

文章编号: 1001 - 893X(2011)01 - 0027 - 04

基于 K - L 距离的改进 D - S 证据合成方法*

魏永超

(中国民用航空飞行学院 飞行技术与飞行安全科研基地, 四川 广汉 618307)

摘要:提出了一种证据距离为基础的冲突证据合成方法。通过改进的 K - L 距离得到证据间间距, 进一步通过相关算法转化为证据权值, 权值反映了证据的重要程度, 以权值为系数对原始的基本可信度分配函数进行重新分配, 再通过改进的合成公式, 得到证据合成结果。最后通过具体的算例比较, 说明新方法的合成结果更确定, 更符合实际情况。

关键词:数据融合; D - S 理论; 证据合成; K - L 距离; 权值

中图分类号: TP274 **文献标识码:** A **doi:** 10.3969/j.issn.1001-893x.2011.01.006

An Improved D - S Evidence Combination Method Based on K - L Distance

WEI Yong-chao

(Academy of Flight Technology and Safety, Civil Aviation Flight University
of China, Guanghan 618307, China)

Abstract: A novel conflict evidence combination method based on evidence distance is proposed. The distance between evidences is obtained with improved K - L distance equation, and is further transformed to weight which reflects the importance of evidence. According to weights, the basic probabilities are reallocated, and then are used to get the final combination result by the improved combination method. The final comparison example shows that the combination results of new method are determined, and are more accordance with the actual conditions.

Key words: data fusion; D - S theory; evidence combination; K - L distance; weight

1 引言

D - S 理论是 Dempster 和 Shafer 共同创立的一种不确定性推理理论, 由于其对不确定信息的表达和合成提供了自然而强有力的方法, 因此在信息融合中获得了广泛的应用^[1-3]。Dempster 证据合成公式要求待合成的证据是相互独立且证据之间不能有较强的冲突, 否则合成结果将有悖常理。冲突证据合成方法一直是研究的一个热点。如何在证据高度

冲突下实现多源信息的有效融合是一个迫切需要解决的问题, 为此, 很多学者提出了多种解决方法^[4]。

目前, 在证据高度冲突下实现多源信息的有效融合的方法总体可分为两大类^[5,6]。第一类方法认为: 证据高度冲突下使用 Dempster 组合规则产生不合理结论, 是由该规则的归一化步骤所产生的。新的组合规则主要是解决如何将冲突重新分配的问题。第二类解决方法的思路是: Dempster 组合规则本身没有错, 在证据高度冲突时应该首先对冲突证据进行预处理, 然后再使用 Dempster 规则。

* 收稿日期: 2010 - 11 - 09; 修回日期: 2010 - 12 - 13

基金项目: 民航局科技项目资助(MHRD200940 & MHRD201024); 中国民航飞行学院博士启动基金资助项目(J2009 - 45)

Foundation Item: Civil Aviation Administration of China Science and Technology Project(No. MHRD200940 & MHRD201024); The Research Foundation for the Doctoral Program of Civil Aviation Flight University of China(No. J2009 - 45)

无论何种方法,在融合之前确定证据间的冲突程度是必须的,冲突系数 k 是表示证据之间冲突程度的一个常见方法。然而,这种方法在表示冲突方面存在一定的问题。本文在分析冲突系数 k 不足的基础之上,提出了一种新的判断冲突的方法,即通过改进的 K-L 距离来判断证据间的冲突程度,并在判断冲突的基础之上,进一步处理得到证据的权值,根据权值对证据进行预处理,从而解决了冲突证据融合的问题。新方法可以快速得到高冲突证据下的融合结果,对于具体的工程应用具有一定的理论和实用价值。

2 Dempster-Shafer (D-S) 理论及存在问题

在证据理论中^[7],一个样本空间称为一个辨识框架,用 Θ 表示。 Θ 由一系列两两相斥对象构成,且包含当前要识别的全体对象,即:

$$\Theta = \{\theta_1, \theta_2, \theta_3, \dots, \theta_n\} \quad (1)$$

式中,对象 θ_i 称为 Θ 的一个单子,只含一个单子的集合称为单子集合。对于 Θ 的每个子集,可以指派一个概率,称为基本概率分配,定义如下:

定义 1: 令 Θ 为一论域集合, 2^Θ 为 Θ 的所有子集构成的集合,称 $m: 2^\Theta \rightarrow [0, 1]$ 为基本概率分配函数,它满足如下公理:

$$\sum_{A \in P(\Theta)} m(A) = 1, m(\Phi) = 0 \quad (2)$$

$m(A)$ 称为事件 A 的基本概率赋值,表示对命题 A 的支持程度。Dempster 组合规则为

$$\begin{cases} m(A) = \frac{1}{1-k} \sum_{\bigcap_{j=1}^n A_j = A} \prod_{1 \leq i \leq n} m_i(A_i), A \neq \Phi \\ m(\Phi) = 0 \end{cases} \quad (3)$$

式中, k 为冲突因子,它反映了证据之间冲突的程度。

$$k = \sum_{\bigcap_{j=1}^n A_j = \Phi} \prod_{1 \leq i \leq n} m_i(A_i) \quad (4)$$

在 Dempster 组合规则中, k 是一个衡量各个证据之间冲突程度的系数。从公式(3)中可以看出,只有两个相交为空集的焦元的基本概率指派才能相乘并累加计算 k 值。如果 $k=1$, 就不能使用 Dempster 组合规则进行信息融合;而当 $k \rightarrow 1$ 时,即对高度冲突的证据进行正则化处理将会导致与直觉相悖的结果。

3 新的证据融合方法

D-S 合成公式要求所有参与合成的证据具有相同的重要程度,而融合系统中,由于各证据的不完全可靠性,不一定具有相同的重要程度,所以在实际

应用中不能将各证据直接利用 D-S 合成公式进行合成。要直接利用该公式,首先需要考虑参与合成的各证据的重要程度,即权值的大小。证据之间发生的严重或者完全冲突往往是由于某个或少数证据的不相容引起的,根据“少数服从多数”的思想,在证据合成过程中,对引起冲突的证据使其作用较小,所以其权值系数就小;反之,其权值系数就越大,对融合结论的影响也越大。

3.1 改进的 Kullback-Leibler (K-L) 距离

K-L 距离又叫相对信息熵,它度量两个概率分布 P 和 Q 之间的距离。假设 P 和 Q 是两个概率分布函数, P 相对于 Q 的 Kullback-Leibler 信息距离即相对信息熵为^[8]

$$D(P \parallel Q) = \sum_{i=1}^n p_i \lg(p_i/q_i) \quad (5)$$

K-L 距离可以测量信息距离,且具有不对称性质。然而,直接用 K-L 到证据距离测量有一定的问题, q_i 的值不能为零,且当 p_i 为零时,测量得到距离为零,也是无效的。因此本文对 K-L 距离进行了改进,使其可以用于证据冲突距离的判断。改进的公式如下:

$$d(m_1, m_2) = \sum_{i=1}^n (m_1(A_i) + \alpha) \lg \frac{m_1(A_i) + \alpha}{m_2(A_i) + \alpha} \quad (6)$$

式中, m_1, m_2 为两个证据的基本可信度分配函数, α 为一趋于零很小的定值。

3.2 权值系数的确定

通过改进的 K-L 判断证据重要程度,下面详细介绍如何根据 K-L 距离得到证据权值。

设融合系统的辨识框架 Θ 包含 N 个完备的互不相容的假设命题,令其幂集 $2^\Theta = \{A_1, A_2, A_3, \dots, A_2^N\}$ 。 n 个证据 $E_1, E_2, E_3, \dots, E_n$ 的基本可信度分配函数分别为 $m_1, m_2, m_3, \dots, m_n$ 。

定义 1: 证据 E_1, E_2 之间的距离定义为

$$D(m_1, m_2) = d(m_1, m_2) + d(m_2, m_1) \quad (7)$$

定义 2: 融合系统中证据 E_i 的基本可信度函数与系统中其它证据的基本可信度分配函数冲突程度之和定义为

$$\delta(m_i) = \sum_{j=1, j \neq i}^n D(m_i, m_j) \quad (8)$$

定义 3: v_i 为证据 E_i 质疑度,值越大,质疑度越高,从而证据的可信度就低。 v_i 的定义如下:

$$v_i = \delta(m_i) / \sum_{i=1}^n \delta(m_i) \quad (9)$$

定义 4:若证据 E_c 满足式(10),则称 E_c 为融合系统中的中心证据。

$$\delta(m_c) = \min_{1 \leq i \leq n} (\delta(m_i)) \quad (10)$$

定义 5:证据的权值 w_i 的定义为

$$w_i = v_c/v_i \quad (11)$$

3.3 权值冲突证据合成方法

所有的权值组成了证据源的权值向量 $W = (w_1, w_2, w_3, \dots, w_n)$ 满足 $w_i \in [0, 1]$ 。权值系数反映了证据源提供的证据在合成过程中的重要程度以及它们对合成结果的影响程度,权值系数的确定方法上面已经给出了详细介绍。对加权之后的基本概率赋值继承文献[9]中合成方法,具体规则如下:

步骤 1:根据上面的权值计算算法,得到证据的权重向量 $W = (w_1, w_2, w_3, \dots, w_n)$;

步骤 2:根据权重向量中权值,按式(12)对证据概率重新分配:

$$m'_k(A_i) = \begin{cases} w_k \times m_k(A_i), & A \neq \Theta \\ 1 - \sum_{A_i \neq \phi} w_k \times m_k(A_i), & A = \Theta \end{cases} \quad (12)$$

步骤 3:根据得到的新概率分配值,分别计算冲突值 k 和证据对命题的平均支持程度 $q(A)$;

$$k = \sum_{i,j:A_i \cap A_j = \Phi} m'_1(A_i) m'_2(A_j) \quad (13)$$

$$q(A) = \frac{1}{n} \sum_{i=1}^n m'_1(A_i) \quad (14)$$

步骤 4:将重新调整的概率分配代入到下面融合公式,计算融合结果。

$$\begin{cases} \tilde{m}(A) = \sum_{i,j:A_i \cap A_j = A} m'_1(A_i) m'_2(A_j) + k \cdot q(A), A \neq \Phi, \Theta \\ \tilde{m}(\Phi) = 0 \\ \tilde{m}(\Theta) = 1 - \sum_{A \in \Theta} \tilde{m}(A) \end{cases} \quad (15)$$

4 实例分析

通过具体的证据源合成实验,对改进的 D-S 证据合成公式进行验证,并和传统的 D-S 合成公式、Yager^[10]推广的合成公式、文献[11]的公式进行合成结果的比对,实验中 α 的值为 0.000 1。以文献[11]给出的 4 个证据为例,设 $\Theta = \{A, B, C\}$:

- $E_1: m_1(A) = 0.98, m_1(B) = 0.01, m_1(C) = 0.01;$
- $E_2: m_2(A) = 0, m_2(B) = 0.01, m_2(C) = 0.99;$
- $E_3: m_3(A) = 0.9, m_3(B) = 0, m_3(C) = 0.1;$
- $E_4: m_4(A) = 0.9, m_4(B) = 0, m_4(C) = 0.1。$

从证据可以看出,证据 2 和其它证据冲突。表 1 是证据分别为 3 个及 4 个时得到的权值,从表 1 中证据距离可以看出,证据 2 和证据 1、证据 3 高度冲突,而证据 1 和证据 3 之间的冲突很小,符合推理。因此在证据合成的过程中,证据 2 的权值相对证据 1 和 3 就要小。从表 1 中还可以看出,证据 2 的权值 0.200 6 为最小,而证据 1 和证据 3 的权值分配要相对大一些。另外,随着证据的增多,冲突证据得到的权值更小,从而使合成结果更加合理。

表 1 权值
Table 1 Weight results

证据	三证据		四证据	
	证据距离	权值	证据距离	权值
1	(13.5, 0.259 4)	0.766	(13.5, 0.259 4, 0.259 4)	0.751 8
2	(13.5, 10.28)	0.443 2	(13.5, 10.28, 10.28)	0.309 4
3	(0.259 4, 10.28)	1	(0.259 4, 10.28, 0)	1
4			(0.259 4, 10.28, 0)	1

表 2 实例分析结果
Table 2 Example analysis results

合成方法	m_1, m_2	m_1, m_2, m_3	m_1, m_2, m_3, m_4
D-S	$k = 0.99, m(A) = 0, m(B) = 0.01, m(C) = 0.99, m(\Theta) = 0$	$k = 0.999\ 01, m(A) = m(B) = 0, m(C) = 1, m(\Theta) = 0$	$k = 0.999\ 901, m(A) = m(B) = 0, m(C) = 1, m(\Theta) = 0$
Yager ^[10]	$k = 0.99, m(A) = 0, m(B) = 0.000\ 1, m(C) = 0.009\ 9, m(\Theta) = 0.99$	$k = 0.999\ 01, m(A) = m(B) = 0, m(C) = 0.000\ 99, m(\Theta) = 0.999\ 01$	$k = 0.999\ 901, m(A) = m(B) = 0, m(C) = 0.000\ 099, m(\Theta) = 0.999\ 901$
文献[11]	$\epsilon = 0.371\ 6, m(A) = 0.18, m(B) = 0.004, m(C) = 0.194, m(\Theta) = 0.622$	$\epsilon = 0.512, m(A) = 0.321, m(B) = 0.003, m(C) = 0.188, m(\Theta) = 0.488$	$\epsilon = 0.604, m(A) = 0.42, m(B) = 0.003, m(C) = 0.181, m(\Theta) = 0.396$
本文	$k = 0.99, m(A) = 0.485\ 1, m(B) = 0.01, m(C) = 0.504\ 9, m(\Theta) = 0$	$k = 0.482\ 51, m(A) = 0.758\ 9, m(B) = 0.001\ 9, m(C) = 0.111\ 9, m(\Theta) = 0.127\ 2$	$k = 0.446\ 51, m(A) = 0.834\ 1, m(B) = 0.001\ 2, m(C) = 0.059\ 9, m(\Theta) = 0.104\ 8$

表 2 为算法分析与其它算法的结果对比,从表中可以看出,算法相对其它算法有了很大的改善,不但极大地降低冲突证据的冲突程度,还能够很好地得到结果。从合成结果可以看出,D-S、Yager、文献[11]与本文算法的优越性逐渐提高,本文算法最优。算法随着证据的增多,证据间的冲突程度逐渐降低,合成结果越确定。

5 结 论

针对传统 D-S 证据合成方法中冲突证据合成带来的问题,把 K-L 距离引入到 D-S 证据合成中,提出了一种新的冲突证据合成方法。该方法不但可以对高度冲突的证据进行合成,也可以用于基本证据合成,且合成结果符合实际推理,因此对于冲突证据的合成具有一定的理论与实际应用价值。下一步将相关理论应用到具体的工程项目中,以验证其在工程应用中的价值。

参考文献:

- [1] Yang J B, Singh M G. An evidential reasoning approach for multiple-attributed decision making with uncertainty [J]. IEEE Transactions on System, Man and Cybernetics, 2004, 34(1):1-18.
- [2] Smarandache F, Dezert J. Four versions of the proportional conflict redistribution rules of combination in information fusion [J]. Information Fusion, 2004, 41(3):386-395.
- [3] Yang J, Sen P. A general multi level evaluation process for hybrid MADM with uncertainty [J]. IEEE Transactions on System, Man and Cybernetics, 2006, 36(10):1458-1473.
- [4] 郭华伟,施文康,邓勇,等.证据冲突:丢弃,发现或化解[J].系统工程与电子技术,2007,29(6):890-898.
GUO Hua-wei, SHI Wen-kang, DENG Yong, et al. Evi-

- dential conflict and its 3D strategy: discard, discover and disassemble[J]. Systems Engineering and Electronics, 2007, 29(6):890-898. (in Chinese)
- [5] Duarte M, Hu Y H. Distance based decision fusion in a distributed wireless sensor network[J]. Telecommunication Systems, 2004, 26(2-4):339-350.
- [6] Xu Lijia, Chen Yangzhou, Cui Pingyuan. Improvement of D-S evidential theory in multi-sensor data fusion system[C]// Proceedings of the 6th World Congress on Intelligent Control and Automation. [S.l.]: IEEE, 2004:580-589.
- [7] Liu W. Analyzing the degree of conflict among belief functions [J]. Artificial Intelligence, 2006, 170(11):909-924.
- [8] Kullback Leibler divergence [EB/OL]. [2010-11-09]. http://en.wikipedia.org/wiki/Kullback-Leibler_divergence.
- [9] 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 Bi-cheng, WANG Bo, WEI Jun, et al. An efficient combination rule of evidence theory[J]. Journal of Data Acquisition & Processing, 2002, 17(1):33-36. (in Chinese)
- [10] Yager R R. On the Dempster-Shafer framework and new combination rules[J]. Information Sciences, 1987, 41(2):93-137.
- [11] 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(8):117-119.
SUN Quan, YE Xiu-qing, GU Wei-kang. A new combination rules of evidence theory[J]. Chinese Journal of Electronics, 2000(8):117-119. (in Chinese)

作者简介:

魏永超(1981-),男,河南禹州人,2009获四川大学博士学位,现为讲师,主要从事光电信息处理方面的研究。

WEI Yong-chao was born in Yuzhou, Henan Province, in 1981. He received the Ph.D. degree from Sichuan University in 2009. He is now a lecturer of Civil Aviation Flight University of China. His research concerns optical information processing.

Email:mylife001@126.com

欢迎订阅全国中文核心期刊《电讯技术》

邮发代号:62-39

全国各地邮局均可订阅!