \theta_2\}, \{\theta_1, \theta_3\}, \dots, \Theta\}$,其中 \emptyset 表示空集。

假设 $\forall A \subset P^\Theta$, 若映射 $m: P^\Theta \rightarrow [0,1]$ 满足 $m(\emptyset) = 0, \sum_{A \subset P^\Theta} m(A) = 1$ (1)

则称 m 为基本概率指派函数(basic probability assignment, BPA)。

$m(A)$ 为命题 A 的基本概率指派值,若 $m(A) > 0$,则称 A 为证据的焦元。

与 BPA 相关的信任度函数(Bel)和似真度函数(Pls)分别定义为

$$\begin{cases} Bel: P(\Theta) \rightarrow [0,1], Bel(A) = \sum_{B \subseteq A} m(B) \\ Pls: P(\Theta) \rightarrow [0,1], Pls(A) = \sum_{B \cap A \neq \emptyset} m(B) \end{cases} \quad (2)$$

此外,D-S 证据理论提供了一个非常有用的证据组合公式。设 m_1 和 m_2 分别是同一识别框架 Θ 上的基本概率指派函数,焦元分别为 $A_i (i = 1, 2, \dots)$ 和 $B_j (j = 1, 2, \dots)$, 则两条证据的组合公式为

$$\begin{cases} m_{1 \oplus 2}(A) = 0, A = \emptyset \\ m_{1 \oplus 2}(A) = \frac{\sum_{A_i \cap B_j = A} m_1(A_i)m_2(B_j)}{1 - K} \end{cases} \quad (3)$$

其中 $K = \sum_{A_i \cap B_j = \emptyset} m_1(A_i)m_2(B_j)$ 反映了证据间冲突的程度, \oplus 表示直和运算。

N 条证据可通过式

$$m(A) = m_1 \oplus m_2 \oplus \dots \oplus m_i \dots \oplus m_N(A) \quad (4)$$

进行组合,得到一个多条证据联合作用的基本概率指派函数,为决策分析等提供更加准确、有效的信息。

1.2 存在的问题

当证据间完全冲突时($K = 1$),无法使用组合公式(3)。此外,在证据高度冲突时(即 K 的取值接近于 1)也可能得到明显不合理的结论。例如,设 $\Theta = \{A_1, A_2, A_3\}$,两条证据的 BPA 分别为

$$m_1(A_1) = 0.9, m_1(A_2) = 0.0, m_1(A_3) = 0.1$$

$$m_2(A_1) = 0.0, m_2(A_2) = 0.9, m_2(A_3) = 0.1$$

则由组合公式(3)可得, $m(A_1) = m(A_2) = 0, m(A_3) = 1$ 。可见,对命题 A_3 支持度都极小的两条证据,组合后却得到了几乎肯定的支持,显然不符合正常的逻辑。

当证据集中的一条证据与其它多条证据完全不相一致时,即存在“异常证据”,会出现一票否决的问

题。另外,D-S 证据理论假设各条证据是 100% 等可靠的,这与实际应用情况往往并不相符。

2 CACERCI 算法

2.1 证据的模型描述

图 1 给出了基于 WSN 的管道安全监控网络结构,该结构采用分簇的设计方式,以降低协议设计和网络管理的复杂度,同时满足新加管道对网络可扩展性的需求。簇由若干的普通检测传感器节点和簇头节点组成,其中簇头由处理能力较强的 sink 节点担当。普通节点沿管道安装,节点间可采用管内无线通信方式;sink 节点可安装于地面上,负责融合处理簇内节点的检测数据,并将最终的诊断结果发送至控制管理中心。

假设管道发生泄漏后,簇内的 N 个节点均可提取到泄漏信号并给出诊断结果。按照 D-S 证据理论,这些节点关于管道状态的诊断信息可构造出 N 条独立的证据,表示为 $\{m_i, i = 1, 2, \dots, N\}$ 。假设识别框架的幂集合 $P^\Theta = \{A_j, j = 1, 2, \dots, M\}$, $m_i(A_j)$ 表示节点 i 对管道的第 j 种状态的基本概率赋值,可通过建立神经网络模型来得到,各检测节点至泄漏点的距离分别为 $d_i (i = 1, 2, \dots, N)$ 。

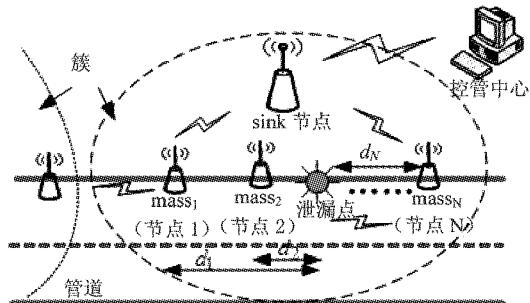


图 1 WSN 监控网络结构

2.2 算法设计

围绕 WSN 在管道安全监控中的实际应用需求,新的组合规则着重考虑证据不可靠性、组合的次序和冲突概率分配机制等因素对最终组合结果的影响,并给出相应的解决方法。

2.2.1 证据集预处理

根据节点上的压力传感器和声传感器所采集到的负压波和声发射信号,各节点可以判断出该节点至管道泄漏点的距离。其中模态声发射泄漏检测方法具有较高的定位精度,泄漏诊断信息很可靠。由

于实际管道泄漏诊断过程中检测节点距离泄漏点越近,提取到的泄漏特征信号越强,诊断结果的可靠性越高,反之,距离越远,泄漏信号中的干扰成分越多,诊断结果越不可靠,因而证据可靠度与距离有关。

由于该应用中汇聚节点是收齐所有簇内节点(证据源)的数据后进行融合处理的,所以,可依据各节点发送的距离值来计算各条证据的可靠度。

假设距离泄漏点最近的传感器节点的可靠度为1,距离大小为 d ,则证据的可靠度 Cre 可通过式

$$\begin{cases} d = \min(d_i) & i = 1, 2, \dots, N \\ Cre_i = \sqrt{Td/d_i} & 0 < Cre_i \leq 1 \end{cases} \quad (5)$$

计算得到,其中, T 为可靠度的影响因子,根据实际诊断情况,经仿真调试后确定为5。为了满足公式(1)中关于各命题基本概率赋值之和为1的要求,经过可靠度预处理后,各证据体的mass函数可表示为

$$\begin{cases} m_i^*(A_j) = Cre_i m_i(A_j), A_j \neq \emptyset \\ m_i^*(\Theta) = 1 - \sum_{A_j \subset \Theta} m_i^*(A_j) \end{cases} \quad (6)$$

由公式(6)可以看出,通过预处理,减小了不可靠证据对正确命题的支持度,转而将这部分概率赋给了未知命题 Θ ,从而减弱了其对组合结果的影响。

2.2.2 证据组合次序的确定

考虑到噪声干扰等因素,传感器的诊断结果可能与事实相违背,致使汇聚节点收集的证据中存在“异常证据”,融合算法难以识别而简单予以丢弃。同时,汇聚节点对收集的各节点信息不分次序进行处理时,由于异常证据随机参与组合,必然使组合结果大幅偏离正确的命题方向。为防止组合结果出现大幅“摆动”,引入证据的总一致强度来调整组合的次序。首先集中组合处理一致强度较低的异常证据,然后组合一致强度较高的其它证据,加快组合结果向正确命题方向的一致收敛速度。

证据集预处理以后,同一识别框架下的两条证据间的完全一致量和冲突量分别为

$$E_{1,2} = \sum_{A_i \cap A_j = A} m_1^*(A_i) m_2^*(A_j) \quad (7)$$

$$K_{1,2} = \sum_{A_i \cap A_j = \emptyset} m_1^*(A_i) m_2^*(A_j) \quad (8)$$

证据间的一致强度 Y 定义为

$$Y_{1,2} = \frac{E_{1,2}}{E_{1,2} + K_{1,2}} \quad (9)$$

由(9)式可以看出, Y 值越大,证据间的冲突越小。 $Y=0$,证据间完全冲突,D-S组合规则失效; $Y \in (0, 0.5]$ 时,证据间高度冲突,直接利用D-S组合方法可能得出不合理的结论; $Y \in (0.5, 1)$ 时,证据

间冲突较小,D-S规则仍然有效;当 $Y=1$ 时,说明证据间完全一致,未有冲突。

推广至整个证据集,可得证据的一致强度矩阵

$$Y = \begin{bmatrix} 1 & Y_{1,2} & \cdots & Y_{1,N} \\ Y_{2,1} & \ddots & & Y_{2,N} \\ \vdots & & \ddots & \vdots \\ Y_{N,1} & Y_{N,2} & \cdots & 1 \end{bmatrix}_{N \times N} \quad (10)$$

对(10)式中矩阵的行向量分别作加性运算,归一化处理后,可得证据 i 总的一致强度值

$$Ym_i = \frac{\sum_{l=1}^N Y_{i,l} - 1}{\sum_{i=1}^N \sum_{l=1}^N Y_{i,l} - N} \quad (11)$$

利用一致强度确定证据组合的次序时,一致强度值越大,证据参与组合的次序越靠后,可得新的组合序列 $\{Z_i, i = 1, 2, \dots, N\}$ 。

2.2.3 基元支持度计算

证据两两组合时,证据间冲突中还包含有一定的有用信息,既不能简单的丢弃,也不能直接的全部分配给未知命题,否则将增加识别的不确定性,需要合理的加以分配。一致强度值可将冲突划分为有用信息(大小为 $Y_{1,2} \cdot K_{1,2}$)及无用信息(大小为 $(1 - Y_{1,2}) \cdot K_{1,2}$)两部分。考虑到基元元素所占的支持度比重对组合结果的影响,本文利用证据集对各基元的支持度来合理分配有用信息,无用信息则全部分配给未知命题。本文采用加性策略来计算证据集对基元元素的支持度。假设幂集合中所有基元构成的集合 $B = \{B_h, h = 1, 2, \dots, M^*\}$,且 $M^* < M$,则各基元的支持度 Sur 为

$$Sur(B_h) = \frac{\sum_{i=1}^N m_i^*(B_h)}{\sum_{h=1}^{M^*} \sum_{i=1}^N m_i^*(B_h)} \quad (12)$$

证据集对某基元的支持度越大,其分配的有用信息越多,从而使组合结果向多数证据支持的命题方向收敛。

2.2.4 证据组合公式

基于上述可考虑,新的证据组合规则定义如下:

$$m(A) = \begin{cases} \sum_{A_i \cap A_j = A} m_1^*(A_i) \cdot m_2^*(A_j) \\ \quad + Sur(A) \cdot Y_{1,2} \cdot K_{1,2} & A \neq \Theta, \emptyset \\ \sum_{A_i \cap A_j = A} m_1^*(A_i) \cdot m_2^*(A_j) \\ \quad + (1 - Y_{1,2}) \cdot K_{1,2} & A = \Theta \end{cases} \quad (13)$$

2.2.5 诊断决策

根据组合得到的最终基本置信度指派值,可通

过以下规则进行决策。设 $\exists A_1, A_2 \subset \Theta$, 且 $m(A_1) = \max\{m(A_k), A_k \subset \Theta\}$, $m(A_2) = \max\{m(A_k), A_k \subset \Theta, \text{且 } A_k \neq A_1\}$, 若满足

$$\begin{cases} m(A_1) - m(A_2) > \epsilon_1 \\ m(\Theta) < \epsilon_2 \\ m(A_1) > m(\Theta) \end{cases} \quad (14)$$

则诊断结果为 A_1 。其中, ϵ_1, ϵ_2 为预先设定的阈值, 根据实际诊断情况, 经大量调试后将 ϵ_1 确定为 0.3, ϵ_2 为 0.1。

3 算例分析

下面以管道泄漏诊断的具体应用实例来验证算法的有效性。假设簇内共有 6 个普通的传感器节点, 各节点至泄漏部位的距离依次为 50m, 160m, 120m, 80m, 200m, 100m。识别框架 $\Theta = \{A_1, A_2, A_3\}$, 其中, A_1 表示“大泄漏”; A_2 表示“小泄漏”; A_3 表示“正常”; 集合 Θ 表示可能泄漏也可能正常。所有证据关于管道状态的基本概率指派值如表 1 所示。

表 1 各条证据的 BPA 值

BPA	A_1	A_2	A_3	Θ
m_1	0.90	0.05	0.05	0.00
m_2	0.00	0.85	0.15	0.00
m_3	0.65	0.20	0.10	0.05
m_4	0.50	0.15	0.25	0.10
m_5	0.35	0.15	0.45	0.05
m_6	0.75	0.10	0.05	0.10

由公式(5)可得证据的可靠度值分别为

$$Cre_1 = 1.0000, Cre_2 = 0.7924, Cre_3 = 0.8394$$

$$Cre_4 = 0.9103, Cre_5 = 0.7579, Cre_6 = 0.8706$$

代入公式(6) – (12), 可得证据组合序列 $\{Z_i\} = \{m_2^*, m_5^*, m_4^*, m_1^*, m_3^*, m_6^*\}$ 和各基元的支持度: $Sur(A_1) = 0.5738$, $Sur(A_2) = 0.2501$, $Sur(A_3) = 0.1761$ 。由新的证据组合公式可得对命题 A_1 的最终支持度为 0.7843, 满足决策规则(14), 即发生了大泄漏。表 2 给出了不同证据组合规则下的诊断结果。对比可以看出, 当证据间高度冲突时, 采用新组合算法可得到收敛性相对较好的正确诊断结果。主要因为:

表 2 不同组合规则下的组合结果

组合法则	组合结果	m_1	m_2	m_1	m_2	m_3	m_1	m_2	m_3	m_1	m_2	m_3
		m_4	m_4	m_4	m_4	m_5	m_4	m_5	m_6	m_4	m_5	m_6
D-S	$m(A_1)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$m(A_2)$	0.8500	0.9043	0.8709	0.7296	0.7825						
	$m(A_3)$	0.1500	0.0957	0.1291	0.2704	0.2175						
	$m(\Theta)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Yager	$m(A_1)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$m(A_2)$	0.0425	0.0085	0.0013	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$m(A_3)$	0.0075	0.0008	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$m(\Theta)$	0.9500	0.9908	0.9985	0.9997	1.0000						
Murphy	$m(A_1)$	0.4880	0.6394	0.6649	0.6292	0.7038						
	$m(A_2)$	0.4880	0.3300	0.2674	0.2360	0.1877						
	$m(A_3)$	0.0241	0.0300	0.0645	0.1311	0.1032						
	$m(\Theta)$	0.0000	0.0003	0.0014	0.0016	0.0025						
文献[12]	$m(A_1)$	0.0214	0.6143	0.5375	0.3918	0.6745						
	$m(A_2)$	0.0713	0.2034	0.1037	0.0846	0.0656						
	$m(A_3)$	0.0049	0.0930	0.0931	0.2174	0.0741						
	$m(\Theta)$	0.9025	0.0893	0.2657	0.3062	0.1857						
CECARCI	$m(A_1)$	0.1085	0.3133	0.6737	0.7028	0.7843						
	$m(A_2)$	0.3121	0.1796	0.0669	0.0971	0.0772						
	$m(A_3)$	0.1610	0.1775	0.0545	0.0595	0.0466						
	$m(\Theta)$	0.4185	0.3295	0.2049	0.1406	0.0919						

(1) D-S 组合规则由于其完全舍弃了证据间的冲突, 其诊断结果为 A_2 , 这与大多数证据都支持命题 A_1 是相违背的。新法则丢弃了组合公式中的归一化因子, 避免了证据间高度冲突时组合结果有

背常理现象。

(2) 尽管新算法和 Yager 组合规则都承认了冲突对融合结果造成的不确定性, 但 Yager 方法仅是

简单的将冲突全部分配给未知命题 Θ , 致使组合结果的模糊性较大, 没有实际意义。新算法则不同, 认为冲突中包含一定的有用信息, 并采用加权分配方法将这部分信息赋给各基元元素, 从而确保了组合结果向多数证据支持的命题的方向收敛。

(3) Murphy 提出的修改模型方法, 将多条证据的基本概率指派函数(BPA)值进行了简单平均, 尽管同样可得到正确的结果, 但由于没有考虑实际应用中存在的证据可靠性对组合结果的影响, 从而降低了证据集关于正确命题的聚焦度。新算法利用证据可靠度对证据集加以预处理, 将不可靠度证据(例如, 证据2和5)对正确命题 A_1 的一部分支持度转而赋给了未知命题, 提高了组合结果向正确命题方向的收敛度。

尽管合成前三条证据后, 文献[12]关于命题 A_1 的信度达到了 0.6143, 但由于未考虑证据组合次序以及证据可靠性等因素的影响, 最终得到的融合结果也并不理想。新算法则考虑了证据组合次序对组合结果的影响, 先集中的处理冲突较大的异常证据, 然后组合一致强度较高的其它证据, 避免了无规则的随机组合所可能导致的收敛方向大幅摆动问题, 提高了证据集对正确命题的聚焦程度。

4 结论

结合天然气管道泄漏诊断的实际应用背景, 提出了一种基于可靠度和一致强度的冲突证据组合算法。该方法减弱了异常的不可靠证据对组合结果的影响, 在同时考虑证据组合次序和基元支持程度的基础上对证据冲突进行了合理的分配, 提高了组合结果向正确命题方向的聚焦度。实验结果表明,

CECARCI 算法可组合证据集中含有高度冲突性的证据, 而且组合结果具有较好的可信性和合理性。

参考文献

- [1] Stoianov I, Nachman L, Madden S, et al. PIPENET: a wireless sensor network for pipeline monitoring. In: Proceedings of the 6th International Symposium on Information Processing in Sensor Networks. Massachusetts, USA: IEEE Press, 2007. 264-273
- [2] Denoeux T, Masson M. EVCLUS: evidential clustering of proximity data. *IEEE Systems, Man and Cybernetics B*, 2004, 34(1):95-109
- [3] Yang J B, Wang Y M, Xu D L, et al. The evidential reasoning approach for MCDA under both probabilistic and fuzzy uncertainties. *European Journal of Operational Research*, 2005, 171(1):309-343
- [4] Laha A, Pal N R, Das J. Land cover classification using fuzzy rules and aggregation of contextual information through evidence theory. *IEEE Transaction on Geoscience and Remote Sensing*, 2006, 44(6):1633-1641
- [5] Yager R R. On the aggregation of prioritized belief structure. *IEEE Transactions on System, Man and Cybernetics*, 1996, 26(6): 708-719
- [6] Murphy C. Combining belief functions when evidence conflicts. *Decision Support Systems*, 2000, 29(1): 1-9
- [7] Lefevre E, Colot O, Vannoorenbergh P. Belief functions combination and conflict management. *Information Fusion*, 2002, 3(2):149-162
- [8] Pu S J, Yang S Y, Hu W W. A combination rule of evidence theory based on reallocation of the basic probability assigned to conflict. In: Proceedings of the 6th World Conference on Control and Automation. Dalian, China: IEEE Press, 2006. 4264-4268
- [9] Dezert J, Smarandache F. Information fusion based on new proportional conflict redistribution rules. In: Proceedings of the 7th International Conference on Information Fusion. Philadelphia, USA: IEEE Press, 2005. 25-29
- [10] Chen L Z, Shi W K, Deng Y, et al. A new fusion approach based on distance of evidences. *Journal of Zhejiang University SCIENCE*, 2005, 6A (5): 476 -482
- [11] Shafer G. A Mathematical Theory of Evidence. Princeton: Princeton University Press, 1976
- [12] 王肖霞, 杨风暴. 基于冲突强度和非正则化的证据合成方法研究. *计算机工程与应用*, 2006, 30:78-80

A conflicting evidence combination algorithm based on reliability and coherence intensity

Wan Jiangwen, Chen Bin*, Feng Renjian, Yu Ning

(School of Instrument Science & Opto-Electronics Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083)

(* School of Computer Science and Technology, Beijing University of Posts and Telecommunications, Beijing 100876)

Abstract

Considering that when wireless sensor networks are applied to security monitoring for gas pipelines, the diagnosis results from different sensor nodes may seriously conflict, so it is difficult for the sink node to make correct decisions using the Dempster-Shafer (D-S) Combination rule or its modified combination rules, this paper proposes a new conflicting evidence combination algorithm based on reliability and coherence intensity (CECARCI). Firstly, the algorithm preprocesses the evidence set according to nodes' reliability in order to decrease the effect of unreliable evidences on the fusion result, and then analyzes all the evidences, and obtains the sequence of combination according to the coherence intensity of evidence. Additionally, the coherence intensity and the support degree of base element are used to manage the conflict. The numerical example shows that the CECARCI can obtain more reasonable results with good convergence compared with D-S, Yager or Murphy combination rules.

Key words: theory of evidence, conflicting evidence, coherence intensity, reliability