不完备目标信息系统中基于描述子的知识约简*)

谢 军[1,2] 杨习贝[1] 杨静宇[1] 孙怀江[1] 宋余庆[2]
(南京理工大学计算机科学与技术学院 南京 210094)[1]
(江苏大学计算机科学与通信工程学院 镇江 212013)[2]

摘 要 以具有遗漏型未知属性值的不完备目标信息系统为研究对象,根据描述子的定义和基于描述子的粗糙集模型,讨论了知识约简问题。给出了求得描述子所有约简的具体操作方法。根据描述子的支持集与决策类之间的关系,提出了描述子的下、上近似相对约简概念,并给出这两种约简的判定定理及区分函数,为从不完备信息系统中获取简化的决策规则提供了新的理论基础与操作手段。

关键词 不完备目标信息系统,描述子,粗糙集,约简

Knowledge Reductions Based on Descriptors in Incomplete Systems

XIE Jun[1,2] YANG Xi-bei[1] YANG Jing-yu[1] SUN Huai-jiang[1] SONG Yu-qing[2]
(School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China)[1]
(School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, China)[2]

Abstract The incomplete information system in which all unknown values were looked as lost was deeply investigated. In such information system, approaches to knowledge reductions were studied from the viewpoint of the descriptor and descriptor-based rough set model. The practical approach to compute all reducts of a descriptor was presented. Based on the relationships between the support set of a descriptor and decision classes, the concept of descriptor-based lower and upper approximate relative reducts were proposed, the judgment theorems, discernibility functions with respect to these reducts were obtained, from which we obtained new theory and practical approach to derive simplified decision rules from the incomplete objective information system.

Keywords Incomplete objective information system, Descriptor, Rough set, Reduction

1 引言

近年来,经典粗糙集理论[1,2]在不完备信息系统[3-7]中的扩充已经成为一个热点研究问题,这也是将粗糙集理论进一步推向实际应用的关键。一个不完备信息系统是指信息系统中出现了未知属性值。就目前已有的研究结果来看,不完备信息系统中的未知属性值可以具有两种不同的语义解释[3,4]:1)所有的未知属性值仅仅是被遗漏的,但又是确实存在的;2)所有的未知属性值被认为是缺席的,是不允许被比较的。当不完备信息系统中所有的未知属性值被认为是遗漏型时,最为常用的拓展粗糙集模型是由 Kryszkiewicz[5]提出的基于容差关系的粗糙集;而当不完备信息系统中所有的未知属性值被认为是缺席型时,最为常用的拓展粗糙集模型是由 Stefanowski[6]提出的基于非对称相似关系的粗糙集。

值得注意的是,不完备信息系统还有另外一种表现形式,即信息系统中的某些对象在某些属性上的取值是不确定的,即取集合值。针对这种集值信息系统,Leung[7]等人提出了基于描述子的粗糙集模型。吴伟志[8,9]等人对基于描述子的粗集模型和基于容差关系的粗集模型进行了对比分析,证明了使用基于描述子的粗集模型可以比基于容差关系的粗集模型获得更高的粗糙近似精度。然而遗憾的是,Leung 和吴伟志等人都没有对不完备信息系统中的描述子和基于描述子的粗集模型进行知识约简的研究工作,这正是本文所要讨论的内容。

笔者的研究对象是具有遗漏型未知属性值的不完备目标信息系统。在这种不完备系统中,笔者根据 Leung 提出的描述子的思想,构建了基于描述子的粗糙集模型,给出了描述子约简的定义,给出了求得描述子所有约简的具体操作手段。描述子的约简是保持描述子的支持集中的元素不发生变化的最小数量原子公式的合取。根据描述子的支持集与决策类之间的关系,笔者提出了描述子的下、上近似相对约简的概念,描述子的下、上近似相对约简分别是保持描述子的支持集与决策类之间的包含、相交关系不发生变化的最小数量原子公式的合取。根据描述子的下、上近似相对约简的定义,笔者给出了这两种约简的判定定理和可辨识属性矩阵,为从不完备目标信息系统中获得基于描述子的简化决策规则提供了具体的操作手段。

综上,本文的主要内容安排如下:第 2 节简要介绍不完备目标信息系统,给出基于描述子的粗糙集模型;第 3 节主要讨论描述子的约简;第 4 节研究描述子的下、上近似相对约简,并进行实例分析;最后总结全文。

2 基本概念

2.1 容差关系

*)国家自然科学基金(No. 60472060,60572034),国家自然科学基金重点项目(No. 60632050),江苏省自然科学基金(No. BK2006081)资助项目。谢 军 博士研究生,研究方向为智能信息处理、生物信息。

目标信息系统是既有条件属性又有目标属性(决策属性)的一种特殊信息系统,主要是研究条件属性和目标属性之间的关系问题。一个目标信息系统为一个四元组:$OIS=\langle U, AT\cup d, V, f\rangle$,其中

• U 是一个被称为论域的非空有限对象集合;

• AT 是非空有限条件属性集合,d 是目标属性且 $AT\cap\{d\}=\varnothing$;

• V_{AT} 是所有条件属性的值域集合,V_d 是目标属性的值域集合,因而就有 $V=V_{AT}\cup V_d$,V 表示全体属性的值域集合;

• f 为信息函数,对于 $\forall a\in AT$,$\forall x\in U$,有 $f(x,a)\in V_a$。

不完备目标信息系统(记为 $IOIS$)是指信息系统中出现了未知属性值。由于假设专家在信息不完备的情况下也能给出决策,因而文中所讨论的未知属性值仅出现在对象的条件属性值上,用"$*$"表示,即 $f(x,a)=*\ (x\in U, a\in AT)$。于是就有 $V=V_{AT}\cup V_d\cup\{*\}$,此时的目标属性构成了论域上的划分,记为 $U/d=\{X_1, X_2,\cdots, X_m\}(\forall 1\in i, j\leqslant m, X_i\cap X_j=\varnothing)$。另一方面,我们认为"$*$"这种未知属性值仅仅是被遗漏的,但又是确实存在的,因而"$*$"被认为与任何已知属性值都是可以比较的。根据这样的解释,可以构建如下所示的容差关系。

定义 1[5] 在不完备目标信息系统 $IOIS$ 中,对于 $A\subseteq AT$,由 A 决定的容差关系记为 $S(A)$,且

$$S(A)=\{(x,y)\in U^2:\forall a\in A, f(x,a)=f(y,a)\vee f(x,a)=*\vee f(y,a)=*\} \tag{1}$$

在不完备目标信息系统 $IOIS$ 中,对于 $\forall x\subseteq U$,x 的容差类可记为 $S_A(x)$,且 $S_A(x)=\{y\in U:(x,y)\in S(A)\}$,即 $S_A(x)$ 是所有与 x 具有容差关系 $S(A)$ 的对象的集合。

定义 2[5] 设 $IOIS$ 为一不完备目标信息系统,其中 $A\subseteq AT$,对于 $\forall X\subseteq U$,X 基于 $S(A)$ 的下、上近似集分别记为 $A_*(X^S)$,$A^*(X^S)$,且

$$A_*(X^S)=\{x\in U:S_A(x)\subseteq X\};$$
$$A^*(X^S)=\{x\in U:S_A(x)\cap X\neq\varnothing\} \tag{2}$$

2.2 描述子

定义 3[7] 令 $IOIS$ 为一不完备目标信息系统,

• 对于 $\forall a\in AT$,$\forall v\in V_a$,称 (a,v) 为一个 AT——原子公式;

• 任意一个 AT——原子公式或几个 AT——原子公式的合取称为一个 AT——描述子;

• 若 $\bigwedge\limits_{a_i\in AT}(\geqslant_{a_i}, v_i)$ 为一个 AT——描述子,令

$$\|\bigwedge_{a_i\in AT}(\geqslant_{a_i}, v_i)\|=\{x\in U:\forall a_i\in AT, f(x,a_i)=*\vee f(x,a_i)=v_i\}$$

称 $\|\bigwedge\limits_{a_i\in AT}(\geqslant_{a_i}, v_i)\|$ 中的对象与描述子 $\bigwedge\limits_{a_i\in AT}(\geqslant_{a_i}, v_i)$ 是相容的,$\|\bigwedge\limits_{a_i\in AT}(\geqslant_{a_i}, v_i)\|$ 称为描述子 $\bigwedge\limits_{a_i\in AT}(\geqslant_{a_i}, v_i)$ 的支持集。

• 若 t 是一个 AT——描述子,则 t 中出现的所有属性的集合记为 $AT(t)$。

设 t 和 s 是两个 AT——描述子,若对于 $\forall(a,v)\in t$,都有 $(a,v)\in s$,换言之,t 是由 s 中出现的原子公式构成,则称 t 比 s 更为粗,或者 s 比 t 更为细,记为 $t\geq s$ 或 $s\leq t$。进一步地,若 t 是由 s 中出现的原子公式的真子集构成,则称 t 比 s 严格粗,或者 s 比 t 严格细,记为 $t>s$ 或 $s<t$。

定义 4 令 $IOIS$ 为一不完备目标信息系统,记

$DES(AT)=\{t: t$ 是 AT——描述子,$\|t\|\neq\varnothing\}$

对于 $\forall t(DES(AT)$,若 $AT(t)=AT$,则称 t 为一个完全 AT——描述子,记

$FDES(AT)=\{t: t\in DES(AT), t$ 是完全 AT——描述子$\}$。

定义 5 设 $IOIS$ 为一不完备目标信息系统,对于 $\forall X\subseteq U$,X 基于描述子的下、上近似集分别记为 $AT_*(X^{des})$,$AT^*(X^{des})$,且

$$AT_*(X^{des})=\{\|t\|:\|t\|\subseteq X, t\in FDES(AT)\} \tag{3}$$

$$AT^*(X^{des})=\{\|t\|:\|t\|\cap X\neq\varnothing, t\in FDES(AT)\} \tag{4}$$

值得注意的是,与基于容差关系的粗糙集模型不同,基于描述子的下、上近似不是论域的子集,而是论域上的子集族[7]。

定理 1[8,9] 设 $IOIS$ 为一不完备目标信息系统,对于 $\forall t\in FDES(AT)$,必定存在 $x\in U$,使得 $\|t\|\subseteq S_{AT}(x)$。

定理 2[8,9] 设 $IOIS$ 为一不完备目标信息系统,对于 $\forall x\in U$,有

$$S_{AT}(x)=\bigcup\{\|t\|: t\in FDES(AT), x\in\|t\|\}$$

定理 3[8,9] 设 $IOIS$ 为一不完备目标信息系统,对于 $\forall X\subseteq U$,有

$$AT_*(X^S)\subseteq\bigcup AT_*(X^{des})\subseteq X\subseteq AT^*(X^S)=\bigcup AT^*(X^{des})$$

定理 3 告诉我们,在不完备目标信息系统中,使用基于描述子的粗集模型比使用基于容差关系的粗集模型可以获得更高的粗糙近似精度。在不完备目标信息系统中,不难发现,$FDES(AT)$ 中所有描述子的支持集构成了论域上的一个覆盖,可以将这种覆盖记为 $U/AT^{des}=\{\|t\|: t\in FDES(AT)\}$。

例 1 表 1 是一个用来进行学生评价的不完备目标信息系统,其中 $AT=\{a,b,c\}=\{$Mathematics, Physics, Literature$\}$ 为条件属性集合,$d=\{$Global evaluation$\}$ 为决策属性,$V_a=V_b=V_c=V_d=\{$Good, Medium, Bad$\}$。于是可以求得 $U/AT^{des}=\{t_1, t_2,\cdots, t_{11}\}$,其中

$t_1=(a,$ Bad$)\wedge(b,$ Medium$)\wedge(c,$ Medium$)$, $\|t_1\|=\{x_4\}$,

$t_2=(a,$ Bad$)\wedge(b,$ Good$)\wedge(c,$ Bad$)$, $\|t_2\|=\{x_5\}$,

$t_3=(a,$ Medium$)\wedge(b,$ Bad$)\wedge(c,$ Bad$)$, $\|t_3\|=\{x_1\}$,

$t_4=(a,$ Medium$)\wedge(b,$ Bad$)\wedge(c,$ Medium$)$, $\|t_4\|=\{x_3\}$,

$t_5=(a,$ Medium$)\wedge(b,$ Medium$)\wedge(c,$ Medium$)$, $\|t_5\|=\{x_3, x_4\}$,

$t_6=(a,$ Medium$)\wedge(b,$ Good$)\wedge(c,$ Bad$)$, $\|t_6\|=\{x_5\}$,

$t_7=(a,$ Medium$)\wedge(b,$ Good$)\wedge(c,$ Medium$)$, $\|t_7\|=\{x_3\}$,

$t_8=(a,$ Good$)\wedge(b,$ Medium$)\wedge(c,$ Bad$)$, $\|t_8\|=\{x_2, x_6\}$,

$t_9=($a, Good$)\wedge(b,$ Medium$)\wedge(c,$ Medium$)$, $\|t_9\|=\{x_2, x_4\}$,

$t_{10}=(a,$ Good$)\wedge(b,$ Medium$)\wedge(c,$ Good$)$, $\|t_{10}\|=\{x_2\}$,

$t_{11}=(a,$ Good$)\wedge(b,$ Good$)\wedge(c,$ Bad$)$, $\|t_{11}\|=\{x_5\}$。

表 1　不完备目标信息系统

Student	Mathematics	Physics	Literature	Global evaluation
x_1	Medium	Bad	Bad	Bad
x_2	Good	Medium	*	Good
x_3	Medium	*	Medium	Medium
x_4	*	Medium	Medium	Medium
x_5	*	Good	Bad	Medium
x_6	Good	Medium	Bad	Medium

3　描述子的约简

在文献[7]中，Leung 定义了描述子约简的概念，但并未给出计算描述子约简的具体方法。

定义 6　设 $IOIS$ 为一不完备目标信息系统，$t\in FDES(AT)$，对于 $\forall t'\in DES(AT)$，若 t' 满足

(1)$t'\geq t$，$\| t'\| = \| t\|$，

(2)对于 $\forall t''>t'$，$\| t''\| \neq \| t\|$，

则称 t' 是 t 的一个约简描述子。

对于一个 AT——描述子 t 来说，它的约简描述子是保持 $\| t\|$ 中的元素不发生变化的最小数量原子公式的合取，t 的所有约简描述子记为 $red(t)$。

在不完备目标信息系统中，$tFDES(AT)$，不妨假设 $t=\bigwedge\limits_{a_i\in AT}(\geqslant_{a_i}, v_i)$，对于 $\forall x\in U$，令

$$D(t,x)=\begin{cases}\{a_i\in AT: f(x,a_i)\text{已知且 } f(x,a_i)(v_i)\} & : x\notin \| t\| \\ AT & : x\in \| t\|\end{cases}$$

则称 $D(t,x)$ 为描述子 t 与对象 x 的分辨属性集，称 $D(t)=\{D(t,x): t\in FDES(AT), x\in U\}$ 为不完备目标信息系统中描述子的分辨矩阵。

定理 4　设 $IOIS$ 为一不完备目标信息系统，若 $t\in FDES(AT)$，$t'\in DES(AT)$ 且 $t'\geq t$，于是就有

$$\| t'\| = \| t\| \Leftrightarrow D(t,x)\cap AT(t')\neq\varnothing\ (\forall x\in U, x\notin \| t\|)$$

证明："⇒"　假设存在 $x\in U$，$x\notin \| t\|$ 使得 $D(t,x)\cap AT(t')=\varnothing$。因为 $t'\geq t$，所以就有 $x\in \| t'\|$。因为根据条件有 $\| t'\| = \| t\|$，所以 $x\in \| t\|$，这与 $x\notin \| t\|$ 矛盾。

"⇐"　因为 $t'\geq t$，所以就有 $\| t'\|\supseteq \| t\|$，故只需证 $\| t'\|\subseteq \| t\|$ 即可。对于 $\forall x\notin \| t\|$，根据条件有 $D(t,x)\cap A(t')\neq\varnothing$，所以必定存在 $a\in D(t,x)\cap AT(t')$ 使得 $f(x,a)$ 已知且 $f(x,a)\neq v(v\in V_a)$，故 $x\notin \| t'\|$，所以可以得出 $x\in \| t'\|\Rightarrow x\in \| t\|$，即 $\| t'\|\subseteq \| t\|$。

定义 7　设 $IOIS$ 为一不完备目标信息系统，$t\in FDES(AT)$，令

$$\Delta(t)=\bigwedge_{x\in U}(\bigvee D(t,x))$$

称 $\Delta(t)$ 为描述子 t 的区分函数。

根据布尔推理理论，由定义 6 及定理 4 可得如下结论。

定理 5　$\Delta(t)$ 的极小析取范式中的每个合取项对应的属性子集即为 t 的约简描述子中出现的属性集合。

例 2　在例 1 所示的不完备目标信息系统中，可求得描述子的分辨矩阵，如表 2 所示。

根据定义 7 可以求得 $\Delta_{AT}(t_1)=(a\wedge b)\vee(a\wedge c)$，根据定理 5 可以求得 t_1 的约简描述子的集合为 $red(t_1)=\{(a,\text{Bad})\wedge(b,\text{Medium}),(a,\text{Bad})\wedge(c,\text{Medium})\}$。类似地，可以求出其他描述子的约简如下：

$red(t_2)=\{(a,\text{Bad})\wedge(b,\text{Good}),(a,\text{Bad})\wedge(c,\text{Bad}),(b,\text{Good})\wedge(c,\text{Bad})\}$，$red(t_3)=\{(b,\text{Bad})\wedge(c,\text{Bad})\}$，$red(t_4)=\{(b,\text{Bad})\wedge(c,\text{Medium})\}$，$red(t_5)=\{(a,\text{Medium})\wedge(b,\text{Medium}),(a,\text{Medium})\wedge(c,\text{Medium})\}$，$red(t_6)=\{(b,\text{Good})\wedge(c,\text{Bad})\}$，$red(t_7)=\{(b,\text{Good})\wedge(c,\text{Medium})\}$，$red(t_8)=\{(b,\text{Medium})\wedge(c,\text{Bad})\}$，$red(t_9)=\{(a,\text{Good})\wedge(c,\text{Medium})\}$，$red(t_{10})=\{(c,\text{Medium})\}$，$red(t_{11})=\{(a,\text{Good})\wedge(b,\text{Good}),(b,\text{Good})\wedge(c,\text{Bad})\}$

表 2　表 1 中描述子的分辨矩阵

Student	x1	x2	x3	x4	x5	x6
t1	AT	a	a	AT	b, c	a, c
t2	a, b	a, b	a, c	b, c	AT	a, b
t3	AT	a, b	c	b, c	b	a, b
t4	c	a, b	AT	b	b, c	AT
t5	b, c	a	AT	AT	b, c	a, c
t6	b	a, b	c	b, c	AT	a, b
t7	b, c	a, b	AT	b	c	AT
t8	a, b	AT	a, c	c	b	AT
t9	AT	AT	a	AT	b, c	c
t10	AT	AT	a, c	c	b, c	c
t11	a, b	b	a, c	b, c	AT	b

4　决策规则与描述子的相对约简

定义 8　设 S 为一不完备目标信息系统，$A=\{a_1,a_2,\cdots,a_m\}\subseteq AT$，$t=\bigwedge\limits_{a_i\in A}(\geqslant_{a_i}, v_i)\in FDES(A)$，对于 $\forall X\subseteq U$，若

(1)若 $\| \bigwedge\limits_{a_i\in A}(\geqslant_{a_i}, v_i)\| \subseteq X$，则称

$$f(x,a_1)=v_1\wedge f(x,a_2)=v_2\wedge\cdots\wedge f(x,a_m)=v_m\rightarrow x\in X$$

为一条基于描述子 t 的确定决策规则；

(2)若 $\| \bigwedge\limits_{a_i\in A}(\geqslant_{a_i}, v_i)\| \cap X\neq\varnothing$，则称

$$f(x,a_1)=v_1\wedge f(x,a_2)=v_2\wedge\cdots\wedge f(x,a_m)=v_m\rightarrow x\in X$$

为一条基于描述子 t 的可能决策规则。

知识约简是粗糙集理论中的核心问题，通过各种各样形式的约简，可以得到满足一定条件的最小属性子集，从而获得简化的决策规则。在基于描述子的不完备目标信息系统中，我们将提出两种不同形式的描述子相对约简的概念。

定义 9　设 $IOIS$ 为一不完备目标信息系统，其中 $t\in FDES(AT)$，令

$$L(t)=\{X_i: \| t\| \subseteq X_i\},\ H(t)=\{X_i: \| t\| \cap X_i\neq\varnothing\}$$

称 $L(t)$ 为描述子 t 的下近似分配集，$H(t)$ 为描述子 t 的上近似分配集。

定义 10　设 $IOIS$ 为一不完备目标信息系统，其中 $t\in FDES(AT)$，$t'\in DES(AT)$，$t'\geq t$，

(1)$L(t)=L(t')$，且对于 $\forall t''>t$，有 $L(t)\neq L(t'')$，称 t' 是 t 的一个下近似相对约简描述子；

(2)$H(t)=H(t')$，且对于 $\forall t''>t$，有 $H(t)\neq H(t'')$，称 t' 是 t 的一个上近似相对约简描述子。

从以上定义可以看出，t 的下近似相对约简描述子是保持满足以下条件的决策类不发生变化的最小数量原子公式的合取：t 的支持集属于某个决策类；而 t 的上近似相对约简描述子是保持所有满足以下条件的决策类的集合不发生变化的最小数量原子公式的合取：t 的支持集与某个决策类相交不

为空。t 的所有下近似相对约简描述子记为 $red_L(t)$，t 的所有上近似相对约简描述子记为 $red_H(t)$。

值得注意的是，对于某个描述子 $t\in FDES(AT)$来说，它的下近似分配集可能为空集，此时在不完备目标信息系统中就没有基于 t 的确定决策规则，因而讨论这种描述子的下近似相对约简是没有意义的。

定义 11 设 *IOIS* 为一不完备目标信息系统，若 $t\in FDES(AT)$，不妨假设 $t\bigwedge_{a_i\in AT}(\geqslant_{a_i}, v_i)$，记

$$D_L(t)=\{x:\forall X_i\in U/d,\ \|t\|\subseteq X_i,\ \forall x\in U-X_i\} \tag{5}$$

$$D_H(t)=\{x:\forall X_i\in U/d,\ \|t\|\cap X_i=\emptyset,\forall x\in X_i\} \tag{6}$$

定义

$$D_L(t,x)=\begin{cases}\{a_i\in AT: f(x,a_i)\text{已知且 } f(x,a_i)\neq v_i\}, & x\in D_L(t)\\ AT, & x\notin D_L(t)\end{cases}$$

$$D_H(t,x)=\begin{cases}\{a_i\in AT: f(x,a_i)\text{已知且 } f(x,a_i)\neq v_i\}, & x\in D_H(t)\\ AT, & x\notin D_H(t)\end{cases}$$

为 S 中描述子 t 的下、上近似的可辨识属性集。

定理 6 设 *IOIS* 为一不完备目标信息系统，若 $t\in FDES(AT)$，$t'\in DES(AT)$，且 $t'\geq t$，于是就有

(1) $L(t)=L(t')\Leftrightarrow D_L(t,x)\cap AT(t')\neq\emptyset$ $(\forall x\in D_L(t))$；

(2) $H(t)=H(t')\Leftrightarrow D_H(t,x)\cap AT(t')\neq\emptyset$ $(\forall x\in D_H(t))$。

证明：(1)“⇒” 假设存在 $X_i\in U/d(\|t\|\subseteq X_i)$且存在 $x\in U-X_i$ 使得 $D(t,x)\cap A(t')=\emptyset$。因为 $t'\geq t$，所以就有 $x\in\|t'\|$。根据条件有 $L(t)=L(t')$，即 $\|t\|\subseteq X_i\Leftrightarrow\|t'\|\subseteq X_i$，此时就有 $x\in X_i$，这与 $x\in U-X_i$ 矛盾。

“⇐” 因为 $t'\geq t$，所以就必定有 $L(t)\supseteq L(t')$，故只需证 $L(t)\subseteq L(t')$即可。对于 $L(t)\neq\emptyset$ 且 $X_i\in L(t)$，即 $\|t\|\subseteq X_i$，此时根据条件有：对于 $\forall x\in U-X_i$，$D_A(t,x)\cap AT(t')\neq\emptyset$ 成立，即 $x\notin\|t'\|$。因为 x 是 $U-X_i$ 中任意选定的，所以此时必定有 $\|t'\|\subseteq X_i$，从而 $L(t)\subseteq L(t')$。

(2)“⇒” 假设存在 $X_i\in U/d(\|t\|\cap X_i=\emptyset)$且存在 $x\in X_i$ 使得 $D_H(t,x)\cap AT(t')=\emptyset$，所以就有 $x\in\|t'\|$。根据条件有 $H(t)=H(t')$，即 $\|t\|\cap X_i=\emptyset\Leftrightarrow\|t'\|\cap X_i=\emptyset$，此时就有 $x\in X_i$，这与 $x\in X_i$ 矛盾。

“⇐” 因为 $t'\geq t$，所以就必定有 $H(t)\subseteq H(t')$，故只需证 $H(t)\supseteq H(t')$即可。对于 $\forall X_i\notin H(t)$，即 $\|t\|\cap X_i=\emptyset$，此时根据条件有：对于 $\forall x\in X_i$，$D_H(t,x)\cap AT(t')\neq\emptyset$ 成立，即 $x\notin\|t'\|$。因为 x 是 X_i 中任意选定的，所以此时必定有 $\|t'\|\cap X_i=\emptyset$，从而 $X_i\in H(t')$，即 $H(t')\subseteq H(t)$。

定义 12 设 *IOIS* 为一不完备目标信息系统，$t\in FDES(AT)$，令

$$\Delta_L(t)=\bigwedge_{x\in D_L(t)}(\vee D_L(t,x)),$$

$$\Delta_H(t)\bigwedge_{x\in D_H(t)}(\vee D_H(t,x))$$

称 $\Delta_L(t)$和 $\Delta_H(t)$为描述子 t 的下、上近似相对约简的区分函数。

根据布尔推理理论，由定义 11 及定理 6 可得如下结论。

定理 7 $\Delta_L(t)(\Delta_H(t))$的极小析取范式中的每个合取项对应的属性子集即为 t 的下(上)近似相对约简描述子中出现的属性集合。

例 3 对于表 1 所示的不完备目标信息系统，目标属性 d 将论域形成划分 $U/d=\{X_{Good},X_{Medium},X_{Bad}\}=\{\{x_2\},\{x_3,x_4,x_5,x_6\},\{x_1\}\}$。

以描述子 t_1 为例，因为 $\|t_1\|\subseteq X_{Medium}$，所以根据定义 12，可以求得

$$\Delta_L(t_1)=(\vee D_L(t_1,x_1))\wedge(\vee D_L(t_1,x_2))=AT\wedge a=a$$

根据定理 7 可知(a,Bad)是 t_1 的下近似相对约简描述子。于是根据定义 8 可以得到一条简化的确定决策规则：

$r_1: f(x,a)=\text{Bad}\rightarrow f(x,d)=\text{Medium}$

类似于上述过程，可以求得表 1 中所有简化的确定决策规则如下：

$r_2: f(x,a)=\text{Bad}\wedge f(x,b)=\text{Good}\rightarrow f(x,d)=\text{Medium}$

$r_3: f(x,b)=\text{Bad}\wedge f(x,c)=\text{Bad}\rightarrow f(x,d)=\text{Bad}$

$r_4: f(x,a)=\text{Medium}\wedge f(x,c)=\text{Medium}\rightarrow f(x,d)=\text{Medium}$

$r_5: f(x,b)=\text{Bad}\wedge f(x,c)=\text{Medium}\rightarrow f(x,d)=\text{Medium}$

$r_6: f(x,a)=\text{Medium}\wedge f(x,b)=\text{Medium}\rightarrow f(x,d)=\text{Medium}$

$r_7: f(x,b)=\text{Good}\rightarrow f(x,d)=\text{Medium}$

$r_8: f(x,c)=\text{Good}\rightarrow f(x,d)=\text{Good}$

表 1 中所有简化的可能决策规则如下：

$r_9: f(x,a)=\text{Good}\wedge f(x,b)=\text{Medium}\rightarrow f(x,d)=\text{Medium or Good}$

结束语 粗糙集理论在不完备信息系统中的扩充一直是粗糙集理论研究的热点和难点问题。在不完备目标信息系统中，Leung 提出了基于描述子的粗糙集模型，杨晓平、吴伟志证明了基于描述子的粗糙集模型要比基于容差关系的粗糙集模型具有更高的粗糙近似精度，但他们都没有进行不完备信息系统中与描述子相关的知识约简。本文首先给出了求得描述子所有约简的具体操作手段，然后在不完备目标信息系统中提出了描述子的下、上近似相对约简的概念。通过约简，可以使用较少的原子公示的合取来保持原描述子与决策类之间的包含与相交关系，从而可以获得简化的决策规则。综上，本文工作都为从不完备系统中获取知识提供了新的理论方法和技术手段。

笔者下一步的研究方向将是对基于描述子的可能规则的度量方法进行讨论与研究。此外，由于本文所研究的不完备目标信息系统中的所有属性为常规属性(属性值之间没有序关系)，因此笔者的另一个研究方向将是讨论不完备序值信息系统中基于描述子的粗糙集模型。

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] 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

[4] Grzymala-Busse J W. Data with Missing Attribute Values: Generalization of Indiscernibility Relation and Rule Induction[J]// Transactions on Rough Sets I, Lecture Notes in Computer Sci-

ence, vol. 3100. Berlin: Springer-Verlag, 2004: 78-95

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

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

[7] Leung Y, Wu W Z, Zhang W X. Knowledge acquisition in incomplete information systems: A rough set approach[J]. European Journal of Operational Research, 2006, 168: 164-180

[8] Wu W Z, Xu Y H. On two types of generalized rough set approximations in incomplete information systems // Hu Xiaohua, Liu Qing, Skowron A, et al., eds. 2005 IEEE International Conference on Granular Computing. Beijing, China, July 2005: 303-306

[9] 杨晓平. 不完备信息系统一种新的粗糙集的性质[J]. 计算机科学, 2004, 31(10A): 64-65, 94

(上接第 147 页)

为标注文档能增加上下文元素。

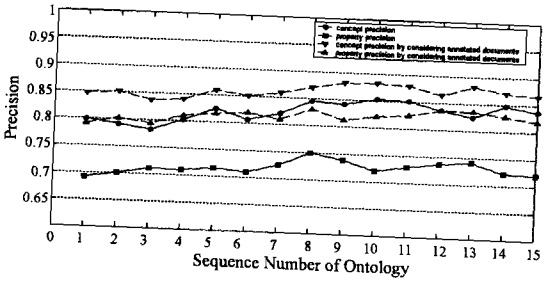

图 3 精度比较图

表 2 平均精度比较

	平均概念精度	平均关系精度	平均关系精度+	平均概念精度+
简单本体	76.9%	70.5%	84.5%	80.3%
复杂本体	83.8%	73.0%	87.1%	82.3%
Overall	82.4%	72.2%	86.3%	81.6%

最后一个实验评估本体澄清算法在半自动过程中的有效性。计算精度的公式为

$$Precision_n = \frac{correct\ disambiguated\ termsintop\ nsenses}{all\ disambiguated\ terms}$$

$$Concept\ Precision_n = \frac{correct\ disambiguated\ concept\ terms\ in\ top\ n\ senses}{all\ disambiguated\ concept\ terms}$$

$$Property\ Precision_n = \frac{correctd\ isambiguated\ property\ terms\ in\ top\ n\ senses}{all\ disambiguated\ propert\ yterms}$$

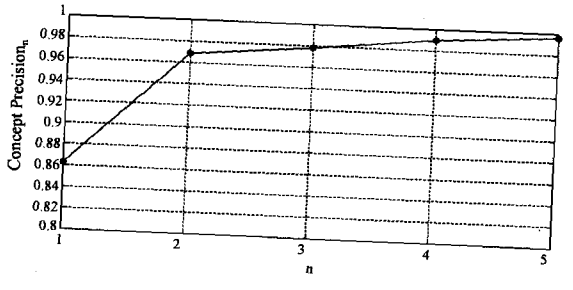

图 4 随 n 变化的 Concept Precision$_n$

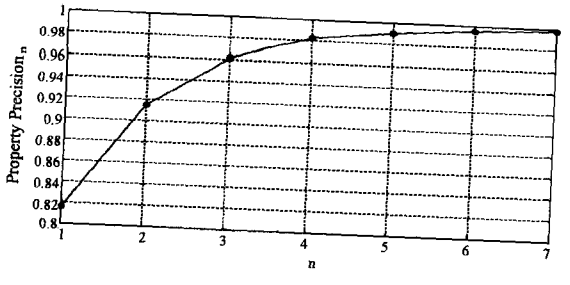

图 5 随 n 变化的 Property Precision$_n$

在上面的计算公式中，词义按照相关度来排序。若目标词的语义解释在最高的 n 个词义内，则认为是正确的。在半自动的本体澄清的过程中，用户能够从最高的 n 个候选词义中选择正确的语义解释。从实验结果图 4 和 5 可看出，concept precision$_3$ 和 property precision$_4$ 能够达到 98%的精度，从而证明了本体澄清算法在半自动过程中的有效性。

结束语 为了提高本体的质量，用 WordNet 中的词义表示本体的元素。本文陈述了考虑本体结构和被标注文档自动对本体元素进行语义消歧的本体澄清过程，实验证明了该方法的有效性。未来的工作分为两方面：首先计划开发本体澄清的工具，其次将调查澄清后的本体给基于本体的应用所能带来的好处。

参 考 文 献

[1] Ushold M, Gruninger M. Ontologies: Principles, methods and applications. The Knowledge Engineering Review, 1996, 11(2): 93-136

[2] Guarino N. Formal ontology and information systems // Proc. of the 1st Int'l Conf. on Formal Ontologies in Information Systems (FOIS98). Trento, Italy, IOS Press, 1998: 3-15

[3] Fellbaum C. Wordnet: An Electronic Lexical Database. Cambridge: MIT Press, 1998

[4] Missikoff M, Navigli R, Velardi P. Integrated approach to Web ontology learning and engineering. IEEE Computer, 2002, 35 (11): 60-63

[5] NavigliR, Velardi P, Gangemi A. Ontology learning and its application to automated terminology translation. IEEE Intelligent Systems, 2003, 18 (1): 22-31

[6] Banerjee S, Pedersen T. Extended gloss overlaps as a measure of semantic relatedness // Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence. Acapulco, 2003: 805-810

[7] Pedersen T, Banerjee S, Patwardhan S. Maximizing Semantic Relatedness top erform Word Sense Disambiguation. research report umsi 2005/25. Supercomputing Institute, University of Minnesota, 2005

[8] Sleator D, Temperley D. Parsing English with a Link Grammar. technical report. CMU-CS-91-196. Carnegie Mellon University, 1991