DOI:10.13374/j.issn1001-053x.2012.05.020

第 34 卷 第 5 期 北 京 科 技 大 学 学 报 Vol. 34 No. 5
2012 年 5 月 Journal of University of Science and Technology Beijing May 2012

基于加权阈值容差关系的不完备信息系统粗糙集模型

武 森[✉] 蒲 立 程 锴 高学东

北京科技大学经济管理学院，北京 100083

✉ 通信作者，E-mail: wusen@ manage. ustb. edu. cn

摘 要 针对已有的不完备信息系统粗糙集扩充模型没有考虑属性的权重的缺点，提出基于加权阈值容差关系的粗糙集扩充模型．该模型根据给出的不完备信息系统信息量计算权重，不需引入系统外知识，权重确定比较客观；同时引入阈值来调整加权阈值容差类判定的严格程度，兼顾了人的主观要求，并能预先排除因自身不满足阈值条件而不可能与任何其他对象划为同一加权阈值容差类的对象，但不影响类的完整性．实例对比分析表明，与其他模型相比，基于加权阈值容差关系的粗糙集扩充模型符合不完备信息系统应用客观实际，具有更强的适应能力．

关键词 粗糙集理论；信息系统；容差分析；加权

分类号 TP 181

Rough set model of incomplete information systems based on the weighted threshold tolerance relation

WU Sen[✉], *PU Li*, *CHENG Kai*, *GAO Xue-dong*

School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China

✉ Corresponding author, E-mail: wusen@ manage. ustb. edu. cn

ABSTRACT In order to overcome the shortage that attribute weights are not taken into consideration in the existing extensions of rough sets under incomplete information systems, a new rough set model is proposed based on the weighted threshold tolerance relation. The new model calculates the weights according to the information quantity of the incomplete information system without outside knowledge, so the weights are objective. Moreover, the model introduces a threshold to adjust the strictness degree of the weighted threshold tolerance class, which not only combines subjective requirements into consideration, but also excludes objects in advance that do not reach the threshold and can not be in the same weighted threshold tolerance class with other objects. This exclusion will not influence the completeness of the classes. Contrastive analysis of an example shows that the proposed extension of a rough set based on the weighted threshold tolerance relation accords with the fact of an incomplete information system and is more applicable compared with other models.

KEY WORDS rough set theory; information systems; tolerance analysis; weighing

粗糙集[1]理论作为一种处理模糊和不确定性知识的新型数学方法，在工业控制与管理[2]、决策科学[3]、医学及生物科学[4]、模式识别[5]、航空航天及军事管理[6]等领域有许多成功的应用．它的基本思想是通过关系数据库分类归纳形成概念和规则，通过等价关系的分类以及分类对于目标的近似实现知识发现[7]．在实际数据挖掘应用中，面对的通常是存在缺失数据的不完备信息系统．由于经典的粗糙集理论中的等价关系不能处理数据中的缺失信息，一般需要采用数据补齐的方法将不完备信息系统转化为完备信息系统或在不完备信息系统下放宽等价关系来扩充粗糙集模型．考虑到补齐数据不一定能真实反映原系统的信息[8]，所以经常采用直接扩充粗糙集模型的方法来处理不完备信息系统．

经典的不完备信息系统扩充模型容差关系[9]、非对称相似关系[10-11]和量化容差关系[10-11]极大地

收稿日期：2011-06-05

基金项目：国家自然科学基金资助项目(70771007)；中央高校基本科研业务费专项(FRF-TP-10-006B)

拓展了粗糙集理论的应用范围. 容差关系认为未知值与任何属性值都相等,可能导致实际没有相同属性值的对象被误判在同一类中; 非对称相似关系因其非对称性,可能使直观上就可以判断相象的对象因不满足相似关系而被判在不同类中; 而量化容差关系需要知道属性值的概率分布情况,应用有一定的困难. 针对上述情况,限制容差关系[12]、联系度容差关系[13]、改进相似关系[14]、限制非对称相似关系[15]和 τ 限制容差关系[16]被相继提出,从不同角度进一步改进和推广了不完备信息系统的粗糙集模型. 但是,这些模型仍然存在着一些局限,并且在上述所有不完备信息系统粗糙集扩充模型中都没有考虑属性权重. 在实际问题中各属性的重要度往往不同,不考虑属性权重是不符合实际情况的.

本文在详细分析限制容差关系、联系度容差关系和 τ 限制容差关系优势与不足的基础上,提出了基于加权阈值容差关系的粗糙集扩充模型. 该模型发扬了上述模型的优点,并克服了这些模型的局限,是对这些模型的进一步推广. 更重要的是加权阈值容差关系在不需引入系统外知识的情况下考虑了不完备信息系统的属性权重,更符合客观实际.

1　几种主要粗糙集扩充模型

1.1　限制容差关系模型

在深入剖析经典模型容差关系、非对称相似关系和量化容差关系的基础上,文献[12]提出的限制容差关系成为后来不完备信息系统扩充模型研究的重要基础之一.

对于一个不完备信息系统 $S=\langle U,A,V,f\rangle$, $U=\{x_1,x_2,\cdots,x_n\}$ 是论域, A 是属性集, $V=\bigcup_{a\in A}V_a$ 是属性值集, V_a 是属性 a 的值域, "*"表示属性值未知, f 是信息函数,即对 $\forall a\in A$, $\forall x\in U$,有 $a(x)\in V_a$. 对 $B\subseteq A$,记 $P_B(x)=\{b\mid b\in B\wedge b(x)\neq *\}$,则限制容差关系[12]定义为

$$\forall_{x,y\in U\times U}(L_B(x,y)\Leftrightarrow\forall_{b\in B}(b(x)=b(y)=*)\vee((P_B(x)\cap P_B(y)\neq\varnothing)\wedge\forall_{b\in B}((b(x)\neq *)\wedge(b(y)\neq *)\rightarrow(b(x)=b(y))))).$$

限制容差关系 $L_B(x,y)$ 包括两种情形: (1) x 和 y 在属性子集 B 上全为"*"; (2) x 和 y 在 B 上有同不为"*"的属性且在所有这样的属性上取值相同. 限制容差关系具有自反性和对称性,但不具有传递性,在信息系统不包含全为"*"的对象时正好介于容差关系太宽松和非对称相似关系太严格两个极端之间,优于经典模型. 限制容差关系的第一种情形可能将全为"*"而实际没有相同属性值的对象误判在同一类中; 对第二种情形,两个对象只要有一个属性取值明确相同而在其余属性上都有"*"时就被判定在同一类,条件仍然比较宽松,尤其在属性数目很多时更为明显.

1.2　联系度容差关系模型

在限制容差关系的基础上,联系度容差关系[13]引入阈值 α_1 和 $\alpha_2(0\leqslant\alpha_2\leqslant\alpha_1\leqslant 1)$. 将不为"*"的属性数目在 B 中所占比例小于阈值 α_1 的对象排除,针对剩余对象构成的新论域 U' 根据下式将取值明确相同属性比例不小于 α_2 且无明确不相同属性值的两个对象判定为同一类.

$$P_B^{\alpha_2}(x)=\{y\in U'\mid u(x,y)=a+b_i,a+b=1,a\geqslant\alpha_2\}.$$

其中, a 和 b 分别为 x 和 y 在属性子集 B 上取值明确相同属性比例和取值不明确属性比例.

联系度容差关系通过引入阈值结合人的主观要求调整类判定的严格程度,改善了限制容差关系条件仍然比较宽松的情况而保持了自反性和对称性,但在联系度容差关系中由于阈值 α_1 的限制,不为"*"的属性数目在 B 中所占比例在 $[\alpha_2,\alpha_1)$ 区间的对象在被预先排除之列,从而可能使满足阈值 α_2 要求的联系度容差类不完整.

1.3　τ 限制容差关系模型

τ 限制容差关系 $\mathrm{ILR}(B,\tau)$[16]引入阈值 $\tau(0\leqslant\tau\leqslant 1)$ 来调整类判定的严格程度,两个对象被判定为同属一类具体分为两种情形: (1) B 中所有属性值都对应一致,即取值明确相同或同为"*"; (2) B 中取值明确相同属性比例不小于 τ 且无明确不相同属性值.

$$\mathrm{ILR}(B,\tau)=\{(x,y)\mid\in U\times U\mid\forall_{b\in B}(b(x)=b(y))\vee(\mid P_B(x)\cap P_B(y)\mid\geqslant\mid A\mid\times\tau)\wedge\forall_{b\in B}((b(x)\neq *)\wedge(b(y)\neq *)\rightarrow(b(x)=b(y)))\}.$$

τ 限制容差关系对限制容差关系进行了改进和推广,保持了自反性和对称性,但由于其第一种情形将所有属性值都对应一致的对象归为一类,所以无论 τ 如何取值,它仍然可能将全部属性值为"*"而实际没有相同属性值的对象误判在同一类中,条件仍然比较宽松.

针对上述模型的优势及局限,本文提出基于加权阈值容差关系的粗糙集扩充模型,吸收了这些模型的优点,摈弃了不足,并在不必引入不完备信息系统外知识的情况下考虑了属性的权重,更符合实际应用情况.

2　基于加权阈值容差关系的粗糙集模型

2.1　不完备信息系统的属性权重

加权阈值容差关系模型基于不完备信息系统信息量计算权重，不需引入系统外知识，权重确定比较客观.

定义 1　对不完备信息系统 $S=\langle U,A,V,f\rangle$，$B\subseteq A$，$[x_i]_B^T$ 为 x_i 的容差类，定义信息量为

$$I(B)=\frac{n}{n-1}-\frac{1}{n(n-1)}\sum_{i=1}^{n}|[x_i]_B^T|.$$

性质 1　$0\leqslant I(B)\leqslant 1$，$\max I(B)=1\Leftrightarrow[x_i]_B^T=\{x_i\}$，$\min I(B)=0\Leftrightarrow[x_i]_B^T=U$.

该性质表明：属性子集 B 的信息量在 $[0,1]$ 区间. 在 B 能区分出每一个对象 x_i，$i=1,2,\cdots,n$ 时信息量达到最大，为 1；在 B 不能区分出 U 中任何对象时信息量达到最小，为 0.

性质 2　对于 $B_1\subset B_2\subseteq A$，有 $I(B_1)\leqslant I(B_2)$.

该性质表明：随着属性子集 B 中属性的增加，信息量单调递增.

定义 2　对不完备信息系统 $S=\langle U,A,V,f\rangle$，属性 $b\in B\subseteq A$ 在 B 中的重要度定义如下：

$$\mathrm{Sig}_B(b)=I(B)-I(B\backslash\{b\}).$$

性质 3　$0\leqslant\mathrm{Sig}_B(b)\leqslant 1$.

该性质表明：属性 $b\in B\subseteq A$ 在 B 中的重要度在 $[0,1]$ 区间. 这是因为由性质 1 可知 $0\leqslant I(B)\leqslant 1$ 及 $0\leqslant I(B\backslash\{b\})\leqslant 1$，由性质 2 可知 $I(B\backslash\{b\})\leqslant I(B)$，所以 $0\leqslant I(B)-I(B\backslash\{b\})\leqslant 1$，即 $0\leqslant\mathrm{Sig}_B(b)\leqslant 1$.

b 在 B 中的权重通过重要度归一化后确定，即

$$w(b)=\frac{\mathrm{Sig}_B(b)}{\sum\limits_{b_i\in B}\mathrm{Sig}_B(b_i)},$$

$$0\leqslant w(b)\leqslant 1,\ \sum_{b\in B}w(b)=1.$$

下面给出不完备信息系统属性权重计算实例. 为了研究比较方便，所用的不完备信息系统为文献[10–13,16]中分析容差关系、量化容差关系、限制容差关系、联系度容差关系和 τ 限制容差关系共同采用的实例，如表 1 所示. 其中，$U=\{a_1,a_2,\cdots,a_{12}\}$ 为论域，为条件属性集，$\{d\}$ 是决策属性集，$V_C=\{0,1,2,3\}$ 为条件属性值集，$V_{\{d\}}=\{\Phi,\Psi\}$ 为决策属性值集.

$$I(C)=0.712,$$

$$\mathrm{Sig}_C(c_1)=I(C)-I(C\backslash\{c_1\})=$$
$$0.712-0.606=0.106,$$

$$\mathrm{Sig}_C(c_2)=I(C)-I(C\backslash\{c_2\})=$$
$$0.712-0.591=0.121,$$

$$\mathrm{Sig}_C(c_3)=I(C)-I(C\backslash\{c_3\})=$$
$$0.712-0.712=0,$$

$$\mathrm{Sig}_C(c_4)=I(C)-I(C\backslash\{c_4\})=$$
$$0.712-0.561=0.151,$$

$$w(c_1)=0.28,$$
$$w(c_2)=0.32,$$
$$w(c_3)=0,$$
$$w(c_4)=0.40.$$

表 1　不完备信息表

Table 1　Table of incomplete information

U	c_1	c_2	c_3	c_4	d
a_1	3	2	1	0	Φ
a_2	2	3	2	0	Φ
a_3	2	3	2	0	Ψ
a_4	*	2	*	1	Φ
a_5	*	2	*	1	Ψ
a_6	2	3	2	1	Ψ
a_7	3	*	*	3	Φ
a_8	*	0	0	*	Ψ
a_9	3	2	1	3	Ψ
a_{10}	1	*	*	*	Φ
a_{11}	*	2	*	*	Ψ
a_{12}	3	2	1	*	Φ

2.2　加权阈值容差关系

定义 3　对不完备信息系统 $S=\langle U,A,V,f\rangle$，$B\subseteq A$，$0\leqslant w(b)\leqslant 1$ 为属性 b 在 B 中的权重，加权阈值容差关系为

$$\mathrm{WT}(\omega)=\{(x,y)\mid x\in U\wedge y\in U\wedge\forall_{b\in B}(b(x)=*\vee b(y)=*\vee b(x)=b(y))\wedge\sum_{b\in B'}w(b)\geqslant\omega\}.$$

其中，

$$B'=\{b\in B\mid(b(x)\neq *)\wedge(b(y)\neq *)\wedge(b(x)=b(y))\},$$ $0\leqslant\omega\leqslant 1$ 为阈值.

此时，记 $[x]_B^{\mathrm{WT}(\omega)}=\{y\in U\mid(x,y)\in\mathrm{WT}(\omega)\}$ 为对象 x 的加权阈值容差类.

由定义可知，对象 x 与 y 只有在属性子集 B 中取值明确相同的属性权重和不小于 ω 且没有明确不相同的属性值时，才能被判定属于同一类. 在实际应用中，可以根据具体问题的实际情况和人的主观要求来调整 ω 的取值，以得到更好的应用效果.

显然，加权阈值容差关系是自反的、对称的，但不一定是传递的. 并且，加权阈值容差关系一般不构成 U 的划分，而构成 U 的覆盖.

定义 4　对不完备信息系统 $S=\langle U,A,V,f\rangle$，

$B \subseteq A$，对象集 X 关于属性子集 B 的加权阈值容差关系的上近似 $B^{WT(\omega)}(X)$ 和下近似 $B_{WT(\omega)}(X)$ 分别为

$$B^{WT(\omega)}(X) = \{x \mid [x]_B^{WT(\omega)} \cap X \neq \varnothing, x \in U\},$$

$$B_{WT(\omega)}(X) = \{x \mid [x]_B^{WT(\omega)} \subseteq X, x \in U\}.$$

由上述定义可知 $B^{WT(\omega)}(X) = \bigcup_{x \in X} [x]_B^{WT(\omega)}$.

关于加权阈值容差关系的下述两个定理成立，证明略.

定理 1　如果 $|B| = 1$ (记为 $B = \{b\}$) 且 $\omega \neq 0$ 时有 $(x, y) \in WT(\omega)$，则

$$(b(x) = b(y)) \wedge (b(x) \neq *) \wedge (b(y) \neq *).$$

此时若存在 $b(z) = *$，则 $[z]_B^{WT(\omega)} = \{z\}$.

该定理表明: 在只有一个属性且阈值不为 0 的情况下，如果两个对象同属一个加权阈值容差类，则属性取值一定明确相等; 如果一个对象属性取值为 "*"，则一定独自属于一个加权阈值容差类.

定理 2　记 $P_B(x) = \{b \mid b \in B \wedge b(x) \neq *\}$，$X = \left\{x \mid \sum_{b \in P_B(x)} w(b) < \omega\right\}$，则对 $\forall x \in X$，有 $[x]_B^{WT(\omega)} = \{x\}$.

该定理表明: 当对象 x 在属性子集 B 中不为 "*" 的属性权重和小于 ω 时，不可能再与其他任何对象同属一个加权阈值容差类，只能被判定独自属于一个加权阈值容差类中，在判定其他对象的加权阈值容差类时可以从论域中将其排除.

2.3　加权阈值容差关系模型与其他模型的联系

特殊地，当 $\forall_{b \in B} w(b) = \frac{1}{|B|}$ 时，加权阈值容差关系即退化为联系度容差关系和 τ 限制容差关系，但不会出现联系度容差关系中取值明确相同的属性比例在 $[\alpha_2, \alpha_1)$ 区间的对象被预先排除的情况，也不会出现 τ 限制容差关系中条件比较宽松的第一种情形. 更进一步地，当 $\forall_{b \in B} w(b) = \frac{1}{|B|}$ 且 $0 < \omega \leqslant \frac{1}{n}$ 时，加权阈值容差关系即退化为限制容差关系的第二种情形，而排除了全部属性为 "*" 的不同对象被判定为同属一类的第一种情形; 当 $\forall_{b \in B} w(b) = \frac{1}{|B|}$ 且 $\omega = 0$ 时，加权阈值容差关系即退化为容差关系. 因此，加权阈值容差关系是联系度容差关系、τ 限制容差关系、限制容差关系和容差关系的推广; 联系度容差关系、τ 限制容差关系、限制容差关系和容差关系都是加权阈值容差关系的特殊情形.

3　实例比较分析

继续采用计算属性权重的实例表 1 来分析加权阈值容差关系. 由于限制容差关系是对容差关系和非对称相似关系的改进和推广，联系度容差关系和 τ 限制容差关系是对限制容差关系的改进和推广，在文献 [12-13,16] 中已有深入详细的对比分析，所以在此仅将加权阈值容差关系与联系度容差关系和 τ 限制容差关系进行比较.

在联系度容差关系下，根据 $0 \leqslant \alpha_2 \leqslant \alpha_1 \leqslant 1$ 的要求取阈值 $\alpha_1 = 0.51$，$\alpha_2 = 0.3$，对于属性子集 C，a_4，a_5，a_7，a_8，a_{10} 和 a_{11} 由于不为 "*" 的属性数目在 C 中的比例小于 0.51 被从论域 U 中删除，剩余对象根据两对象取值明确相同属性比例不小于 0.3 且无明确不相同属性值被判定为一类，得到如下联系度容差类:

$$P_C^{0.3}(a_1) = \{a_1, a_{12}\},$$
$$P_C^{0.3}(a_2) = P_C^{0.3}(a_3) = \{a_2, a_3\},$$
$$P_C^{0.3}(a_6) = \{a_6\},$$
$$P_C^{0.3}(a_9) = \{a_9, a_{12}\},$$
$$P_C^{0.3}(a_{12}) = \{a_1, a_9, a_{12}\}.$$

上述实例中，联系度容差关系没有考虑属性权重，而且由于阈值的限制，不为 "*" 的属性数目在 C 中所占比例在 $[0.3, 0.51)$ 区间的对象被预先删除，从而使满足不小于阈值 $\alpha_2 = 0.3$ 要求的联系度容差类不完整. 例如: 根据不小于阈值 $\alpha_2 = 0.3$ 的要求，a_7 应该被包含在 a_9 的联系度容差类中，但是由于 a_7 根据小于阈值 $\alpha_1 = 0.51$ 的要求已被删除，因而 a_9 的联系度容差类不完整.

在 τ 限制容差关系下，当阈值 $\tau = 0.3$ 时，对于属性子集 C，所有属性值都对应一致 (明确相同或同为 "*") 的两对象同属一类，取值明确相同的属性比例不小于 0.3 且无明确不相同属性值的两对象也同属一类.

$$IL_C^{0.3}(a_1) = \{a_1, a_{12}\},$$
$$IL_C^{0.3}(a_2) = IL_C^{0.3}(a_3) = \{a_2, a_3\},$$
$$IL_C^{0.3}(a_4) = IL_C^{0.3}(a_5) = \{a_4, a_5\},$$
$$IL_C^{0.3}(a_6) = \{a_6\},$$
$$IL_C^{0.3}(a_7) = \{a_7, a_9\},$$
$$IL_C^{0.3}(a_8) = \{a_8\},$$
$$IL_C^{0.3}(a_9) = \{a_7, a_9, a_{12}\},$$
$$IL_C^{0.3}(a_{10}) = \{a_{10}\},$$
$$IL_C^{0.3}(a_{11}) = \{a_{11}\},$$
$$IL_C^{0.3}(a_{12}) = \{a_1, a_9, a_{12}\}.$$

从上述实例可知: τ 限制容差关系也没有考虑属性权重. 在该例中，属性子集 C 中全部属性为 "*" 的对象没有出现. 如果有，则无论 τ 如何取值，

都有可能将全部属性值为“*”而实际没有相同属性值的对象误判在同一类中.

在加权阈值容差关系下,取阈值 $\omega=0.3$ 时,对象 a_{10} 不为“*”的属性权重和为 $\sum_{b\in P_C(a_{10})} w(b)=w(c_1)=0.28<\omega$,不可能再与其他任何对象同属于一个类中,独自属于一个加权阈值容差类,即 $[a_{10}]_C^{\mathrm{WT}(0.3)}=\{a_{10}\}$,$a_{10}$ 被从论域 U 中排除. 剩余各对象的加权阈值容差类分别为

$$[a_1]_C^{\mathrm{WT}(0.3)}=\{a_1,a_{11},a_{12}\},$$
$$[a_2]_C^{\mathrm{WT}(0.3)}=[a_3]_C^{\mathrm{WT}(0.3)}=\{a_2,a_3\},$$
$$[a_4]_C^{\mathrm{WT}(0.3)}=[a_5]_C^{\mathrm{WT}(0.3)}=\{a_4,a_5,a_{11},a_{12}\},$$
$$[a_6]_C^{\mathrm{WT}(0.3)}=\{a_6\},$$
$$[a_7]_C^{\mathrm{WT}(0.3)}=\{a_7,a_9\},$$
$$[a_8]_C^{\mathrm{WT}(0.3)}=\{a_8\},$$
$$[a_9]_C^{\mathrm{WT}(0.3)}=\{a_7,a_9,a_{11},a_{12}\},$$
$$[a_{11}]_C^{\mathrm{WT}(0.3)}=\{a_1,a_4,a_5,a_9,a_{11},a_{12}\},$$
$$[a_{12}]_C^{\mathrm{WT}(0.3)}=\{a_1,a_4,a_5,a_9,a_{11},a_{12}\}.$$

与联系度容差关系和 τ 限制容差关系相比,加权阈值容差关系不仅考虑了属性的权重,而且可以将不满足加权阈值条件的对象预先排除,但不影响类的完整性,类的判定更为合理. 下面以几个对象为例说明.

(1) a_{10} 只有属性 c_1 具有明确值,不为“*”的属性权重和即为属性 c_1 的权重 0.28,小于阈值 $\omega=0.3$,根据定理 2 可知,应该被预先排除,独自属于一个加权阈值容差类.

(2) a_{11} 只有属性 c_2 具有明确值,不为“*”的属性权重和即为属性 c_2 的权重 0.32,大于阈值 $\omega=0.3$,根据定理 2 不应该被预先排除,最终结果是其被判定为分别属于 a_1,a_4,a_5,a_9,a_{11} 和 a_{12} 的加权阈值容差类中.

(3) a_9 的加权阈值容差类为 $\{a_7,a_9,a_{11},a_{12}\}$,与其 τ 限制容差类 $\{a_7,a_9,a_{12}\}$ 相比多 a_{11},这是因为虽然 a_{11} 和 a_9 只有属性 c_2 取值明确相同,但是由于 c_2 的权重较大,$w(c_2)=0.32$,大于阈值 $\omega=0.3$,所以 a_{11} 被判定在 a_9 的加权阈值容差类中. a_9 的加权阈值容差类与其联系度容差类 $\{a_9,a_{12}\}$ 相比多 a_7 和 a_{11},这不仅是由于属性权重的影响,还因为 a_7 和 a_{11} 由于阈值 $\alpha_1=0.51$ 的限制在联系度容差类中已经被预先删除了,联系度容差类 $\{a_9,a_{12}\}$ 是不完整的. 显而易见,与联系度容差类和 τ 限制容差类相比,a_9 的加权阈值容差类更完整,更合理.

4 结论

粗糙集模型在不完备信息系统的扩充,对于粗糙集理论进一步应用于实际有着非常重要的意义. 本文在深入分析几种主要粗糙集扩充模型的基础上,提出了基于加权阈值容差关系的粗糙集扩充模型,在同时考虑属性权重和阈值的情况下确定加权阈值容差类. 该模型是对容差关系、限制容差关系、联系度容差关系和 τ 限制容差关系的改进和推广,适应性更强. 属性权重基于不完备信息系统的信息量确定,不需引入系统外知识,比较客观,而阈值 ω 的引入又兼顾了人的主观要求,可以根据具体问题来调整 ω 的取值,以取得更好的应用效果. 该模型还能预先排除因自身不满足阈值要求而不可能与任何其他对象判定为同一类的对象,但不影响加权阈值容差类的完整性,且能保证加权阈值容差类的自反性和对称性. 实例对比分析表明,与其他几种主要模型相比,本文提出的粗糙集扩充模型具有更强的适应能力.

参 考 文 献

[1] Pawlak Z. Rough sets. *Int J Comput Inf Sci*, 1982, 11(5): 341

[2] Pawlak Z. AI and intelligent industrial applications: the rough set perspective. *Cybern Syst*, 2000, 31(3): 227

[3] Xie G, Zhang J L. Group decision-making for risk avoidance in software project bidding based on VPRS. *Chin J Manage Sci*, 2006, 14(2): 71
(谢刚,张金隆. 基于 VPRS 的软件项目投标风险规避群决策研究. 中国管理科学,2006,14(2): 71)

[4] Hirano S, Tsumoto S. Segmentation of medical images based on approximations in rough set theory // *Proceedings of the Third International Conference on Rough Sets and Current Trends in Computing*. Malvern, 2002: 554

[5] Song X Y, Liu F, Sun H L. Autonomous threshold selection based on rough set theory in clustering algorithm. *Syst Eng Electron*, 2010, 32(1): 192
(宋晓宇,刘锋,孙焕良. 基于粗糙集的聚类算法中阈值自动选取. 系统工程与电子技术,2010,32(1): 192)

[6] Wojcik Z M. Detecting spot s for NASA space programs using rough sets // *Proceedings of the Second International Conference on Rough Sets and Current Trends in Computing*. Banff, 2000: 531

[7] Wang G Y, Yao Y Y, Yu H. A survey on rough set theory and applications. *Chin J Comput*, 2009, 32(7): 1229
(王国胤,姚一豫,于洪. 粗糙集理论与应用研究综述. 计算机学报,2009,32(7): 1229)

[8] Wang G Y, Guan L H, Hu F. Rough set extensions in incomplete information systems. *Front Electr Electron Eng China*, 2008, 3(4): 399

[9] Kryszkiewicz M. Rough set approach to incomplete information system. *Inf Sci*, 1998, 112(1-4): 39

[10] Stefanowski J, Tsoukiás A. On the extension of rough sets under incomplete information // *Proceedings of the Seventh International Workshop on New Directions in Rough Sets, Data Mining, and*

Granular-Soft Computing. Yamaguchi, 1999: 73

[11] Stefanowski J, Tsoukiàs A. Incomplete information tables and rough classification. *Comput Intell*, 2001, 17(3): 545

[12] Wang G Y. Extension of rough set under incomplete information systems. *J Comput Res Dev*, 2002, 39(10): 1238
(王国胤. Rough 集理论在不完备信息系统中的扩充. 计算机研究与发展, 2002, 39(10): 1238)

[13] Huang B, Zhou X Z. Extension of rough set model based on connection degree under in complete information systems. *Syst Eng Theory Pract*, 2004, 24(1): 88
(黄兵, 周献中. 不完备信息系统中基于联系度的粗集模型拓展. 系统工程理论与实践, 2004, 24(1): 88)

[14] Rady E A, Abd El-Monsef M M E, Abd El-Latif W A. A modified rough set approach to incomplete information systems. *J Appl Math Decis Sci*, 2007, 10: 1155

[15] Qu B B, Lu Y S. Rough set model based on limited non-symmetric similarity relation. *J Chin Comput Syst*, 2007, 28(6): 1084
(瞿彬彬, 卢炎生. 基于限制非对称相似关系的粗糙集模型. 小型微型计算机系统, 2007, 28(6): 1084)

[16] Liang M L, Liang J R, Li T Z, et al. Extension of rough set under incomplete information system based on τ limited tolerance relation. *Comput Eng Appl*, 2007, 43(31): 53
(梁美莲, 梁家荣, 李天志, 等. 基于 τ 限制容差关系的不完备信息粗糙集模型. 计算机工程与应用, 2007, 43(31): 53)