第8卷第5期
2005年10月

管理科学学报
JOURNAL OF MANAGEMENT SCIENCES IN CHINA

Vol.8 No.5
Oct. 2005

基于 Rough Sets 理论的证据获取与合成方法①

杨善林，刘业政，李亚飞
(合肥工业大学管理学院，合肥 230009)

摘要：证据理论是处理不确定性问题的有力工具，它处理的证据来源于专家.专家的知识经验是有限的，获取较困难，且可能存在一定的主观性.针对上述问题，提出了一种基于粗糙集理论的证据获取的新方法，并对证据合成和应用进行了研究.首先研究了大型决策表分解问题.利用粗糙集理论分析条件属性间的依赖关系，对条件属性集进行聚类，形成多个条件属性集相对独立的子决策表；其次对各子决策表进行分析，利用粗糙集的分类思想和隶属度概念，计算证据的基本可信度分配；最后文章对证据的合成及其在决策分析中的应用进行了研究，提出了相应的解决方法.

关键词：证据理论；粗糙集；依赖度；基本可信度分配；信度函数

中图分类号：C931.9　　**文献标识码**：A　　**文章编号**：1007－9807(2005)05－0069－07

0 引 言

证据理论[1]的应用，主要存在两个困难，一是证据获取，二是证据合成.Shafer 认为证据并不是实证据，而是指专家利用经验、知识对问题观察和研究的结果.对于某个实际问题，领域专家的经验、知识以及他对该问题的观察、研究都是推理的证据.目前证据理论中的证据获取主要是向专家咨询，根据专家的知识、经验在辨别框架上产生一个基本可信度分配[2].由于是依赖专家提供证据，这必然面临下列一些问题：(1)专家的知识、经验是有限的；(2)专家的知识、经验往往有一定的主观性，特别在面向决策领域时个人倾向会更明显；(3)专家的知识、经验获取有时是很困难的.在证据合成方面，Dempster 为证据的合成作出了奠基性的贡献[3,4].但 Dempster 合成法则也存在明显的局限性[5]：(1)它要求被合成的证据是独立的证据，这往往不符合客观实际，当两个证据具有相关性时，应用 Dempster 合成法则进行证据合成，将会导致结果的超估计；(2)Dempster 合成法则对各信息源提供的证据是平等对待的，在合成时各信息源提供证据的重要性与可靠性无优劣之分，这也与客观实际不太相符，特别当出现证据高度冲突时，其合成结果会出现逆序或有悖常理的现象.这些局限性严重影响了该合成法则在实际问题中的应用.针对 Dempster 合成法则的局限性，人们做了很多的研究，也取得了一些成果[6~10]，这些成果不同程度地克服了 Dempster 合成法则的缺陷，并且在一些领域取得了较好的应用，但也存在着明显的缺陷[11].

Pawlak[12]认为，决定人类行为的一个基本特征是对对象的分类能力.而分类既是 Rough 集理论的核心，也是证据理论的核心[13,14].有鉴于此，利用信息系统所积累的数据和粗糙集理论，本文提出了一种获取独立证据的方法，使证据满足 Dempster 合成法则.首先应用粗糙集理论分析条件属性间的依赖关系，对条件属性集进行聚类，形成多个条件属性集相对独立的子决策表；其次对各子决策表进行分析，利用粗糙集的分类思想和决策表对象间的隶属关系[15]对辨识框分析，获取

① 收稿日期：2003－09－25；修订日期：2005－06－12.
基金项目：教育部科学技术研究资助重点项目(02127)；安徽省重点研究资助项目(03021057).
作者简介：杨善林(1948—)，男，安徽怀宁人，教授，博士生导师.

证据在辨识框上的基本可信度分配,从而解决传统证据理论中证据获取困难、证据独立性问题和由于专家的经验判断所带来的主观不确定问题.最后利用 Rough 集理论分析证据的可靠度或支持度以及不同证据的重要度,修正 Dempster 合成规则.

1 决策表

对于一个具有两个或两个以上决策属性的决策表,可以把它转化为具有单决策属性特征的决策表[16],所以本文只考虑具有单决策属性的决策表.

定义 设有限集合 U 为论域,C 为关于 U 的条件属性集,d 为决策属性,则决策表可表示为 $I=(U,C\cup\{d\})$.其 s 行 a 列的项记为 $a(s)$.称 $d(U)=\{k\in N, s\in U: d(s)=k\}$ 为 d 的阶,记为 $r(d)$.d 的值域记为 $V_d=\{1,2,\cdots,r(d)\}$.

根据决策属性 d 可以对决策表 I 作如下划分 $CL_I(d)=\{X_1,X_2,\cdots,X_{r(d)}\}$,其中

$$X_k=d^{-1}(\{k\})=\{s\mid s\in U\wedge d(s)=k\},\quad 1\leqslant k\leqslant r(d) \tag{1}$$

那么决策表 I 由 d 决定的两个一一对应集合

$$\Theta_I=\{1,2,\cdots,r(d)\}\text{和}\quad CL_I(d)=\{X_1,X_2,\cdots,X_{r(d)}\} \tag{2}$$

称为 d 辨识框和 d 分类.

设 $X\subseteq U$ 是任一非空子集,R 是 U 上的等价关系,R 的所有等价类的集合记为 U/R,含有 x 元素的等价类记为 $[x]_R$,则 X 的下近似和上近似分别为

$$\underline{R}X=\{x\in U:[x]_R\subseteq X\}\text{和}\quad \overline{R}X=\{x\in U:[x]_R\cap X\neq\varnothing\} \tag{3}$$

X 的 R-边界域(R-boundary region)为 $BN_R(X)=\overline{R}X-\underline{R}X$.

粗糙集理论中,Pawlak 定义了两个依赖函数

$$\gamma_R=\frac{|\underline{R}X|}{|U|}\text{和}\ \overline{\gamma}_R=\frac{|\overline{R}X|}{|U|} \tag{4}$$

从证据理论的角度看,γ_R 和 $\overline{\gamma}_R$ 分别是 X 的信度函数和似然函数[15].由于证据理论将命题概率推断的基础建立在结论的幂集上而不是结论本身,将命题推断转化为集合运算,将从客观问题获取的知识(证据)用分布于$[0,1]$之间的数值描述命题为真的信任程度并依此支持决策,因此基于这种思想,本文利用粗糙集方法对证据进行分类,再利用上面的两个依赖函数过渡到证据理论中,对信息进行处理.

2 决策表分解

在实际决策中,涉及的条件属性很多,且这些属性间往往存在相互依赖的关系,显然无论是进行粗糙集数据分析还是证据合成,这都是不恰当的.特别地,作为证据理论中主要的合成规则——Dempser 合成规则,它要求被合成的证据之间必须是独立的,为了利用这个主要的合成法则,本文提出了一种基于属性依赖度的属性聚类方法,对决策表进行分解即:利用属性聚类,把决策表的条件属性划分成几个相互独立的类,形成几个比较简单的决策表,然后对简单决策表进行证据提取,再按照 Dempster 合成规则进行证据合成.

属性聚类在数据分析中又叫变量聚类、$\boldsymbol{R}$ 型聚类,它一般是从数据的相关矩阵 $\boldsymbol{R}$ 出发,经如下变换

$$\boldsymbol{R}=\begin{bmatrix}1 & & & & \\ r_{21} & 1 & & & \\ r_{31} & r_{32} & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ r_{p1} & r_{p2} & r_{p3} & \cdots & 1\end{bmatrix}\xrightarrow{d_{ij}=f(r_{ij})}\boldsymbol{D}=\begin{bmatrix}0 & & & & \\ d_{21} & 0 & & & \\ d_{31} & d_{32} & 0 & & \\ \vdots & \vdots & \vdots & \ddots & \\ d_{p1} & d_{p2} & d_{p3} & \cdots & 0\end{bmatrix} \tag{5}$$

然后可以采取谱系聚类的方法,实现对属性的聚类.

由于需要分析的决策表中各属性通常是离散的,因此表征连续属性相关性的相关系数在这里并不太合适.为了解决这个问题,借鉴粗糙集理论中的依赖度思想,以任意两属性间的依赖度代替相关矩阵中的相关系数.所谓依赖,它表征的是两属性集之间的一种支持关系,由于依赖有高低程度之分,那么就需要依赖度 k 这样一种指标来衡

量其度的大小：设有属性集(或等价关系)D 和 C，若 D 以依赖度 k 依赖于 C，则

$$k = \gamma(C,D) = \sum_{X \in U/D} \frac{|\underline{C}(X)|}{|U|} \tag{6}$$

设决策表 $I=(U, C \cup \{d\})$，条件属性分别为 $c_1, c_2, \cdots, c_p$，决策属性为 d. 从集合的角度看，每一条件属性又形成一等价关系. 那么等价关系 c_i、c_j 划分决策表分别为 U/c_i 和 U/c_j，则属性 c_i 对 c_j 的依赖度

$$k_{ij} = \gamma(c_i, c_j) = \sum_{X \in U/c_j} \frac{|\underline{c_i}(X)|}{|U|} \tag{7}$$

根据这个结果，式(5)就可转化为以 k_{ij} 为元素的依赖矩阵 $\boldsymbol{R}$. 由于 k_{ij} 通常与 k_{ji} 并不相等，因此依赖矩阵 $\boldsymbol{R}$ 不是对称阵，为此做如下转化

$$k'_{ij} = \max\{k_{ij}, k_{ji}\} \tag{8}$$

使依赖矩阵 $\boldsymbol{R}$ 转为对称阵 $\boldsymbol{R}'$. 令 $d_{ij} = f(k'_{ij}) = 1 - k'_{ij}$，则 $\boldsymbol{R}'$ 可转化为式(5)的距离阵 $\boldsymbol{D}$，这样就可以利用谱系聚类的方法，将 C 划分成几个相对独立的条件属性子集 C_l，$(l = 1,2,\cdots,n)$ 且 $C_l \cap C_k = \emptyset$，$\bigcup_{l=1}^{n} C_l = C$. 至此每个相对独立的条件属性集 C_l 与决策属性 d 就可以构成简单决策表 $I_l = (U, C_l \cup \{d\})$，$(l = 1,2,\cdots,n)$.

3 证据获取

3.1 证据信息的获取

对于每一简单决策表(为了表述方便，仍用 $I = (U, C \cup \{d\})$ 表示)，所有条件属性的并可以看作是一等价关系 C，那么对于任意 $X_i \in U/d, i = 1,2,\cdots,r(d)$，$X_i$ 关于 C 的下近似 $\underline{C}X_i$ 中的对象明确支持其决策值 i. 而边界域中的对象对决策来说是模糊不确定的，因此难以满足证据理论的需要. 考虑到粗糙集与证据理论的联系与差别，可以利用等价关系 C 对其作进一步的划分

$$I = \{C(\theta) : C(\theta) = \bigcap_{i \in \theta} \overline{C}X_i - \bigcup_{\theta' \supset \theta}(\bigcap_{j \in \theta'} \overline{C}X_j), \quad \theta', \theta \subseteq \Theta_I\} \tag{9}$$

其中：Θ_I 表示决策表 I 的辨识框；θ、θ' 表示辨识框中的元素，它可以是决策值的任一决策组合. 这样可以变对象模糊为决策目标模糊，从而可以把获取的证据用到证据理论中. 式(9)中，等价关系 C 明确支持的对象都是在 $\underline{C}X_i (i = 1,2,\cdots,r(d))$ 中，而这些下近似是没有交集的，因此对上式可以作进一步的变换

$$I = \{\underline{C}X_1, \cdots, \underline{C}X_{r(d)}\} \cup \{Bd_C(\theta) : \theta \subseteq \Theta_I, |\theta| > 1\} \tag{10}$$

其中：

$$Bd_C(\theta) = \bigcap_{i \in \theta} BN_C(X_i) - \bigcup_{\theta' \supset \theta}(\bigcap_{j \in \theta'} BN_C(X_j)),$$

其中 $\theta', \theta \subseteq \Theta_I$，$|\theta| > 1$

至此决策表中的任一对象，都可以根据其属性值进行归类：如果某一对象包含于 $\underline{C}X_i$，那么该对象精确归属到 X_i 类，否则它就应划分到相应的 $Bd_C(\theta)$ 中，因为没有足够的信息去确定它到底属于哪个集合 X_i，只能说它属于 $X_i(i \in \theta)$ 的边界域，也即它的决策值是模糊的.

对于任一决策组合 θ，它代表的信息可以表示为式(11).

$$F_I(\theta) = \begin{cases} \underline{C}X_i & \theta = \{i\}, 1 \leqslant i \leqslant r(d) \\ \emptyset & \theta = \emptyset \\ Bd_C(\theta) & |\theta| > 1 \end{cases} \tag{11}$$

为了表述方便，定义一个映射 $\delta_C : U \to P(\Theta_I)$，则 $\delta_C(s)$(其中 $s \in F_I(\theta)$)就是 Θ_I 中 θ 的一个子集即 $\delta_C(s) \subseteq \theta$，上式表明若 $\theta = \{i\}$，那么 $s \in \underline{C}X_i$，如果 $|\theta| > 1$，则 $s \in Bd_C(\theta)$. 这样函数 δ_C 就可以作为一个决策属性来考虑，即对于任一决策表 $I = (U, C \cup \{d\})$ 都可以定义这样一个新的决策表 $I' = (U, C \cup \{\delta_C\})$ 与之相对应.

令 $\Theta = \{\theta_i, \cdots, \theta_k\}$ 为与决策表 I 相对应的辨识框，$\chi : \Theta \to \Theta_I$ 为 Θ 到 Θ_I 的标准映射，即 $\chi(\theta_i) = i, i = 1,2,\cdots,k$，则

$$m_I(\theta) = \frac{|F_I(\chi(\theta))|}{|U|} \tag{12}$$

为决策表 I 的基本可信度分配. 有了基本可信度分配函数，根据 $Bel_I(\theta) = \sum_{\Delta \subseteq \theta} m(\Delta)$、$Pl_I(\theta) = \sum_{\Delta \cap \theta \neq \emptyset} m(\Delta)$ 便可分别得到信度函数和似然函数.

3.2 示例

设决策表 A 为分解后的一个简单决策表，如表 1.

表 1 决策表 A

Table 1 Decision table A

U \ A	a	b	c	e	d	U \ A	a	b	c	e	d
1	1	2	2	2	1	13	1	2	2	1	3
2	1	2	2	2	1	14	2	1	1	1	3
3	2	1	2	2	1	15	2	2	2	1	3
4	3	2	2	2	1	16	2	2	2	1	3
5	1	1	1	1	2	17	2	2	2	1	3
6	1	2	1	2	2	18	2	2	2	2	3
7	2	1	1	2	2	19	3	1	2	1	3
8	2	2	1	2	2	20	3	1	1	2	3
9	3	2	1	2	2	21	3	1	2	1	3
10	1	1	1	1	3	22	3	2	1	2	3
11	2	1	2	1	3	23	3	2	2	2	3
12	1	2	1	2	3	24	3	2	2	2	3

令 $X_i = d^{-1}(\{i\}) = \{k \in U : d(k) = i; i = 1,2,3\}$，则

$\underline{A}X_1 = \{1,2,3\}$; $\underline{A}X_2 = \{7,8\}$; $\underline{A}X_3 = \{11,13,14,15,16,17,18,19,20,21\}$

$Bd_A(\{1,2\}) = \{\emptyset\}$; $Bd_A(\{1,3\}) = \{4,23,24\}$; $Bd_A(\{2,3\}) = \{5,6,9,10,12,22\}$; $Bd_A(\{1,2,3\}) = \{\emptyset\}$

那么表 1 就可以转换为以 δ_A 为决策属性的决策表，见表 2.

表 2 转换后的中间决策表

Table 2 Decision table with decision attribute

U \ A	a	b	c	e	δ_A	U \ A	a	b	c	e	δ_A
1	1	2	2	2	{1}	13	1	2	2	1	{3}
2	1	2	2	2	{1}	14	2	1	1	1	{3}
3	2	1	2	2	{1}	15	2	2	2	1	{3}
4	3	2	2	2	{1,3}	16	2	2	2	1	{3}
5	1	1	1	1	{2,3}	17	2	2	2	1	{3}
6	1	2	1	2	{2,3}	18	2	2	2	2	{3}
7	2	1	1	2	{2}	19	3	1	2	1	{3}
8	2	2	1	2	{2}	20	3	1	1	2	{3}
9	3	2	1	2	{2,3}	21	3	1	2	1	{3}
10	1	1	1	1	{2,3}	22	3	2	1	2	{2,3}
11	2	1	2	1	{3}	23	3	2	2	2	{1,3}
12	1	2	1	2	{2,3}	24	3	2	2	2	{1,3}

辨识框 $\Theta = \{\theta_1, \theta_2, \theta_3\}$，则可得基本可信度分配 $m_A(\theta)$ 和信度函数 $Bel_A(\theta)$ 如表 3：

表 3 基本可信度分配和信度函数

Table 3 Basic probability assignment and belief function

θ	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_3\}$	$\{\theta_1,\theta_2\}$	$\{\theta_1,\theta_3\}$	$\{\theta_2,\theta_3\}$	$\{\theta_1,\theta_2,\theta_3\}$
$\chi(\theta)$	{1}	{2}	{3}	{1,2}	{1,3}	{2,3}	{1,2,3}
$m_A(\theta)$	1/8	1/12	5/12	0	1/8	1/4	0
$Bel_A(\theta)$	1/8	1/12	5/12	5/24	2/3	3/4	1

4 证据合成及决策分析

4.1 证据重要度

由于 Dempster 合成规则将来自于不同信息源的证据平等对待，对冲突证据的合成效果较差. 实际上，不同的证据一般有不同的重要性，其对决策的支持因此也不同. 鉴于此，受粗糙集理论中属性重要度思想的启发，本文把它引入到证据合成中来，以此衡量证据的重要性问题，从而改善冲突证据的合成效果.

定义(属性重要度) 设决策表 $I = (U, C \cup D)$，对于 C 的非空子集 B，其重要度为

$$\lambda_B = \sigma_{(C,D)}(B) = \frac{\gamma_C(D) - \gamma_{C-B}(D)}{\gamma_C(D)} = 1 - \frac{\gamma_{C-B}(D)}{\gamma_C(D)} \tag{13}$$

其中：γ 为近似质量.

则对于给定的决策表 $I = (U, C \cup \{d\})$，

$$\lambda_l = \sigma_{(C,d)}(C_l) = \frac{\gamma_C(d) - \gamma_{C-C_l}(d)}{\gamma_C(d)} = 1 - \frac{\gamma_{C-C_l}(d)}{\gamma_C(d)} \tag{14}$$

其中：C_l 是决策表 I 分解后子决策表 I_l 的条件属性集，λ_l 即为来自于子决策表 I_l 的证据的重要度.

4.2 证据支持度

传统证据理论中，证据信息是专家根据已知的信息(条件、状况)给出的关于决策目标的基本可信度分配. 而本文所求出的基本可信度分配和信度函数，是对历史数据的归纳，对目前的决策来说是可能决策组合 θ 的先验概率. 因而传统证据理论中的决策支持方法在这里并不适用. 为此本文提出一种利用粗糙集理论中分明矩阵的思想实现决策支持的方法.

设子决策表 $I_l = (U, C_l \cup \{d\})(l = 1,2,\cdots,n; C_l = \{c_1, c_2, \cdots, c_m\})$ 所对应的中间决策表为 $I'_l = (U, C_l \cup \{\delta_{C_l}\})$，对于待决策问题 x，构建 $m \times 2$ 阶矩阵 $\boldsymbol{M}(I'_l) = (b_{ij})$，称其为 I'_l 的关于 $F_{I'_l}(\theta)$ 隶属度分明矩阵，其中

$$b_{i1} = \{c_i \in c : c_i(u) \neq c_i(v), u, v \in U\} \tag{15}$$

$$b_{i2} = 1 - \frac{|c_i|}{k} \tag{16}$$

对于任一 $\theta \subseteq \Theta$，都有 $s_l(\theta) = \max\{b_{i2}\}$，称 $s_l(\theta)$ 是第 l 个子决策表对 θ 的支持度.

4.3 证据合成与决策支持

设 Bel_l 是子决策表 $I_l = (U, C_l \cup \{d\})(l = 1,2,\cdots,n)$ 在辨识框 Θ 上的信度函数，m_l 是对应的基本可信度分配，如果 $\sum\limits_{\substack{\theta_1 \cap \theta_2 \cap \cdots \cap \theta_n = \emptyset \\ \theta_1, \theta_2, \cdots, \theta_n \subseteq \Theta}} m_1(\theta_1) m_2(\theta_2) \cdots m_n(\theta_n) <$

1，且基本可信度分配为 m，则

$$m(\theta)=\begin{cases}0 & \theta=\varnothing\\ K\sum\limits_{\substack{\theta_1,\theta_2,\cdots,\theta_n\subseteq\Theta\\ \theta_1\cap\theta_2\cap\cdots\cap\theta_n=\varnothing}} s_1(\theta_1)\lambda_1 m_1(\theta_1)\cdots \\ \qquad s_n(\theta_n)\lambda_n m_n(\theta_n) & \theta\neq\varnothing\end{cases} \tag{17}$$

其中：$K=\left(\begin{array}{l}1-\sum\limits_{\substack{\theta_1,\theta_2,\cdots,\theta_n\subseteq\Theta\\ \theta_1\cap\theta_2\cap\cdots\cap\theta_n=\varnothing}} s_1(\theta_1)\lambda_1 m_1(\theta_1)\cdots\\ \qquad s_n(\theta_n)\lambda_n m_n(\theta_n)\end{array}\right)^{-1}$，$\lambda_l$ 是来自第 l 个子决策表证据的重要度，$s_l(\theta)$ 为第 l 个子决策表对决策问题关于 θ 的支持度．这就是本文得到的证据合成规则．

最大信度函数 $Bel(\theta_i)$ 是期望的结果，如果 $|\theta_i|=1$，那么对决策最有利，若 $|\theta_i|>1$，则对决策帮助不大，因此希望通过得到的集合 θ_i 及其信度 $Bel(\theta_i)$ 找到真值，或将所得到的集合缩小，使得包含真值的集合元素个数达到最少．基本思想是：对于集合 θ_i，去掉若干个元素以后为集合 θ'_i，信度为 $Bel(\theta'_i)$，如果 $Bel(\theta_i)$ 与 $Bel(\theta'_i)$ 相差不大，说明去掉这几个元素对信度影响不大，因此可以移去这几个元素．详细算法参考文献[17,3]．

5　实证研究

针对本文所提的方法，取 2003 年 5 月 12 号到 2003 年 9 月 30 号之间 100 个交易日的 10 只股票，每个股票取 14 个指标属性作为本次的试验数据，决策值按每日的涨跌幅分为暴跌（日跌幅 > 1%）、下跌（1% > 日跌幅 > 0.5%）、不变（－0.5% < 日涨跌 < 0.5%）、上涨（0.5% < 日涨幅 < 1%）和暴涨（日涨幅 > 1%）五种情况．对于每只股票取前 60 个交易日交易数据作为训练数据，后 40 个作为测试集．在经过预处理、决策表分解、证据信息提取和证据合成之后，对每条测试记录的测试结果进行验证．正确率为 40.29%，比 5 种情况平均预测准确率 20% 高出一倍．

部分源数据如图 1 所示．部分证据合成结果如图 2 所示（其中 result1，result2，result3… 代表决策值 1、2、3…，表中数值是对应证据的基本可信度分配）．

stock name	stock code	Datel	MA5	MA10	MA20	Total	DIF	MACD	J	K	D	MB	UPPER1	LOW1
中兴通讯	000063	2003.9.12	18.61	18.99	19.03	979	-0.18	-0.13	12.86	21.97	26.52	19.03	19.81	18.20
中兴通讯	000063	2003.9.15	18.53	18.88	18.97	1 076	-0.22	-0.16	2.21	16.10	23.05	18.97	19.77	18.10
中兴通讯	000063	2003.9.16	18.52	18.77	18.91	1 445	-0.23	-0.15	4.17	14.96	20.35	18.91	19.68	18.1
中兴通讯	000063	2003.9.17	18.41	18.62	18.85	1 842	-0.27	-0.19	-3.87	9.97	16.89	18.85	19.70	17.9
中兴通讯	000063	2003.9.18	18.27	18.46	18.97	1 625	-0.30	-0.2	-1.99	8.80	14.19	18.79	19.71	17.8
中兴通讯	000063	2003.9.19	18.15	18.38	18.73	688	-0.32	-0.18	2.29	9.09	12.49	18.73	19.67	17.7
中兴通讯	000063	2003.9.2	19.06	19.05	19.41	3 359	-0.07	-0.07	79.64	46.72	30.26	19.41	20.30	18.5
中兴通讯	000063	2003.9.22	18.10	18.32	18.70	490	-0.33	-0.16	6.99	10.14	11.71	18.70	19.68	17.7
中兴通讯	000063	2003.9.23	18.03	18.27	18.66	1 319	-0.33	-0.13	28.55	18.93	14.11	18.66	19.68	17.6
中兴通讯	000063	2003.9.24	18.09	18.25	18.63	1 937	-0.31	-0.07	55.42	31.82	20.01	18.63	19.67	17.6
中兴通讯	000063	2003.9.25	18.15	18.21	18.62	987	-0.29	-0.02	68.45	40.77	26.93	18.62	19.66	17.5
中兴通讯	000063	2003.9.26	18.21	18.18	18.58	430	-0.27	0.01	84.29	51.51	35.13	18.58	19.62	17.5
中兴通讯	000063	2003.9.29	18.20	18.15	18.52	1 461	-0.27	0	65.19	48.01	39.42	18.52	19.51	17.5
中兴通讯	000063	2003.9.3	19.21	19.08	19.38	890	-0.04	-0.01	82.75	52.76	37.76	19.38	20.21	18.5
中兴通讯	000063	2003.9.30	18.23	18.13	18.45	652	-0.26	0.03	71.90	53.34	44.06	18.45	19.34	17.5
中兴通讯	000063	2003.9.4	19.39	19.12	19.35	906	-0.01	0.04	87.51	59.08	44.87	19.35	20.12	18.5
中兴通讯	000063	2003.9.5	19.37	19.08	19.31	4 340	-0.04	-0.02	49.24	46.74	45.49	19.31	20.08	18.5
中兴通讯	000063	2003.9.8	19.23	19.08	19.26	2 515	-0.08	-0.08	19.56	34.38	41.79	19.26	20.05	18.4
中兴通讯	000063	2003.9.9	19.02	19.04	19.18	1 358	-0.13	-0.14	5.08	26.06	36.54	19.18	19.97	18.3
格力电器	000651	2003.5.12	8.90	9.08	9.27	1 680	-0.01	-0.18	27.29	24.14	22.57	9.27	10.13	8.41
格力电器	000651	2003.5.13	8.85	8.95	9.25	2 216	-0.06	-0.21	12.38	18.20	21.11	9.25	10.15	8.35
格力电器	000651	2003.5.14	8.76	8.88	9.23	1 760	-0.09	-0.22	18.55	20.01	20.75	9.23	10.17	8.29
格力电器	000651	2003.5.15	8.79	8.83	9.22	2 080	-0.09	-0.17	57.62	36.55	26.01	9.22	10.18	8.27
格力电器	000651	2003.5.16	8.78	8.82	9.21	1 646	-0.08	-0.13	78.29	48.42	33.48	9.21	10.18	8.29
格力电器	000651	2003.5.19	8.76	8.83	9.20	2 240	-0.07	-0.09	95.65	60.13	42.36	9.20	10.16	8.23
格力电器	000651	2003.5.20	8.82	8.84	9.16	1 261	-0.08	-0.08	85.29	60.76	48.49	9.16	10.13	8.19
格力电器	000651	2003.5.21	8.86	8.81	9.11	1 089	-0.08	-0.07	77.11	60.76	52.58	9.11	10.03	8.19

图 1　部分实验源数据

Fig.1 Source data

6 结 论

证据理论作为处理不确定性信息的一种有效方法,是当前国内外人工智能领域的热点研究课题,其理论和技术都得到了很大的发展.本文提出了基于粗糙集理论的证据信息获取方法,克服了传统证据理论中证据信息获取的局限性;并对所获证据的合成和决策支持进行了研究.研究成果拓宽了证据理论中证据信息的来源,能较有效地克服 Dempster 合成法则的不足,具有较大的理论意义和实际应用价值.

result1	result2	result3	result4	result5	result12	result13	result14	result15	result23	result24	result25	result34	result35	result45	result
0.88	2.09	4.31	1.35	0.95	0.33	1.37	0.15	0.16	1.87	0.94	0.44	1.90	0.79	0.23	0.66
0.95	2.07	3.71	1.22	1.05	0.32	1.30	0.11	0.23	1.69	0.89	0.46	1.51	0.77	0.24	0.65
0.96	2.54	3.55	2.04	1.16	0.59	1.56	0.23	0.08	1.52	2.02	0.39	1.70	1.19	0.35	0.65
0.81	1.58	3.63	1.33	1.17	0.32	1.62	0.15	0.13	1.89	0.98	0.52	2.08	1.12	0.37	0.80
1.06	2.64	4.63	2.01	0.72	0.62	1.37	0.11	0.17	2.59	1.93	0.56	1.72	1.53	0.22	0.56
0.65	2.92	2.83	2.00	0.85	0.54	0.87	0.17	0.16	1.53	1.37	0.52	1.72	1.09	0.34	0.37
0.85	1.62	3.68	1.85	1.18	0.54	1.62	0.18	0.13	1.88	1.36	0.53	2.12	1.16	0.31	0.61
0.83	2.30	3.65	1.88	0.94	0.30	1.29	0.17	0.14	2.46	1.00	0.67	2.48	0.85	0.38	0.65
1.14	2.25	3.50	1.51	1.05	0.33	1.22	0.16	0.15	1.40	1.34	0.52	2.10	0.79	0.23	0.39
1.10	2.57	3.14	2.52	1.21	0.96	1.68	0.07	0.25	2.26	1.17	0.39	1.65	1.12	0.42	0.91
0.75	2.01	2.32	1.54	0.81	0.42	0.81	0.05	0.15	1.67	0.45	0.23	0.71	1.08	0.25	0.57
0.75	2.01	2.32	1.54	0.81	0.42	0.81	0.05	0.15	1.67	0.45	0.23	0.71	1.08	0.25	0.57
0.91	2.40	2.92	2.27	1.23	0.74	1.18	0.07	0.23	2.51	1.07	0.51	1.54	1.79	0.42	0.77
0.95	2.56	3.17	2.41	1.19	0.77	1.19	0.07	0.25	2.66	1.19	0.50	1.67	1.67	0.39	0.66
0.86	2.17	3.26	2.22	1.02	0.64	1.03	0.13	0.19	2.46	0.92	0.43	1.50	1.47	0.39	0.77
0.86	2.17	3.26	2.22	1.02	0.64	1.03	0.13	0.19	2.46	0.92	0.43	1.50	1.47	0.39	0.77
0.36	0.84	1.87	0.85	0.50	0.29	0.92	0.07	0.09	0.99	0.69	0.23	0.75	0.56	0.10	0.43
0.40	0.83	1.86	0.86	0.51	0.28	0.90	0.08	0.09	0.99	0.59	.022	0.75	0.56	0.11	0.43
0.39	0.69	1.88	0.77	0.58	0.27	0.83	0.06	0.12	0.95	0.83	0.26	0.67	0.64	0.08	0.40
0.89	1.65	2.96	1.32	0.83	0.48	1.30	0.08	0.14	1.42	1.70	0.45	1.72	1.15	0.26	0.50
0.88	2.59	2.69	1.86	1.18	0.98	1.17	0.05	0.21	2.13	1.53	0.37	1.17	1.648	0.42	0.68
0.85	2.37	2.77	2.23	0.91	0.72	0.81	0.03	0.17	1.39	1.07	0.20	1.11	1.10	0.28	0.48
1.03	3.03	2.59	1.76	1.07	0.56	0.79	0.12	0.04	1.34	1.37	0.21	0.69	1.09	0.19	0.22
0.74	3.14	2.74	1.83	0.90	0.94	1.15	0.11	0.14	1.93	1.45	0.35	1.07	0.78	0.21	0.30
1.02	2.66	3.53	1.73	0.74	0.45	1.02	0.10	0.14	1.88	1.22	0.30	1.79	0.71	0.18	0.59
0.86	1.56	3.43	1.42	0.74	0.33	1.10	0.12	0.12	1.98	1.30	0.44	1.96	0.69	0.30	0.67
0.71	2.27	3.04	1.67	0.85	0.45	1.31	0.13	0.19	1.84	1.34	0.60	1.76	1.08	0.21	0.35
0.90	2.35	3.59	2.23	0.86	0.53	1.08	0.13	0.14	1.84	2.07	0.34	1.46	1.33	0.34	0.44
0.97	2.56	3.00	1.54	0.80	0.45	0.96	0.06	0.18	1.64	1.12	0.32	1.45	0.70	0.19	0.56
0.81	2.27	3.59	1.86	0.83	0.29	1.27	0.15	0.14	2.42	1.00	0.55	2.48	0.84	0.38	0.61
0.59	3.25	3.65	2.19	0.68	0.69	0.82	0.04	0.05	3.12	1.41	0.40	1.09	1.23	0.18	0.12
0.45	1.75	2.53	1.22	0.69	0.29	0.92	0.07	0.11	1.31	1.09	0.42	1.42	1.13	0.15	0.27
0.91	2.23	2.96	1.79	0.94	0.50	1.18	0.07	0.11	1.81	1.27	0.35	1.93	1.22	0.32	0.62
0.70	1.61	2.26	1.77	0.59	0.35	0.98	0.12	0.11	1.04	0.80	0.24	1.32	0.64	0.19	0.75

图2 证据合成部分结果

Fig.2 Results of evidence combination

参考文献:

[1]Glenn Shafer. A Mathematical Theory of Evidence[M]. Princeton: Princeton University Press, 1976.

[2]段新生. 证据理论与决策、人工智能[M]. 北京: 中国人民大学出版社, 1993. 115—125.
Duan Xinsheng. Evidence Theory, Decision and Artificial Intelligence[M]. Beijing: China Renmin University Press, 1993. 115—125. (in Chinese)

[3]Dempster A P. Upper and lower probabilities induced by a multivalued mapping[J]. Annals of Mathematical Statistics, 1967, 38: 325—339.

[4]Dempster A P. A generalization of Bayesian inference[J]. Journal of the Royal Statistical Society(Series B), 1968, 30: 205—247.

[5]Voorbraak F. On the justification of Dempster's rule of combination[J]. Artificial Intelligence, 1991, 48: 171—197.

[6]肖人彬, 王 雪, 费 奇等. 相关证据合成方法的研究[J]. 模式识别与人工智能, 1993, 6(3): 227—234.
Xiao Renbin, Wang Hue, Fei Qi, *et al*. Research on combination method for dependent evidences[J]. Pattern Recognition and Artificial Intelligence, 1993, 6(3): 227—234. (in Chinese)

[7]孙怀江, 杨静宇. 一种相关证据合成方法[J]. 计算机学报, 1999, 22(9): 1004—1007.
Sun Huaijiang, Yang Jingyu. A combination method for dependendent evidences[J]. Chinese Journal of Computers, 1999, 22(9): 1004—1007. (in Chinese)

[8]杜文吉，陈彦辉，谢维信．加权 Dempster 证据组合算法[J]．西安电子科技大学学报，1999，26(5)：549—551．
Du Wenji, Chen Yanhui, Xie Weixin. A weighted Dempster's rule of combination[J]. Journal of Xidian University, 1999, 26(5): 549—551. (in Chinese)
[9]Yager RR. On the Dempster-Shafer framework and new combination rules[J]. Information Sciences, 1987, 41: 93—137.
[10]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 Xiuqing, Gu Weikang. A new combination rules of evidence theory[J]. Acta Electronic Sinica, 2000, 8: 117—119. (in Chinese)
[11]杨善林，朱卫东，任明仑．基于可变参数优化的相关证据合成方法研究[J]．管理科学学报，2003，6(5)：12—16．
Yang Shanlin, Zhu Weidong, Ren Minglun. Combination theory and method for interrelated evidences based optimal adjustmetn coefficient[J]. Journal of Management Sciences in China, 2003, 6(5): 12—16. (in Chinese)
[12]Pawlak Z. Rough Sets—Theoretical Aspects of Reasoning about Data[M]. Dordrecht: Kluwer Academic Publishers, 1991. 1—30.
[13]Skowron A, Grzymala-Busse J W. From rough set theory to evidence theory[J]. In: Advances in the Dempster-Shafer Theory of Evidence. Yeager RR, Fedrizzi M, Kcprzyk J. New York: John Wiley and Sons, 1994. 193—235.
[14]Skowron A. The relationship between rough sets theory and evidence theory[J]. Bull. Polish. Acad. Sci. Math, 1989, 37: 87—90.
[15]张文修，吴伟志．粗糙集理论介绍和研究综述[J]．模糊系统与数学，2000，14(4)：1—12．
Zhang Wenxiu, Wu Weizhi. An introduction and a survey for the studies of rough set theory[J]. Fuzzy Systems and Mathematics, 2000, 14(4): 1—12. (in Chinese)
[16]刘业政．基于粗糙集数据分析的智能决策支持系统研究[D]．合肥：合肥工业大学，2003．23—24．
Liu Yezheng. Research on Rough Set Data Analysis Based Intelligent Decision Support Systems[D]. Hefei: Hefei University of Technology, 2003. 23—24. (in Chinese)
[17]王伟强，顾国昌．利用 Shafer 证据理论进行决策的改进算法[J]．哈尔滨工程大学学报，1997，18(1)：70—74．
Wang Weiqiang, Gu Guochang. Improved policy decision algorithm based on shafer evidence theory[J]. Journal of Harbin Engineering University, 1997, 18(1): 70—74. (in Chinese)

Research on rough sets-based evidence acquirement and combination of DST

YANG Shan-lin, LIU Ye-zheng, LI Ya-fei
Management School of Hefei University of Technology, Hefei 230009, China

Abstract: As a powerful tool in dealing with uncertainty questions, the evidence using in the evidence theory is given by experts. The expert's knowledge is limited, subjective and sometimes difficult to obtain. To solve these problems, this paper proposes a new method of knowledge acquirement and presents a method of evidence combination and application. Firstly, we research the problem of how to partition the huge decision table. By analysing of the dependency degree among the attributes based on the Rough Sets (RS), condition attributes are classified and the original decision table turns into several small tables, which are independent from one another. Secondly, with the classification method of RS, we analyses the small table, and get the basic probability assignment for evidence theory. Lastly, the problem of how to combine and apply the evidence is discussed and a solution is given as well.
Key words: evidence theory; rough sets; degree of dependency; basic probability assignment; belief function