文章编号：1000-6788(2008)05-0116-06

不完备信息系统中的可变精度分类粗糙集模型

杨习贝[1]，杨静宇[1]，於东军[1]，吴 陈[2]
(1. 南京理工大学 计算机科学与技术学院，南京 210094；2. 德瑞索大学 信息科学与技术学院，美国 费城 PA19104)

摘要： 在不完备信息系统中，容差关系过于宽松，而相似关系则过于严格. 根据这样的解释，提出了一种新的基于可变精度分类的拓展粗糙集模型，其中的分类方式相比较于容差关系和相似关系显得更为灵活，是这两者的一种广义化表现形式，且可变精度分类也是限制容差关系的一种改进形式. 在此基础上，将这种拓展粗集模型与基于容差关系和相似关系的拓展粗集模型进行了对比分析. 最后在不完备信息系统中使用新的拓展粗集模型讨论了确定和可能性规则的直接生成方法，并进行了实例分析以说明其有效性.

关键词： 不完备信息系统；容差关系；相似关系；可变精度分类；粗糙集
中图分类号： TP18 **文献标志码：** A

Rough set model based on variable parameter classification in incomplete information systems

YANG Xi-bei[1], YANG Jing-yu[1], YU Dong-jun[1], WU Chen[2]
(1. School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China; 2. College of Information Science and Technology, Drexel University, Philadelphia, PA 19104, USA)

Abstract: In incomplete information systems, the tolerance relation is too loose while the similarity relation is too strict. According to these explanations, a new kind of rough set model that is based on the variable parameter classification of objects is proposed. The variable parameter classification is a kind of generalized classification form and it is more flexible than tolerance and similarity relations in practical applications. Furthermore, it can be regarded as an improvement of the limited tolerance relation. From what have been discussed above, the new-defined generalized rough set model is compared with expanded rough sets those are on the basis of tolerance and similarity relations. Finally, the immediate method of certain and possible rules' induction is studied by the new expanded rough set model, an illustrative example is used to show its validity.

Key words: incomplete information system; tolerance relation; similarity relation; variable parameter classification; rough sets

1 引言

近年来，随着实际工程应用的需要，使用粗糙集理论[1~3](Rough Set Theory，简称 RST)从不完备信息系统[4~12](Incomplete Information System，简称 IIS)中获取知识已成为一个热点研究问题. 由于数据填充或删除[5]处理 IIS 中的未知属性值(间接处理方法)会影响到原有的数据分布情况，使得挖掘出来的规则具有较大的不确定性[4]，因此将传统 RST 中的不可分辨关系(等价关系)拓展为其他较弱的二元关系，从而使用各种拓展的粗集模型处理 IIS(直接处理方法)正受到越来越多的学者的关注.

一般来说，IIS 中的未知属性值可能具有两种不同的解释. 首先，所有的未知属性值仅仅是被遗漏的，但又是确实存在的[5~8]，根据这样的解释，Kryszkiewicz 构建了满足自反和对称性的容差关系[6]，并研究了 IIS 中的知识约简问题. 另一方面，所有的未知属性值被认为是丢失的，是不允许被比较的[9,10]，据此，Stefanowski 等人构建了非对称相似关系[10]并建立了近似集的概念. 在对容差关系和相似关系做进一步深

收稿日期：2007-01-07
资助项目：国家自然科学基金(60472060，60572034，60632050)；江苏省自然科学基金(BK2006081)
作者简介：杨习贝(1980－)，男，博士研究生，研究方向为智能信息处理、粗糙集理论、人工智能.

入研究后，王国胤教授认为容差关系的要求过于宽松，而相似关系的要求过于严格，因此他提出了限制容差关系[4]，根据限制容差关系所得到的分类结果介于容差关系和相似关系之间．但是对于限制容差关系来说，也存在一定的不足之处，例如它将两个没有任何已知属性值的对象划分到一类中，或者当信息系统中属性的数目很多时，两个对象仅仅具有很少的已知相同属性，亦可以被认为是处于同一类中的．对于上述两种情况来说，对象之间只能说是具有同类的可能性，并且由于未知属性值的大量存在，这种可能性较小．

由于 RST 建立在分类机制的基础上，它把知识看作是一种对现实或抽象的对象进行分类的能力，而分类则是推理、学习与决策中的关键问题[3]，因此为了使得 IIS 中对象的分类结果更加符合客观实际的需要，更具有灵活性，笔者构建了一种新的可变精度分类关系，通过设置一个阈值来衡量两对象之间相似的程度，从而判断这两个对象是否属于同一类．这种可变精度分类关系仅满足自反性，是容差关系和相似关系的一种推广形式，并且可变精度分类关系也是限制容差关系的一种改进形式，因此研究基于这种分类方式的粗糙集模型具有更为广泛的意义．

本文的主要内容安排如下：第二节简要介绍 IIS 的基本概念及已有的几种拓展粗集模型．第三节提出了可变精度分类方法，在此基础上构建了新的拓展粗集模型，并与现有的几种拓展粗集模型进行了对比分析．第四节进行了规则直接生成方法的讨论，并进行了实例分析．第五节总结全文．

2　基本概念

一个 IIS 为一个四元组：$S=\langle U, AT, V, f\rangle$．其中 U 是一个被称为论域的非空有限的对象集合；AT 是一个非空有限的属性集合；对于 $\forall a\in AT$，有 $a: U\to V_a$，其中 V_a 是属性 a 的值域（可包含未知属性值，文中用‘ $*$ ’表示），集合 $V=\bigcup_{a\in AT}V_a$，V 被称为全体属性的值域；定义 f 为信息函数，即对于 $\forall a\in AT$，$\forall x\in U$，有 $f(x,a)\in V_a$．

当 IIS 中的所有未知属性值都是遗漏型时，可构建如下所示的容差关系．

定义 1[6]　令 S 为一 IIS，对于 $\forall A\subseteq AT$，由 A 决定的容差关系记为 $T(A)$ 且

$$T(A)=\{(x,y)\in U^2:\forall a\in A, f(x,a)=f(y,a)\vee f(x,a)=*\vee f(y,a)=*\}. \tag{2.1}$$

在容差关系 $T(A)$ 中，未知属性值‘ $*$ ’被认为是一种任何值都有可能的值，这个解释与这样的值仅仅是被遗漏但又确实存在的解释相对应，此时，个体对象具有潜在的完备信息，而当前只是遗漏了这些值，因此当两对象在所有取值同不为空的属性上有相同值时，就可被认为是同类的．对于 $\forall x\in U$，x 的容差类记为 $T_A(x)$ 且 $T_A(x)=\{y\in U:(x,y)\in T(A)\}$．

定义 2[6]　令 S 为一 IIS，其中 $A\subseteq AT$，对于 $\forall X\subseteq U$，X 基于 $T(A)$ 的下、上近似集定义为

$$\underline{A^{\mathrm{T}}}(X)=\{x\in U:T_A(x)\subseteq X\};\quad \overline{A^{\mathrm{T}}}(X)=\{x\in U:T_A(x)\cap X\neq\emptyset\}. \tag{2.2}$$

当 IIS 中的所有未知属性值都被认为是丢失型时，可构建如下所示的非对称相似关系．

定义 3[10]　令 S 为一 IIS，对于 $\forall A\subseteq AT$，由 A 决定的相似关系记为 $S(A)$ 且

$$S(A)=\{(x,y)\in U^2:\forall a\in A, f(x,a)=*\vee f(x,a)=f(y,a)\}. \tag{2.3}$$

在相似关系 $S(A)$ 中，对象可能被不完全描述的原因不仅可能是由于知识不精确，还可能是由于干脆就不可能用所有的属性来描述它们，因此不认为未知属性值‘ $*$ ’是不确定的，而是当前不存在的，是不允许比较的未知值，从而各对象可能有或多或少的完全描述，这取决于可能采用多少个属性．因此只要两个对象的已知属性值相同，就可以认为一个对象 x 相似于另一个对象 y[4]．根据定义 3 所示的相似关系，Stefanowski[10] 给出了非对称相似于 x 的集合 $S_A(x)=\{y\in U:(y,x)\in S(A)\}$，$x$ 与之非对称相似的集合 $S_A^{-1}(x)=\{y\in U:(x,y)\in S(A)\}$．

定义 4[10]　令 S 为一 IIS，其中 $A\subseteq AT$，对于 $\forall X\subseteq U$，X 基于 $S(A)$ 的下、上近似集定义为

$$\underline{A^{S}}(X)=\{x\in U:S_A^{-1}(x)\subseteq X\};\overline{A^{S}}(X)=\bigcup\{S_A(x):x\in X\}. \tag{2.4}$$

通过对容差关系与非对称相似关系的深入研究，王国胤[4]认为容差关系的要求过于宽松，容易将没有任何已知相同属性值的对象划分到一类中去，而相似关系的要求则过于严格，容易将具有大量已知相同属性值的对象分离开来，为此，他提出了限制容差关系如下所示．

定义 5[4]　令 S 为一 IIS,对于 $\forall A \subseteq AT$,由 A 决定的限制容差关系记为 $L(A)$ 且

$$L(A)=\{(x,y)\in U^2:\forall a\in A, f(x,a)=f(y,a)=*\ \vee\ ((P_A(x)\cap P_A(y)\neq\varnothing)\ \wedge \forall_{a\in A}(f(x,a)\neq *)\wedge(f(y,a)\neq *)\rightarrow(f(x,a)=f(y,a))\}.\tag{2.5}$$

其中 $P_A(x)=\{a\in A: f(x,a)\neq *\}$.

容差关系和相似关系分别是对传统粗集理论中的不可分辨关系扩充所得到的两种极端情况,而限制容差关系所得到的分类结果介于这两者之间.从直观上来看,限制容差关系有两种情形:1) x,y 在 A 上全为空值;2) x,y 在 A 上存在同不为空值的属性,且取值相同.

3　可变精度分类粗集模型

在容差关系中,两个对象在没有任何已知相同属性值的情况下可以被认为是同类的,如 $x=\{*,1,*,2,3\}$, $y=\{1,*,0,*,*\}$,有 $(x,y)\in T(A)$. 在相似关系中,两个具有大量已知相同属性值的对象往往容易被分离开来,如 $x=\{*,1,0,2,3,2,1,1\}$, $y=\{1,*,0,2,3,2,1,1\}$,有 $(x,y)\notin S(A)$ 且 $(y,x)\notin S(A)$.而对于限制容差关系的第一种情况来说,两个不同的对象在所有属性上取值都为空,仅仅说明这两个对象有不可区分的可能性且由于未知属性值的大量存在,这种可能性极小;对于限制容差关系的第二种情况来说,两对象只要有一个属性值取值相同,而在其余属性上取值无法比较时仍视为同类,在属性成百上千的情形下,这种条件仍然过于宽松.为了解决上述问题,我们分析并提出了如下所示的可变精度分类关系.

定义 6　令 S 为一 IIS,对于 $\forall A \subseteq AT$,由 A 决定的可变精度分类关系记为 $V^\alpha(A)$ 且

$$V^\alpha(A)=\left\{(x,y)\in U^2:\forall a\in P_A(x)\cap P_A(y), f(x,a)=f(y,a)\wedge\frac{|P_A(x)\cap P_A(y)|}{|P_A(x)|}\geqslant\alpha\right\}\cup I_U.\tag{3.1}$$

其中 $\alpha\in[0,1]$, $|\cdot|$ 表示集合的基数, I_U 为恒等关系且 $I_U=\{(x,x):x\in U\}$.

在定义 6 中,特别地,若有 $P_A(x)=\varnothing$,则认为 $|P_A(x)\cap P_A(y)|/|P_A(x)|=0$,但恒等关系 I_U 的存在则保证了 x 与其自身仍然满足 $V^\alpha(A)$.容易看出, $V^\alpha(A)$ 仅满足自反性,而对称性和传递性则不一定满足.从直观上来看,可变精度分类关系要求两对象在所有同不为空值的属性上必须取值相同,并且这些同不为空的属性必须达到一定的数量.根据定义 6, x 的自反邻域记为 $V_A^\alpha(x)$ 且 $V_A^\alpha(x)=\{y\in U:(x,y)\in V^\alpha(A)\}$.根据可变精度分类关系 $V^\alpha(A)$,只要合理地设置 α 的取值,就可以解决由容差关系和相似关系分类时所带来的问题.例如,若设 $\alpha>0$,则可以将 $x=\{*,1,*,2,3\}$, $y=\{1,*,0,*,*\}$ 两者分离开来,若设 $\alpha=0.6$, $x=\{*,1,0,2,3,2,1,1\}$, $y=\{1,*,0,2,3,2,1,1\}$ 两者则可以被认为是同类的.

对于限制容差关系 $L(A)$ 来说,当 $\alpha>0$ 时, $V^\alpha(A)$ 排除了 $L(A)$ 的第一种情形;而对于限制容差关系的第二种情形,只要合理地设置阈值 α, $V^\alpha(A)$ 就可以避免两对象因只有极少数相同的属性值就被认为是同一类的情况,例如 $x=\{*,1,*,2,3,*,1,*\}$, $y=\{1,*,0,*,*,*,1,*\}$,限制容差关系认为这两者是属于同类的,然而在所有八个属性上, x 与 y 仅在一个属性上具有相同的属性值,因此这两者属于同类的可能性较小,把它们归为一类就显得牵强,然而若设 $\alpha=0.5$,则有 $(x,y)\notin V^\alpha(A)$ 且 $(y,x)\notin V^\alpha(A)$,即使用可变精度分类关系可以将这两者分离开来.综上,可变精度分类关系实际上是限制容差关系的一种改进形式,更加符合客观实际.

定理 1　令 S 为一 IIS,其中 $A\subseteq AT$,则 $V^0(A)=T(A)$, $V^1(A)=S(A)$,形式地可以写成

$$T(A)=\left\{(x,y)\in U^2:\forall a\in P_A(x)\cap P_A(y), f(x,a)=f(y,a)\wedge\frac{|P_A(x)\cap P_A(y)|}{|P_A(x)|}\geqslant 0\right\}\cup I_U,\tag{3.2}$$

$$S(A)=\left\{(x,y)\in U^2:\forall a\in P_A(x)\cap P_A(y), f(x,a)=f(y,a)\wedge\frac{|P_A(x)\cap P_A(y)|}{|P_A(x)|}=1\right\}\cup I_U.\tag{3.3}$$

定理 1 说明了通过设置不同的阈值,可以使得根据可变精度分类关系在容差关系和相似关系之间变

化,因而使用可变精度分类关系处理 IIS 具有更好的灵活性,而容差关系和相似关系则是可变精度分类关系的两种极端表现情况,因此 $V_A^\alpha(x)$实际上是 $T_A(x)$和 $S_A^{-1}(x)$一种广义化表现形式.

定义 7　令 S 为一 IIS,其中 $A\subseteq AT$,对于$\forall X\subseteq U$,X 基于 $V^\alpha(A)$的下、上近似集分别记为$\underline{A^\alpha}(X)$,$\overline{A^\alpha}(X)$且

$$\underline{A^\alpha}(X)=\{x\in U:V_A^\alpha(x)\subseteq X\};\overline{A^\alpha}(X)=\{x\in U:V_A^\alpha(x)\cap X\neq\emptyset\}. \tag{3.4}$$

定理 2　令 S 为一 IIS,其中 $A\subseteq AT$,若 $0\leqslant\alpha_1\leqslant\alpha_2\leqslant 1$,则对于$\forall X\subseteq U$ 有

$$V_A^{\alpha_2}(x)\subseteq V_A^{\alpha_1}(x),\underline{A^{\alpha_1}}(X)\subseteq\underline{A^{\alpha_2}}(X),\overline{A^{\alpha_2}}(X)\subseteq\overline{A^{\alpha_1}}(X). \tag{3.5}$$

证明　对于$\forall y\in V_A^{\alpha_2}(x)$,有$|P_A(x)\cap P_A(y)|/|P_A(x)|\geqslant\alpha_2$,所以$|P_A(x)\cap P_A(y)|/|P_A(x)|\geqslant\alpha_1$,即 $y\in V_A^{\alpha_1}(x)$,由于 y 是 $V_A^{\alpha_2}(x)$中任意取得,所以 $V_A^{\alpha_2}(x)\subseteq V_A^{\alpha_1}(x)$成立,再根据粗糙集的相关性质,易证$\underline{A^{\alpha_1}}(X)\subseteq\underline{A^{\alpha_2}}(X)$,$\overline{A^{\alpha_2}}(X)\subseteq\overline{A^{\alpha_1}}(X)$.

定理 3　令 S 为一 IIS,其中 $A\subseteq AT$,对于$\forall x\in U$,$\forall X\subseteq U$,有

$$S_A^{-1}(x)\subseteq V_A^\alpha(x)\subseteq T_A(x),\underline{A^\alpha}(X)\subseteq X\subseteq\overline{A^\alpha}(X); \tag{3.6}$$

$$\underline{A^{\mathrm{T}}}(X)\subseteq\underline{A^\alpha}(X)\subseteq\underline{A^S}(X),\overline{A^S}(X)\subseteq\overline{A^\alpha}(X)\subseteq\overline{A^{\mathrm{T}}}(X). \tag{3.7}$$

证明　由定理 1 和 2,$S_A^{-1}(x)\subseteq V_A^\alpha(x)\subseteq T_A(x)$显然成立,故只需证式 3.7 即可.

对于$\forall x\in U$,若 $V_A^\alpha(x)\subseteq X$,则必定有 $S_A^{-1}(x)\subseteq X$,反之则不一定成立,因此$\underline{A_\alpha}(Z)\subseteq\underline{A_S}(Z)$,类似地可以证得$\underline{A^{\mathrm{T}}}(X)\subseteq\underline{A^\alpha}(X)$.对于$\forall x\in U$,若 $V_A^\alpha(x)\cap X\neq\emptyset$,则必定有 $T_A(x)\cap X\neq\emptyset$,而反之则不一定成立,所以$\overline{A^\alpha}(X)\subseteq\overline{A^{\mathrm{T}}}(X)$成立.由于 $S_A(x)=\{y\in U:x\in S_A^{-1}(y)\}$,因而$\overline{A^S}(X)=\{y\in U:x\in X\wedge x\in S_A^{-1}(y)\}$.对于$\forall y\in\overline{A^S}(X)$,因为 $S_A^{-1}(y)\subseteq V_A^\alpha(y)$,所以必定有 $x\in V_A^\alpha(y)$,又因为 $x\in X$,所以 $V_A^\alpha(y)\cap X\neq\emptyset$,即 $y\in\overline{A^\alpha}(X)$.又由于 y 为$\overline{A^S}(X)$中任意取得,因此$\overline{A^S}(X)\subseteq\overline{A^\alpha}(X)$成立.

定理 4　令 S 为一 IIS,其中 $A\subseteq AT$,对于$\forall X,Y\subseteq U$,有

$$\underline{A^\alpha}(X\cap Y)=\underline{A^\alpha}(X)\cap\underline{A^\alpha}(Y),\underline{A^\alpha}(X\cup Y)\supseteq\underline{A^\alpha}(X)\cup\underline{A^\alpha}(Y); \tag{3.8}$$

$$\overline{A^\alpha}(X\cap Y)\subseteq\overline{A^\alpha}(X)\cap\overline{A^\alpha}(Y),\overline{A^\alpha}(X\cup Y)=\overline{A^\alpha}(X)\cup\overline{A^\alpha}(Y). \tag{3.9}$$

容差关系、相似关系、可变精度分类关系实际上分别代表了 3 种不同的知识粒度思想.由以上的分析可以看出,容差关系实际上表示了一种较粗粒度的覆盖,相似关系表示了一种较细粒度的覆盖,而可变精度分类关系所表示的知识粒度介于这两者之间,并且设置不同的阈值可以得到不同的知识粒度,这恰好符合人类处理现实世界中信息的需要.

4　规则生成与实例分析

粗糙集理论的核心问题是知识约简与知识推理,通过知识约简,可以得到简化规则以更好地指导决策.但是对于一个信息系统来说,可能存在有多个约简,并且不同的约简可以生成不同的决策规则,在实际应用中,很难确定选择哪个约简导出的规则.鉴于此,笔者在文[13]方法的基础上,改进了决策矩阵的形式,利用决策类的下、上近似,可以不用计算约简而得到所有的确定和可能决策规则.

定义 8　一个不完备决策系统为 $S=\langle U,AT\cup D,V,f\rangle$,其中 AT 为条件属性集合,D 为决策属性集合,且 $AT\cap D\neq\emptyset$,$*\notin V_D$.设 $A\subseteq AT$,$U/IND(D)=\{D_1,D_2,\cdots,D_q\}$表示论域上的划分,则决策类 D_l 的决策矩阵形式化定义为 $M_{x,y}^l$ 且

$$M_{x,y}^l=\begin{cases}(a,f(x,a)),f(x,a)\neq *\wedge f(y,a)\neq *\wedge f(x,a)\neq f(y,a)\\ \emptyset,\text{otherwise}\end{cases} \tag{4.1}$$

其中 $x\in\underline{A^\alpha}(D_l)$,$y\in U-D_l(1\leqslant l\leqslant q)$,$a\in P_A(x)$.

决策矩阵表示把某类对象与其他对象分离开来的所有已知信息,利用决策矩阵计算 Boolean 表达式 $B_l=\bigwedge_y\bigvee M_{x,y}^l$(其中 $M_{x,y}^l\neq\emptyset$),$\wedge$和$\vee$分别表示合取和析取运算,B_l 称为知识表示系统的决策函数.将决策函数化为析取范式,如利用 Boolean 代数的吸收律化简,其中每个合取子式对应一条规则.重复上述步骤,就可以得到 D_l 类的所有确定决策规则集.若要生成可能决策规则,只需要修改决策矩阵中的对象 $x\in$

$\overline{A^\alpha}(D_l)$，$y\in U-\overline{A^\alpha}(D_l)$即可．

为了便于对比分析，下面使用文[10]中的一个实际的不完备信息系统进行讨论，其中 $AT=\{a,b,c,d\}$为条件属性集合，而 e 表示决策属性．

若设 $\alpha=0$，由定义 6 可得 $V_{AT}^0(O_1)=\{O_1,O_{11},O_{12}\}$，$V_{AT}^0(O_2)=\{O_2,O_3\}$，$V_{AT}^0(O_3)=\{O_2,O_3\}$，$V_{AT}^0(O_4)=\{O_4,O_5,O_{10},O_{11},O_{12}\}$，$V_{AT}^0(O_5)=\{O_4,O_5,O_{10},O_{11},O_{12}\}$，$V_{AT}^0(O_6)=\{O_6\}$，$V_{AT}^0(O_7)=\{O_7,O_8,O_9,O_{11},O_{12}\}$，$V_{AT}^0(O_8)=\{O_7,O_8,O_{10}\}$，$V_{AT}^0(O_9)=\{O_7,O_9,O_{11},O_{12}\}$，$V_{AT}^0(O_{10})=\{O_4,O_5,O_8,O_{10},O_{11}\}$，$V_{AT}^0(O_{11})=\{O_1,O_4,O_5,O_7,O_9,O_{10},O_{11},O_{12}\}$，$V_{AT}^0(O_{12})=\{O_1,O_4,O_5,O_7,O_9,O_{11},O_{12}\}$．这与根据容差关系所得的分类结果是一致的，同样亦可以计算出 $V_{AT}^1(x)$的结果与 $S_A^{-1}(x)$是一致的．

表 1 不完备信息系统

U	a	b	c	d	e
O_1	3	2	1	0	Φ
O_2	2	3	2	0	Φ
O_3	2	3	2	0	Ψ
O_4	*	2	*	1	Φ
O_5	*	2	*	1	Ψ
O_6	2	3	2	1	Ψ
O_7	3	*	*	3	Φ
O_8	*	0	0	*	Ψ
O_9	3	2	1	3	Ψ
O_{10}	1	*	*	*	Φ
O_{11}	*	2	*	*	Ψ
O_{12}	3	2	1	*	Φ

若设 $\alpha=0.5$，则根据定义 6 可以求得 $V_{AT}^{0.5}(O_1)=\{O_1,O_{11},O_{12}\}$，$V_{AT}^{0.5}(O_2)=\{O_2,O_3\}$，$V_{AT}^{0.5}(O_3)=\{O_2,O_3\}$，$V_{AT}^{0.5}(O_4)=\{O_4,O_5,O_{11},O_{12}\}$，$V_{AT}^{0.5}(O_5)=\{O_4,O_5,O_{11},O_{12}\}$，$V_{AT}^{0.5}(O_6)=\{O_6\}$，$V_{AT}^{0.5}(O_7)=\{O_7,O_9,O_{12}\}$，$V_{AT}^{0.5}(O_8)=\{O_8\}$，$V_{AT}^{0.5}(O_9)=\{O_7,O_9,O_{12}\}$，$V_{AT}^{0.5}(O_{10})=\{O_{10}\}$，$V_{AT}^{0.5}(O_{11})=\{O_4,O_5,O_9,O_{11},O_{12}\}$，$V_{AT}^{0.5}(O_{12})=\{O_1,O_9,O_{12}\}$．与 $V_{AT}^0(x)$相比较，这种分类方式更为合理，例如 O_4 与 O_{10}没有任何已知的相同属性值，但 $O_4\in V_{AT}^0(O_{10})$，而 $O_4\notin V_{AT}^{0.5}(O_{10})$，同样，$O_7$ 与 O_8，O_7 与 O_{11}，O_8 与 O_{10}也是这样情况．限于篇幅起见，此处不再赘述 $V_A^\alpha(x)$相比较于 $S_A^{-1}(x)$的优越性．

根据决策属性进行分类，可得 $D_\Phi=\{O_1,O_2,O_4,O_7,O_{10},O_{12}\}$，$D_\Psi=\{O_3,O_5,O_6,O_8,O_9,O_{11}\}$．$\underline{A^{0.5}}(D_\Phi)=\{O_{10}\}$，$\underline{A^{0.5}}(D_\Psi)=\{O_6,O_8\}$，$\overline{A^{0.5}}(D_\Phi)=\{O_1,O_2,O_3,O_4,O_5,O_7,O_9,O_{10},O_{11},O_{12}\}$，$\overline{A^{0.5}}(D_\Psi)=\{O_1,O_2,O_3,O_4,O_5,O_6,O_7,O_8,O_9,O_{11},O_{12}\}$．根据定义 8，可以得到表 2、表 3 所示的决策矩阵．

表 2 决策类 Φ 的决策矩阵(确定规则)

x \ y	O_3	O_5	O_6	O_8	O_9	O_{11}
O_{10}	$(a,1)$	$\varnothing$	$(a,1)$	$\varnothing$	$(a,1)$	$\varnothing$

表 3 决策类 Φ 的决策矩阵(可能规则)

x \ y	O_6	O_8	O_{12}
O_1	$(a,3)(b,2)(c,1)(d,0)$	$(b,2)(c,1)$	$\varnothing$
O_2	$(d,0)$	$(b,3)(c,2)$	$(a,2)(b,3)(c,2)$
O_3	$(d,0)$	$(b,3)(c,2)$	$(a,2)(b,3)(c,2)$
O_4	$(b,2)$	$(b,2)$	$\varnothing$
O_5	$(b,2)$	$(b,2)$	$\varnothing$
O_7	$(a,3)(d,3)$	$\varnothing$	$\varnothing$
O_9	$(a,3)(b,2)(c,1)(d,3)$	$(b,2)(c,1)$	$\varnothing$
O_{10}	$(a,1)$	$\varnothing$	$(a,1)$
O_{11}	$(b,2)$	$(b,2)$	$\varnothing$

于是根据表 2 可以得到简化的确定性决策规则：$(a,1)\rightarrow(e,\Phi)$．

由表 3 可以得到简化的可能性决策规则：$(b,2)\vee(c,1)\rightarrow(e,\Phi)$；$(b,3)\wedge(d,0)\rightarrow(e,\Phi)$；$(c,2)\wedge$

$(d,0)\rightarrow(e,\Phi)$；$(a,3)\vee(d,3)\rightarrow(e,\Phi)$；$(a,1)\rightarrow(e,\Phi)$.

类似上述过程，可以得到关于决策类 Ψ 的确定规则为：$(a,2)\wedge(b,3)\wedge(d,1)\rightarrow(e,\Psi)$；$(b,0)\rightarrow(e,\Psi)$.

关于决策类 Ψ 的可能规则为：$(a,2)\vee(a,3)\rightarrow(e,\Psi)$.

5 结束语

粗糙集理论发展至今，以其独特的优势在智能信息处理领域占据着重要的地位. 由于现实世界可能存在的数据不完备性，因而促进了各种拓展粗集模型的产生与发展. 其中最为常用的扩展二元关系有容差关系与相似关系. 为了使得对于 IIS 中的分类更具有一般性，本文构建了一种可变精度分类关系，设置一个阈值来控制 IIS 中的分类，使得建立起来的粗集模型更加符合客观实际的需要. 而由于可变精度分类所得结果介于容差和相似关系分类结果之间，对于知识粒度的把握更为灵活，因此本文的工作具有更为广泛的意义. 然后，笔者研究了不进行约简而直接从 IIS 中推导出确定和可能性规则的方法，并用实例进行了验证.

在本文所做工作的基础上，下一步的工作就是对挖掘出来规则的可信度进行度量. 另一方面，从不完备信息系统中挖掘高阶规则[14]也是笔者的一个研究重点.

参考文献：

[1] Pawlak Z. Rough set theory and its applications to data analysis[J]. Cybernetics and Systems, 1998, 29: 661 – 688.

[2] Pawlak Z. Rough sets and intelligent data analysis[J]. Information Sciences, 2002, 147: 1 – 12.

[3] 刘清. Rough 集及 Rough 推理[M]. 科学出版社, 2001, 3.
Liu Q. Rough sets and rough reasoning[M]. Science Press, 2001, 3.

[4] 王国胤. Rough 集理论在不完备信息系统中的扩充[J]. 计算机研究与发展, 2002, 39(10): 1238 – 1243.
Wang G Y. Extension of rough set under incomplete information systems[J]. Journal of Compute Research and Development, 2002, 39(10), 1238 – 1243.

[5] Grzymala-Busse J W. On the unknown attribute values in learning from examples[J]. In Proceeding of the 6th International Symposium on Methodologies for Intelligent Systems(ISMIS-91), Charlotte, North Carolina, October 16 – 19, 1991. Lecture Notes in Artificial Intelligence, vol. 542, Springer-Verlag, Berlin, Heidelberg, New York (1991) 368 – 377.

[6] Kryszkiewicz M. Rough set approach to incomplete information systems[J]. Information Sciences, 1998, 112: 39 – 49.

[7] Yee Leung, Deyu Li. Maximal consistent block technique for rule acquisition in incomplete information systems[J]. Information Sciences, 2003, 15: 85 – 106.

[8] 吴陈, 杨习贝, 傅凡. 基于全相容性粒度的粗糙集模型[J]. 系统工程学报, 2006, 21(3): 292 – 298.
Wu C, Yang X B, Fu F. Rough set model based on the granulated view of complete compatibility[J]. Journal of System Engineering, 21(3): 292 – 298.

[9] Grzymala-Busse J W, Wang A Y. Modified algorithms LEM1 and LEM2 for rule induction from data with missing attribute values[C]//In Proceeding of the Fifth International Workshop on Rough Sets and Soft Computing (RSSC '97) at the Third Joint Conference on Information Sciences (JCIS '97), Research TrianglePark, NC, March 2 – 5, 1997, 69 – 72.

[10] Stefanowski J, Tsoukias A. Incomplete information tables and rough classification[J]. Computational Intelligence, 2001, 17: 545 – 566.

[11] Wu W Z, Zhang W X, Li H Z. Knowledge acquisition in incomplete fuzzy information systems via the rough set approach[J]. Expert Systems, 2003, 20(5): 280 – 286.

[12] Tzung-Pei H. Learning rules from incomplete training examples by rough sets[J]. Expert Systems with Application, 2002, 22: 285 – 293.

[13] Ning Shan, Wojciech Ziarko. Data-based acquisition and incremental modification of classification rules[J]. Computational Intelligence, 1995, 11(2): 357 – 370.

[14] Yao Y Y. Probabilistic approaches to rough sets[J]. Expert Systems, 2003, 20: 287 – 297.