一种新的基于证据理论的合成公式

孙 全, 叶秀清, 顾伟康
(浙江大学信息与通信工程研究所, 杭州 310027)

摘 要: 由于 D-S 证据合成公式所存在的不足, 使证据理论的应用受到了一定的限制. Yager 对此作了改进, 但改进后合成公式又存在着新的问题. 鉴于此, 本文在引入证据可信度概念的基础上, 提出了一个新的证据合成公式. 新的合成公式弥补了 D-S 证据理论和 Yager 合成公式所存在的不足, 使冲突证据合成的结果更为理想.

关键词: D-S 证据理论; 证据; 合成公式
中图分类号: O21, TP274 文献标识码: A 文章编号: 0372-2112(2000)08-0117-03

A New Combination Rules of Evidence Theory

SUN Quan, YE Xiu-qing, GU Wei-kang
(*Institute of Information and Communication Engineering, Zhejiang University, Hangzhou 310027, China*)

Abstract: D-S evidence theory is a very useful tool in dealing with the uncertainty problems, but its application is limited by the shortcomings of combination rules. Yager had provided some the modified rules, but his rules have some deficiencies. So in this article the new combination rules are proposed which are better than D-S' s and Yager' s rules in combining the conflicted evidences.

Key words: D-S theory of evidence; evidence; combination rules

1 引言

D-S 证据理论作为一种不确定性推理方法, 正在受到越来越多的关注. 这不仅是因为 D-S 证据理论比传统概率论能更好地把握问题的未知性和不确定性, 还因为 D-S 证据理论提供了一个非常有用的合成公式, 使我们能融合多个证据源提供的证据. 但是, 正如 Yager 在文[1]中指出的那样, D-S 合成公式存在着不足, 例如在合成高度冲突的证据时, 合成结果将有悖常理. Yager 因此提出了一个新的合成公式, 新公式应用在二个证据源时效果较好, 但当证据源多于二个时, 合成结果有时却并不理想.

本文较为详细地分析了 D-S 与 Yager 合成公式的优缺点, 并引入了证据可信度概念, 在此基础上提出了新的合成公式, 使不同冲突程度的证据的合成结果更为理想.

2 D-S 证据理论

D-S 证据理论首先定义一个空间 X, 称为辨识框架(Frame of Discernment), 由一些互斥且穷举的元素组成. 对于问题域中任何命题 A, 都应包含于 2^X. 定义映射 $m: 2^X \to [0, 1]$, 为基本概率赋值函数 BPAF(Basic Probability Assignment Function), m 满足: (1) $m(\Phi)=0$; (2) $0 \le m(A) \le 1$, $\forall A \subset X$; (3) $\sum_{A\subset X} m(A)=1$. 如果 $A \subset X$, 且 $m(A)>0$, 则称 A 为焦点元素(Focus Element). 在 D-S 证据理论中, 对事件 A 的描述采用区间[$Bel(A)$, $Pl(A)$], Bel 和 Pl 分别称为信任函数(Belief Function)和似然函数(Plausibility Function)[2,3].

D-S 证据理论提供了一个非常有用的合成公式, 使我们能合成多个证据源提供的证据. 公式定义如下:

$$m(A)=\frac{1}{1-k}\sum_{A_i\cap B_j\cap C_l\cap\cdots=A} m_1(A_i)\circ m_2(B_j)\circ m_3(C_l)\cdots \quad (1)$$

其中 $k=\sum_{A_i\cap B_j\cap C_l\cap\cdots=\phi} m_1(A_i)\circ m_2(B_j)\circ m_3(C_l)\cdots$, 它的大小反映了证据冲突程度. 系数 $1/(1-k)$ 称为归一化因子, 它的作用是为了避免在合成时将非零的概率赋给空集.

3 Yager 的合成公式

对于式(1), 当 $k=1$ 时将无法使用, 因为此时分母为零. 并且 $k\to 1$ 时, 即证据高度冲突时, 式(1)将会产生违反常理的结果. 设 $X=\{A, B, C\}$, 有二个 BPAF 如下:

例 1: m_1: $m_1(A)=0.99$, $m_1(B)=0.01$; m_2: $m_2(A)=0.01$, $m_2(B)=0.99$

由 D-S 合成公式计算出: $k=0.9999$, $m(A)=m(C)=0$, $m(B)=0.0001/(1-k)=1$. 尽管 m_1、m_2 对 B 的支持率都非常低, 但合成结果却认为 B 为真. 鉴于此, Yager 对 D-S 合成公式做了改进, 改进后公式如下(2 个证据源).

$$m(\Phi)=0$$

$$m(A)=\sum_{A_i\cap B_j=A} m_1(A_i)\circ m_2(B_j),\ A\ne\Phi,\ X$$

$$m(X)=\sum_{A_i\cap B_j=X} m_1(A_i)\circ m_2(B_j)+k$$

上式中去掉了归一化因子 $1/(1-k)$, 并且反映冲突程度的因

收稿日期: 1999-05-31; 修回日期: 1999-10-31

子 k 在合成后被赋给了 $m(X)$. Yager 认为: 既然我们对冲突的证据无法作出合理的抉择,就应将其归入未知邻域. 对于上面的例 1, Yager 合成的结果如下: $m(A)=m(C)=0$, $m(B)=0.0001$, $m(X)=0.9999$. 可以看出原来冲突的证据在合成后被完全否定,原来支持率低的证据在合成后支持率仍然比较低. 而几乎为 1 的概率合成后被分给了 X. 这实际上表明: 尽管有多个证据提供了证据,但由于它们是相互冲突的,因此我们仍然所知甚少.

当 $k=0$ 时,即证据没有冲突时, Yager 合成公式与 D-S 合成公式是一样的.

虽然 Yager 提出的合成公式能合成高度冲突的证据,但是由于对冲突的证据是完全否定的,因此在证据源多于二个时,合成结果有时并不理想. 下面的例 2 可以看出这一点: 设有二个证据源如下:

m_1: $m_1(A)=0.98$, $m_1(B)=0.01$, $m_1(C)=0.01$

m_2: $m_2(A)=0$, $m_2(B)=0.01$, $m_2(C)=0.99$

由 Yager 合成公式计算得: $k=0.99$, $m(A)=0$, $m(B)=0.0001$, $m(C)=0.0099$, $m(X)=0.99$. 如果有新证据加入: $m_3(A)=0.9$, $m_3(B)=0$, $m_3(C)=0.1$, 则 m_1、m_2、m_3 合成结果为: $k=0.99901$, $m(A)=0$, $m(B)=0$, $m(C)=0.00099$, $m(X)=0.99901$. 此后即使有更多的支持 A 的证据加入,但合成结果变化并不大: $m(A)=0$, $m(B)=0$, $m(C)\to 0$, $m(X)\to 1$. 可见虽然不断有新的证据支持 A, 但合成结果对 A 的支持率却始终为 0. 也就是: 尽管绝大多数的证据证明 A 是正确的,但由于某一个证据否定了 A, 合成结果也否定 A. 因此对于多传感器系统,会由于一个或少数传感器的出错,而导致整个系统无法正常工作. 所以有必要对 Yager 的合成公式作进一步的改进,以弥补这一点不足,这也是本文所做的工作.

4 新的合成公式

设 $m_1, m_2, \cdots, m_n$ 所对应的证据集为: $F_1, F_2, \cdots, F_n$, 并设证据集 i 和 j 之间的冲突大小为 k_{ij}, 则:

$$k_{ij}=\sum_{\substack{A_i\cap A_j=\Phi\\ A_i\in F_i, A_j\in F_j}} m_i(A_i)m_j(A_j)$$

定义 ε 为证据的可信度, $\varepsilon=e^{-\bar{k}}$, 其中 $\bar{k}=\frac{1}{n(n-1)/2}\sum_{i<j}k_{ij}$, $i, j\leq n$, n 为证据源的个数. $\bar{k}$ 是 n 个证据集中每对证据集总和的平均,它反映了证据两两之间的冲突程度. ε 是 $\bar{k}$ 的减函数,反映了证据的可信度,也就是当证据之间的冲突增大时,证据的可信度将降低. $\bar{k}$ 与 D-S 理论中 k 不同, k 反映证据总体上冲突程度,当 k 大时, $\bar{k}$ 不一定大. 例如对于前面的例 2, 假设有 4 个证据源,则 $k=0.99901$, $\bar{k}=0.504$, $\varepsilon=e^{-\bar{k}}=0.604$, 可见尽管证据在总体冲突较大,但 $\bar{k}$ 却不大,这是因为此时除少数证据外,大部分证据之间的冲突并不是很大. $\varepsilon=0.604$ 表明此时证据的可信度,并随着有利于 A 的证据不断加入而增大.

新的合成公式定义如下:

$$\begin{aligned}&m(\Phi)=0\\&m(A)=p(A)+k*\varepsilon*q(A),\ A\neq\Phi, X\\&m(X)=p(X)+k*\varepsilon*q(X)+k(1-\varepsilon)\end{aligned}\tag{2}$$

其中:

$$p(A)=\sum_{\substack{A_i\in F_i\\ \cap_{i=1}^{n}A_i=A}} m_1(A_1)m_2(A_2)\cdots m_n(A_n),\ q(A)=\frac{1}{n}\sum_{i=1}^{n}m_i(A).$$

式(2)又可写成如下形式:

$$m(A)=(1-k)\frac{p(A)}{1-k}+k*\varepsilon*q(A)\tag{3}$$

可以发现其中第一项的 $p(A)/(1-k)$ 正是 D-S 证据合成公式. 因此,新的合成公式实际上是一个加权和形式, $1-k$ 和 k 是加权系数. 当 k 较小时,即证据冲突较小,式(3)中第一项起主要作用,合成结果近似于 D-S 合成结果. 当 $k=0$ 时,新的合成公式等同于 D-S 的合成公式; 当 $k\to 1$ 时,即证据高度冲突时,合成结果将主要有式(3)中的 $\varepsilon*q(A)$ 决定. ε 为证据可信度, $q(A)$ 为证据对 A 的平均支持度,因此,对于高度冲突的证据,其合成结果将主要由证据可信度 ε 和证据平均支持度 $q(A)$ 的乘积决定.

与 Yager 一样,我们也认为由于证据冲突的存在,将导致未知程度的增加,这反映在式(2)中的 $k(1-\varepsilon)$ 项,当冲突 k 增大或证据可信度 ε 的减小,都会使未知程度增加.

新的合成公式与 Yager 的主要区别在: Yager 在合成冲突证据时,将支持冲突证据的那部分概率全部赋给了 X, 即认为冲突的证据不能提供任何有用信息. 而我们认为,即使证据之间存在着冲突,它们也是部分可用的,并且可用程度取决于证据的可信度 ε.

5 实验结果

本节中我们将三个合成公式的合成结果列表作比较. 对于前面的例 1, 三个公式的合成结果见表 1.

表 1

	K	$\bar{k}$	ε	$m(A)$	$m(B)$	$m(C)$	$m(X)$
D-S 证据理论合成公式	0.9999	—	—	9	1	0	0
Yager 合成公式	0.9999	—	—	0	0.0001	0	0.9999
本文的合成公式	0.9999	0.9999	0.3679	0.182	0.004	0.182	0.632

表 2

	m_1, m_2	m_1, m_2, m_3	m_1, m_2, m_3, m_4
D-S 合成公式	$k=0.99$, $m(A)=0$, $m(B)=0.01$, $m(C)=0.99$, $m(X)=0$	$k=0.99901$, $m(A)=m(B)=0$, $m(C)=1$, $m(X)=0$	$k=0.999901$, $m(A)=m(B)=0$, $m(C)=1$, $m(X)=0$
Yager 合成公式	$k=0.99$, $m(A)=0$, $m(B)=0.0001$, $m(C)=0.0099$, $m(X)=0.99$	$k=0.99901$, $m(A)=m(B)=0$, $m(C)=0.00099$, $m(X)=0.99901$	$k=0.999901$, $m(A)=m(B)=0$, $m(C)=0.000099$, $m(X)=0.999901$
本文合成公式	$\varepsilon=0.3716$, $m(A)=0.18$, $m(B)=0.004$, $m(C)=0.194$, $m(X)=0.622$	$\varepsilon=0.512$, $m(A)=0.321$, $m(B)=0.003$, $m(C)=0.188$, $m(X)=0.488$	$\varepsilon=0.604$, $m(A)=0.42$, $m(B)=0.003$, $m(C)=0.181$, $m(X)=0.396$

可以看出：对于高度冲突的证据 A 和 C，D-S 和 Yager 的合成结果为 0，而本文公式合成结果不为 0，但由于此时证据可信度 ε 较低，合成后它们的支持率也是比较低的．总体上这三个公式得出的结果基本是一致的．

对于前面的例 2，三个公式后的合成结果见表 2．

在本例中可以发现，随着高度支持 A 的证据的不断加入，ε 在增大的同时 $m(A)$ 也在增大，而 D-S 和 Yager 的合成结果中 $m(A)$ 却是始终为 0 的，相比可见，新的合成公式输出结果更为合理．

6　结论

由于采用了加权和的形式，新的合成公式能有效地合成冲突证据．同时证据可信度概念的引入，使我们对于冲突的证据不再是盲目地全盘否定，而是根据可信度将其一分为二，这样极大地提高了合成结果的可靠性．通过比较可以发现，新的合成公式不仅能有效地合成高度冲突的证据，并且合成效果好于 Yager 公式．

参考文献：

[1]　Ronald R. Yager. On the dempster-shafer framework and new combination rules [J]. Information Sciences, 1987, 41: 93—137.

[2]　G. Shafer. A mathematical theory of evidence [M]. Princeton U. P., Princeton, 1976.

[3]　A. P. Dempster. Upper and lower probabilities induced by a multi-valued mapping [J]. Ann. Math. Statist. 1967, 38: 325—339.

作者简介：

孙　全　1992 年毕业于浙江大学光电与科学仪器工程学系，获学士学位．1992～1994 年工作于南京五二八厂．1997 年毕业于浙江大学光电与科学仪器工程学系，获硕士学位．1997 年在浙江大学信电系攻读博士学位．研究方向：多传感器信息融合，图像处理．

叶秀清　教授，浙江大学智能与信息系统研究所，研究方向为图像处理，计算机视觉．

顾伟康　教授，博士生导师，浙江大学副校长，智能与信息系统研究所，研究方向为模式识别，视觉智能机器人．