广义邻域关系下不完备混合决策系统的约简

徐久成 张灵均 孙 林 李双群
(河南师范大学计算机与信息工程学院 新乡 453002)

摘 要 为了能够直接处理不完备的、数值和符号混合的数据,对相容关系和相对邻域关系进行广义化表示,提出一种新的广义邻域关系。在广义邻域关系下,基于信息熵提出一种适用于不完备混合决策系统的条件熵,并证明基于该条件熵的属性重要性包含基于正区域的属性重要性,进而构造基于该条件熵的启发式属性约简算法。采用UCI数据库中6组混合型属性数据集进行仿真实验,通过对比约简后的属性数目、分类精度和运行时间,验证了该约简算法比同类型的其它算法更准确有效。

关键词 不完备混合决策系统,广义邻域关系,粗糙集,条件熵
中图法分类号 TP18 **文献标识码** A

Reduction in Incomplete Hybrid Decision System Based on Generalized Neighborhood Relationship

XU Jiu-cheng ZHANG Ling-jun SUN Lin LI Shuang-qun
(College of Computer & Information Engineering, Henan Normal University, Xinxiang 453002, China)

Abstract In order to deal with the incomplete, symbol and numeric hybrid data directly, a new kind of generalized neighborhood relationship was constructed by combining with relative neighborhood relationship and tolerance relationship. Under the general neighborhood relationship, the conditional entropy used for incomplete hybrid decision system was defined on the basis of information entropy. It was proved that the attibute significance of the condition entropy contains that of the positive regions in this paper. And then the reduction algorithm based on conditional entropy of incomplete hybrid decision system was constructed. The experiments on six hybrid attribute UCI datasets were made, and the proposed method and the similar methods were compared in aspects of feature gene number, classification accuracy and run-time. The results show that the method of feature gene selection based on the proposed extended rough set model is effective.

Keywords Incomplete hybrid decision system, Generalized neighborhood relation, Rough set, Conditional entropy

1 引言

粗糙集理论[1]是一种处理不确定信息的有效工具,目前已广泛应用于机器学习、模式识别、数据挖掘等领域。属性约简是数据挖掘的前提,也是粗糙集理论的核心内容。经典的粗糙集理论基于等价关系,其研究对象是完备的符号型信息系统或决策系统,对于不完备的和数值型的决策系统,经典粗糙集理论无法直接处理。因此,寻求基于扩展粗糙集模型的快速高效约简算法成为近年来粗糙集理论的研究热点。

目前,针对不完备决策系统约简的主要研究成果有:Kryszkiewicz[2]提出了基于容差关系的广义决策约简的辨别矩阵方法;周献中[3]在不完备决策系统中构造了分配序约简;管延勇等人[4]使用最大全相容类技术获取了不完备信息系统中的最优可信规则。针对数值、符号混合型决策系统约简的主要研究成果有:T. Y. Lin[5]提出了邻域粗糙集的基本概念;Y. Y. Yao及吴伟志等人[6,7]研究了邻域系统的基本性质;胡清华等人利用度量空间的邻域构造了邻域粗糙集模型[8,9],扩展了经典粗糙集模型,可直接处理数值型属性。以上研究是在单独考虑不完备和数值型决策系统的基础上进行的,而同时对不完备和数值型混合决策系统的研究并不多见。赵佰亭等人提出一种广义邻域粗糙集模型用于不完备混合决策系统约简[10,11],重点研究同时包含丢失型和遗漏型的不完备决策系统,对不完备数值型属性的约简算法改进较少,只是简单融合了邻域关系下的基于正域的属性约简算法[8]。目前,基于邻域粗糙集模型的约简算法大多基于正区域[8,9]和区分矩阵[12,13]。基于正区域的约简依据集合包含关系确定知识的不确定性程度,其本质含义不易被理解。因此,一些学者利用信息熵来描述知识的不确定性程度。苗夺谦等[14]讨论了知识粗糙性与信息熵的关系,得到偏序关系和粗糙熵的单调性相反的结论。王国胤等[15]证明了在不一致决策系统中[16],基于条件熵的属性重要性定义包含基于正区域的属性重要性定义,但是在等价关系下进行讨论的。黄兵等[17]提出基于一

到稿日期:2012-06-09 返修日期:2012-10-13 本文受国家自然科学基金(60873104,61040037),河南省科技攻关重点项目(112102210194),河南省教育厅自然科学基金(2008B520019)资助。

徐久成(1964—),男,博士,教授,硕士生导师,主要研究方向为粗糙集理论、粒计算、数据挖掘等;张灵均 女,硕士生,主要研究方向为粗糙集理论、数据挖掘等;孙 林 男,博士生,主要研究方向为粗糙集理论、粒计算等;李双群 男,硕士生,副教授,主要研究方向为粒计算、图像处理等。

般二元关系的知识粗糙熵和粗集粗糙熵，但仅研究了信息系统的约简。目前，对基于条件熵的不完备混合决策系统约简的研究较少。

针对上述讨论，本文提出一种新的广义邻域关系，对相对邻域关系和相容关系进行了广义化表示，得到的广义邻域粗糙集模型不仅能够处理完备的混合型决策系统，也能处理不完备的混合型决策系统。提出广义邻域关系下的条件熵，并给出其属性重要性定义。在广义邻域关系下，通过讨论基于条件熵的属性重要性和基于正区域的属性重要性[8]，得出前者定义包含后者定义，适用范围更广，最后构造基于条件熵的启发式约简算法。

2 基于不完备混合型决策系统的广义邻域关系

给定决策系统 $T=\langle U,C\cup D,V,f\rangle$，其中 $U=\{x_1,x_2,\cdots,x_n\}$ 表示非空有限样本集；$C\cap D=\varnothing$，C 和 D 分别为条件属性集和决策属性集，如果 $D=\varnothing$，则决策系统转换为信息系统；$V=\bigcup\limits_{a\in C\cup D}V_a$ 为属性值域，其中 V_a 为属性 a 的值域；$x_i(a)$ 为样本 x_i 在属性 a 上的取值。V 如果既包含连续数值数据，又包含离散符号数据，则称为混合决策系统 $MT=\langle U,C\cup D,V,f\rangle$。如果混合决策系统中至少存在一个属性 $a\in C$ 使得 V_a 含有空值，则称为不完备的混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$。不完备混合决策系统中，用符号 * 表示遗漏的属性值。

目前对粗糙集扩展模型的研究大多基于以下两类表：不完备的符号型决策系统，数值和符号型混合决策系统。基于相容关系的粗糙集模型可以直接处理不完备的数据，相对邻域粗糙集模型可以直接处理数值型和符合型混合数据，融合相对邻域关系和相容关系的优势，不仅克服了经典粗糙集只能直接处理符号型属性的局限性，还可直接处理不完备决策系统。本文对相容关系和相对邻域关系[19]进行广义化表示，并提出一种新的广义邻域关系。

定义 1[18] 给定不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$a\in C\cup D$，$\forall x_i,x_j\in U(i,j=1,2,\cdots,n,i\neq j)$，$x_i$ 和 x_j 在属性集 a 上的相异度定义为：

$$d_a(x_i,x_j)=\frac{|x_i(a)-x_j(a)|}{\max_{1\leqslant k\leqslant n}\{x_k(a)\}-\min_{1\leqslant k\leqslant n}\{x_k(a)\}}$$

定义 2[18] 给定不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$B\subseteq C\cup D$，$\forall x_i,x_j\in U(i,j=1,2,\cdots,n,i\neq j)$，$x_i$ 和 x_j 在属性集 B 上的相异度定义为：

$$d_B(x_i,x_j)=\frac{\sum\limits_{a\in B}d(x_i(a),x_j(a))}{|B|}$$

定义 3 给定不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$B\subseteq C\cup D$，$\forall x_i,x_j\in U(i,j=1,2,\cdots,n,i\neq j)$，广义邻域关系 TN_B 定义为：

$$TN_B=\{(x_i,x_j)\in U^2\mid\forall a\in B,d_B(x_i,x_j)\leqslant\frac{1}{H}\bigcup x_i(a)=x_j(a)\bigcup x_j(a)=*\bigcup x_j(a)=*\}$$

式中，$H(H\geqslant 1)$ 为量化级数[18]。

定义 4 给定不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$B\subseteq C\cup D$，$\forall x_i\in U$，样本 x_i 在属性子集 B 上的广义邻域定义为：

$$\omega_B(x_i)=\{x_j\in U\mid(x_i,x_j)\in TN_B\}$$

$\frac{U}{TN_B}=\{\omega_B(x_i)\mid x_i\in U\}$ 是由 B 生成的一个分类，$\omega_B(x_i)$ 称为广义邻域粒。广义邻域关系是邻域关系和相容关系的扩展，既可处理完备的混合型属性，又可处理不完备混合型属性。

性质 1 给定不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$B\subseteq C\cup D$，$\forall x_i\in U$，样本 x_i 在属性子集 B 上的广义 ω 邻域满足：

(1) $\omega_B(x_i)\neq\varnothing$；

(2) $\bigcup\limits_{i=1}^{|U|}\omega_B(x_i)=U$；

(3) $x_j\in\omega_B(x_i)\Leftrightarrow x_i\in\omega_B(x_j)$。

3 广义邻域关系下的两种观点的属性重要性比较

3.1 广义邻域关系下基于条件熵的属性重要性

引入基于等价关系的信息熵，提出基于广义邻域关系的信息熵，进而定义条件熵及其属性重要性。基于条件熵的属性重要性，考虑属性对论域中的不确定分类样本的影响[15]，如果添加一个属性，不改变信息的不确定性，那么这个属性关于属性集的重要性为 0。

定义 5[15] 信息系统中，给定论域 U，属性集 B 对 U 的划分为 $B=\{X_1,X_2,\cdots,X_k\}$，则 B 的信息熵定义为：

$$H(B)=-\sum_{i=1}^{k}\frac{|X_i|}{|U|}\log_2\frac{|X_i|}{|U|}$$

信息熵是利用等价关系对论域的划分进行定义。如果将等价关系放宽为广义邻域关系，邻域类对论域的划分变成了覆盖，则利用分块大小衡量信息量大小就不再恰当。将每一个样本单独看待，其所在的等价类则可以看作它的邻域，那么可以用一个样本在所有样本的邻域中出现的次数来定义信息熵[17]。而广义邻域关系是对称的，那么在广义邻域关系下，可以利用样本的邻域大小来定义信息熵。决策系统可以看作信息表的特殊情况，因此可以在信息熵上的基础上通过属性的并来构造决策系统的条件熵[19]。

定义 6 给定不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$B\subseteq C\cup D$，$\omega_B(x_i)$ 表示在 B 下样本 x_i 生成的广义邻域，则在广义邻域关系下 B 的信息熵定义为：

$$H(B)=-\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|\delta_B(x_i)|}{|U|}$$

性质 2 给定不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$P,Q\subseteq C\cup D$，$U/TN_P=\{\omega_P(x_1),\omega_P(x_2),\cdots,\omega_P(x_{|U|})\}$ 和 $U/TN_Q=\{\omega_Q(x_1),\omega_Q(x_2),\cdots,\omega_Q(x_{|U|})\}$ 分别表示 P,Q 对论域构成的两个分类，则这两个分类的交运算满足：$U/TN_P\cap U/TN_Q=U/TN_{Q\cup P}$。

定义 7 给定不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$P,Q\subseteq C\cup D$，则属性子集 P 和属性子集 Q 的条件熵定义为：

$$H(P\cup Q)=-\sum_{i=1}^{|U|}\sum_{j=1}^{|U|}\frac{1}{|U|}\log_2\frac{|\omega_P(x_i)\bigcap\omega_Q(x_j)|}{|U|}$$

依据性质 2 对上式进行如下推导：

$U/TN_P\cap U/TN_Q=\{\omega_P(x_1)\cap\omega_Q(x_1),\omega_P(x_1)\cap\omega_Q(x_2),\cdots,\omega_P(x_{|U|})\cap\omega_Q(x_{|U|-1}),\omega_P(x_{|U|})\cap\omega_Q(x_{|U|})\}=U/TN_{Q\cup P}=\{\omega_{Q\cup P}(x_1),\omega_{Q\cup P}(x_2),\cdots,\omega_{Q\cup P}(x_{|U|})\}=U/TN_P\cap$

$U/TN_Q=\{\omega_P(x_1)\cap\omega_Q(x_1),\omega_P(x_2)\cap\omega_Q(x_2),\cdots,\omega_P(x_{|U|})\cap\omega_Q(x_{|U|})\}$。因此属性子集 P 和属性子集 Q 的条件熵定义可以简化为：

$$H(P\cup Q)=-\sum_{i=1}^{|U|}\sum_{j=1}^{|U|}\frac{1}{|U|}\log_2\frac{|\omega_P(x_i)\cap\omega_Q(x_j)|}{|U|}$$
$$=\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|\omega_P(x_i)\cap\omega_Q(x_i)|}{|U|}$$

定义 8 给定不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$B\subseteq C\cup D$，则属性子集 B 和决策属性 D 的条件熵定义为：

$$H(D\cup B)=-\sum_{i=1}^{|U|}\sum_{j=1}^{|U|}\frac{1}{|U|}\log_2\frac{|\omega_D(x_i)\cap\omega_B(x_j)|}{|U|}$$
$$=\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|\omega_D(x_i)\cap\omega_B(x_i)|}{|U|}$$

定义 9 给定不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$P,Q\subseteq C\cup D$，如果 $\forall\omega_P(x_i)\in U/TN_P\Rightarrow\exists\omega_Q(x_i)\in U/TN_Q$，满足 $\omega_P(x_i)\subseteq\omega_Q(x_i)$，那么就称分类 U/TN_P 比 U/TN_Q 细，记作 $U/TN_P\prec U/TN_Q$。

定理 1 给定不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$M,N\subseteq C\cup D$。如果 $U/TN_M\prec U/TN_N$，则 $H(D\cup N)<H(D\cup M)$。

证明：$U/TN_M=\{\omega_M(x_1),\omega_M(x_2),\cdots,\omega_M(x_{|U|})\}$

$U/TN_N=\{\omega_N(x_1),\omega_N(x_2),\cdots,\omega_N(x_{|U|})\}$

$U/TN_D=\{\omega_D(x_1),\omega_D(x_2),\cdots,\omega_D(x_{|U|})\}$

如果 $U/TN_M\prec U/TN_N$，则对 $\forall x_i\in U$，满足 $\omega_M(x_i)\subset\omega_N(x_i)$，且 $\omega_D(x_i)\cap\omega_M(x_i)\subset\omega_D(x_i)\cap\omega_N(x_i)$。即对 $\forall x_i\in U$，有 $1\leqslant|\omega_M(x_i)|\leqslant|\omega_N(x_i)|$ 且 $1\leqslant|\omega_D(x_i)\cap\omega_M(x_i)|\leqslant|\omega_D(x_i)\cap\omega_N(x_i)|$。因此，

$$H(D\cup N)-H(D\cup M)$$
$$=-\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|\omega_D(x_i)\cap\omega_N(x_i)|}{|U|}+\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|\omega_D(x_i)\cap\omega_M(x_i)|}{|U|}$$
$$=\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|\omega_D(x_i)\cap\omega_M(x_i)|}{|\omega_D(x_i)\cap\omega_N(x_i)|}<0$$

推论 1 给定不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$B\subseteq C\cup D$，$a\subseteq B$，则 $H(D\cup B)<H(D\cup B\cup\{a\})$。

定义 10 给定不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$B\subseteq C\cup D$，则任意属性 $a\in C-B$ 相对于决策属性 D 的重要性定义为：

$$SGF_H(a,B,D)=H(D\cup B\cup\{a\})-H(D\cup B)$$

3.2 广义邻域关系下基于正区域的属性重要性

正区域的大小反映各类样本的可分离程度，正区域越大，表明确定分类样本和不确定分类样本的重叠区域越小[8]。基于正区域的属性重要性考虑对确定分类样本的影响。如果添加一个属性，不改变确定分类样本，则该属性相对于属性集的重要性为 0。

定义 11 给定不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$B\subseteq C\cup D$，$X\subseteq U$，则等价类 X 在条件属性子集 B 上的广义邻域下近似和上近似分别定义为：

$$\underline{B}(X)=\{x_i\mid\omega_B(x_i)\subseteq X,x_i\in U\}$$
$$\overline{B}(X)=\{x_i\mid\omega_B(x_i)\cap X\neq\varnothing,x_i\in U\}$$

定义 12 给定不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$B\subseteq C\cup D$，$X\subseteq U$，决策属性 D 将 U 划分成个 M 等价类：$X_1,X_2,\cdots,X_M$，则 D 在条件属性子集 B 上的广义邻域下近似和上近似分别定义为：

$$\underline{B}(D)=\bigcup_{i=1}^{M}\underline{B}(X_i),\overline{B}(D)=\bigcup_{i=1}^{M}\overline{B}(X_i)$$

决策属性 D 的下近似也称决策正域，记作 $POS_B(D)$。

定义 13 给定不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$B\subseteq C\cup D$，则决策属性 D 对条件属性子集 B 的依赖性定义为：

$$\gamma_B(D)=|POS_B(D)|/|U|$$

定义 14 给定不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$B\subseteq C\cup D$，则任意属性 $a\in C-B$ 相对于决策属性 D 的重要性定义为：

$$SGF_\gamma(a,B,D)=\gamma_{B\cup\{a\}}(D)-\gamma_B(D)$$

3.3 两种观点的属性重要性比较

在等价关系下，如果添加一个属性，不改变原本确定分类和不确定分类的样本，仅改变不确定性，则该属性基于正区域的重要性为 0，但基于条件熵的重要性不为 0，因此基于条件熵的属性重要性包含基于正区域的属性重要性[15]。如果不在等价关系下，这种包含关系还存在吗？下面我们讨论在广义邻域关系下，这两种观点的属性重要性的关系。

例 1 设不完备混合决策系统 $IMT=\langle U,C\cup D,V,f\rangle$，$B\subseteq C\cup D$，则广义邻域覆盖 $U/TN_D=\{\{x_1,x_2\},\{x_2,x_1\},\{x_3,x_4\},\{x_4,x_3\}\}$，$U/TN_B=\{\{x_1,x_2\},\{x_2,x_1,x_3\},\{x_3,x_2,x_4\},\{x_4,x_3\}\}$，$U/TN_{B\cup\{a\}}=\{\{x_1\},\{x_2,x_3\},\{x_3,x_2,x_4\},\{x_4,x_3\}\}$。属性 a 相对于 D 的基于正区域的重要性 $SGF_\gamma(a,B,D)=\gamma_{B\cup\{a\}}(D)-\gamma_B(D)=0$。属性 a 相对于 D 的基于条件熵的重要性 $SGF_H(a,B,D)=H(D\cup B\cup\{a\})-H(D\cup B)=2-1=1$。

从例 1 可以看出，该属性的基于正区域的重要性为 0，但基于条件熵的重要性不为 0，即后者的定义是包含前者的。

定理 2 给定不完备混合决策系统 $IMT=\langle U,A,V,f\rangle$，$A=C\cup D$，$B\subseteq C$，$a\in C-B$，如果 $SGF_H(a,B,D)=0$，则 $SGF_\gamma(a,B,D)=0$。

证明：假设 $U/TN_{B\cup\{a\}}=\{\omega_{B\cup\{a\}}(x_1),\omega_{B\cup\{a\}}(x_2),\cdots,\omega_{B\cup\{a\}}(x_{|U|})\}$，$U/TN_B$ 可以由 $U/TN_{B\cup\{a\}}$ 中某两个邻域的合并得到。例如将 $\omega_{B\cup\{a\}}(x_i)$ 和 $\omega_{B\cup\{a\}}(x_j)$ 合并，如果在 $\omega_{B\cup\{a\}}(x_i)$ 中添加了 x_j，由于邻域关系的对称性，则 $\omega_{B\cup\{a\}}(x_j)$ 中也要相应地添加 x_i。因此可以得到，$U/TN_B=\{\omega_{B\cup\{a\}}(x_1),\omega_{B\cup\{a\}}(x_2),\cdots,\omega_{B\cup\{a\}}(x_i)\cup\{x_j\}$，则属性子集 $B\cup\{a\}$ 和 B 的条件熵分别为：

$$H(D\cup B\cup\{a\})=-\frac{1}{|U|}\left(\log_2\frac{|\omega_{B\cup\{a\}}(x_1)\cap\omega_D(x_1)|}{|U|}+\cdots+\log_2\frac{|\omega_{B\cup\{a\}}(x_i)\cap\omega_D(x_i)|}{|U|}+\log_2\frac{|\omega_{B\cup\{a\}}(x_j)\cap\omega_D(x_j)|}{|U|}+\cdots+\log_2\frac{|\omega_{B\cup\{a\}}(x_{|U|})\cap\omega_D(x_{|U|})|}{|U|}\right.$$

$$H(D\cup B)=-\frac{1}{|U|}\left(\log_2\frac{|\omega_{B\cup\{a\}}(x_1)\cap\omega_D(x_1)|}{|U|}+\cdots+\log_2\frac{|\omega_{B\cup\{a\}}(x_i)\cup\{x_j\}\cap\omega_D(x_i)|}{|U|}+\log_2\frac{|\omega_{B\cup\{a\}}(x_j)\cup\{x_i\}\cap\omega_D(x_j)|}{|U|}+\cdots+\log_2\right.$$

$$\frac{|\omega_{B\cup\{a\}}(x_{|U|})\cap\omega_D(x_{|U|})|}{|U|}$$

如果 $SGF_H(a,B,D)=0$，则 $H(D\cup B\cup\{a\})=H(D\cup B)$，进而得到：

$$\frac{|\omega_{B\cup\{a\}}(x_i)\cup\{x_j\}\cap\omega_D(x_i)|}{|\omega_{B\cup\{a\}}(x_i)\cap\omega_D(x_i)|}=\frac{|\omega_{B\cup\{a\}}(x_j)\cup\{x_i\}\cap\omega_D(x_j)|}{|\omega_{B\cup\{a\}}(x_j)\cap\omega_D(x_j)|}$$

由此可看出，样本邻域合并后，条件属性构成的样本邻域对于决策属性的隶属度不会发生变化，即 $SGF_\gamma(a,B,D)=0$。

由此看出，属性基于条件熵的重要性为 0 时，其基于正区域的重要性也为 0，但反过来未必满足。也即基于条件熵的属性重要性比基于正区域的属性重要性适用范围更广。

4 基于条件熵的属性约简算法

依据定理 2 的结论，利用条件熵为启发因子构造启发式的前向约简算法。为达到提高算法效率的目的，该算法在以下两方面对算法进行改进：通过属性的重要性降序排序，首先对条件熵最大的属性计算属性重要性；求最大的 $SGF_H(a,B,D)$，实际上是求最大的 $H(D\cup B\cup\{a\})$，因为每轮都可以省去计算 $H(D\cup B)$ 的时间，该算法的时间复杂度为 $O(|C|^2|U|\log|U|)$，且考虑了约简的完备性。

算法 1 基于条件熵的启发式属性约简算法

输入：$MT=\langle U,A,V,f\rangle$，$A=C\cup D$，$D=\{d\}$，$C=\{c_1,c_2,\cdots,c_{|C|}\}$；

输出：约简 B；

Step 1 计算每个样本 $x_i\in U$ 的 $\delta_C(x_i)$ 和 $\delta_D(x_i)$，利用快速排序求出 $H(D\cup C)$；

Step 2 依据 $H(D\cup\{c_i\})$ 的值，将 $c_i\in C$ 从大到小排序组成新的序列 $S=\{a_1,a_2,\cdots,a_{|C|}\}$；

Step 3 令 $B=\varnothing$；

Step 4 对任意 $a_i\in S-B$ 计算

$SGF_H(a_i,B,D)=H(D\cup B\cup\{a_i\})-H(D\cup B)$

Step 5 求 $\max(SGF_H(a_i,B,D))$，等价于求：$H=\{a_i\in S-B|\max(H(D\cup B\cup\{a_i\}))\}$，if $H\neq 1$ then 选择 $a_i\in H$ 满足 $\min(|U/TN_{B\cup\{a_i\}}|)$；

Step 6 $B=B\cup\{a_i\}$

Step 7 if $H(D\cup B)\neq H(D\cup C)$

go to Step 5

else

//此处考虑算法的完备性

令 $B=S\cap B=\{a_1,a_2,\cdots,a_{|B|}\}$

for ($i=1;i\leqslant|B|;i++$)

$a_i\in B$

$B=B-\{a_i\}$

if $H(D\cup B)\neq H(D\cup C)$

$B=B\cup\{a_i\}$

Step 8 返回约简 B。

5 实验结果及分析

为验证模型的合理性和算法的有效性，采用 6 组 UCI 机器学习数据库中的数据进行了约简和分类的仿真实验。为了充分验证反正效果，在数据中随机删除了部分样本的邻域值，增大了样本的不完备性。数据约简程度是比较不同约简算法的一个指标，但对于分类问题而言，更重要的是选择的属性不能显著降低分类能力。一个优秀的约简算法要在保持或提高分类能力的基础上，尽可能地减少分类建模所需的属性数目。表 1 显示了数据集的属性约简的测试结果，其中的 r、a 和 t 分别定义了约简后的属性数目、分类精度和运行时间。从图 1 中可以直观地看出，本文方法（标记为 F2）较文献[8]中的方法（标记为 F1）在这 3 方面都有更好的表现。

表 1 属性约简的测试结果

编号	数据集	样本数	属性数	F1			F2		
				r	a	t	r	a	t
1	Zoo	101	17	5	92.39	0.09	5	93.56	0.01
2	Iono	351	34	16	95.17	0.3	15	95.78	0.09
3	Diabetes	768	20	8	77.47	2	7	81.23	0.4
4	Image	2310	19	8	73.4	10	8	80.42	2
5	Abalone	4177	8	1	16.55	16	1	17.86	4
6	Mushroom	8124	22	5	96.65	23	4	96.50	6

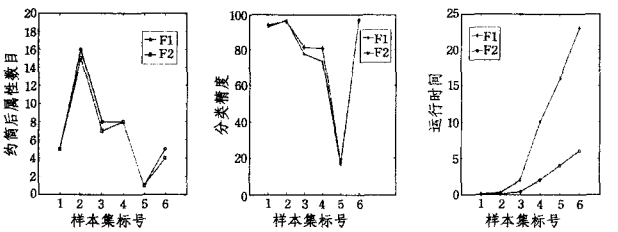

图 1 约简后属性数目、分类精度和运行时间对比图

结束语 本文提出了一种新的广义邻域关系，其适用于针对不完备的、数值和符号型混合的决策系统，定义了广义邻域关系下的条件熵及其属性重要性，并证明了基于条件熵的属性重要性定义比基于正区域的属性重要性定义适用范围更广，最后构建了基于条件熵的启发式约简算法。通过仿真实验验证了该方法是准确有效的。

参 考 文 献

[1] Pawlak Z. Rough sets [J]. International Journal of Information Computer Science, 1982, 11(5): 341-356

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

[3] 周献中，黄兵. 基于粗集的不完备信息系统属性约简[J]. 南京理工大学学报，2003，27(5)：630-635

[4] Guan Yan-yong, Wang Hong-kai. Set-valued information systems [J]. International Journal of Information Sciences, 2006, 176(17): 2507-2525

[5] Lin T Y. Granular Computing on binary relations I: Data mining and neighborhood systems [C]// Skoworn A, Polkowshi L, eds. Proc. of the Rough Sets in Knowledge Discovery. Physica-Verlag, 1998: 107-121

[6] Yao Y Y. Relational interpretation of neighborhood operators and rough set approximation operators [J]. Information Scien-ces, 1998, 111(198): 239-259

[7] Wu Wei-zhi. Neighborhood operator systems and approximations [J]. Information Sciences, 2002, 144(1-4): 201-217

[8] 胡清华，于达仁，谢宗霞. 基于邻域粒化和粗糙逼近的数值属性约简[J]. 软件学报，2008，19(3)：640-649

[9] Hu Qing-hua, Yu Da-ren, Xie Zong-xia. Neighborhood classifiers [J]. Expert systems with applications, 2008, 34(2): 866-876

[10] 赵佰亭，陈希军，曾庆双. 广义不完备混合决策系统的知识约简[J]. 四川大学学报，2009，41(6)：177-182

[11] Zhao Bai-ting, Chen Xi-jun, Zeng Qing-shuang. Incomplete hu-

brid attributes reduction based on neighborhood granulation and approximation [C]// 2009 IEEE International Conference on Mechatronics and Automation. 2009,2:2066-2071

[12] 林俊伟,叶东毅. 基于邻域辨识矩阵的属性约简增量式算法[J]. 计算机应用,2009,29(06):119-121

[13] 舒文豪,徐章艳. 不完备决策表的差别矩阵属性约简算法[J]. 计算机工程于应用,2011,47(24):103-105

[14] 苗夺谦,王珏. 粗糙集理论中知识粗糙性与信息熵关系的讨论[J]. 模式识别与人工智能,1998,11(3):34-40

[15] 王国胤,于洪,杨大春. 基于条件信息熵的决策表约简[J]. 计算机学报,2002,25(7):759-766

[16] Qian Yu-hua, Liang Ji-ye, Li De-yu. Approximation reduction in-consistent incomplete decision tables [J]. Knowledge-Based Systems, 2010, 23: 427-433

[17] 黄兵,周献中,史迎春. 基于一般二元关系的知识粗糙熵与粗集粗糙熵[J]. 系统工程理论与实践,2004,24(1):93-96

[18] Xu Jiu-cheng, Zhang Ling-jun, Sun Lin, et al. Gene Selection Algorithm Combining ReliefF and Relative Neighborhood Rough Set[C]// IEEE International Conference on Granular Computing. 2011: 745-749

[19] Sun Lin, Xu Jiu-cheng, Xue Zhan'ao, et al. Rough entropy-based feature selection and its application [J]. Journal of Information and Computational Science, 2011, 8(9): 1525-1532

(上接第 220 页)

表 2　3 个机器人进行 500 次实验仿真的平均数据

		平均路径长度(cm)	平均规划时间(s)
ICCEAA	r_1	27.6012	16.1806
	r_2	32.2931	21.7778
	r_3	28.3429	16.2504
APF	r_1	28.7024	15.8056
	r_2	41.2790	31.2083
	r_3	29.6243	16.2639

结束语　本文研究了一种在动态环境下的协作多机器人路径规划算法,采用集中式与分布式相结合的混合式作为多机器人系统的体系结构,并在体系结构中通过融合免疫协同进化算法与 APF 算法,提出 ICCEAA 算法解决全局路径规划与局部路径规划问题。ICCEAA 算法的优越性主要体现在:1. 通过免疫协同进化算法解决全局路径规划,既能够实现机器人之间的协调合作,也能保证机器人全局协调能力,避免陷入局部最优,弥补了 APF 算法的缺陷;2. 采用 APF 算法的局部路径规划,能够实现实时避障,机器人拥有较好的灵活性,也相应弥补了全局路径规划实时避障的不足。仿真实验也表明,ICCEAA 算法在动态环境下进行路径规划有较好的优越性,能够合理地分配任务,在较短的时间内完成任务。

参 考 文 献

[1] Fukuda T, Nakagawa S. A dynamically reconfigurable robotic system(Concept of a system and optimal configurations)[C]// Industrial Application of Robotics & Machine Vision. 1987: 588-595

[2] Parker L E. Multiple mobile robot systems[M]. Springer Handbook of Robotics. 2008: 921-941

[3] Zhu A, Yang S X. Path planning of multi-robot systems with cooperation[C]// IEEE International Symposium on Computational Intelligence in Robotics and Automation. 2003: 1028-1033

[4] Parker L E. ALLIANCE: An architecture for fault tolerant, cooperative control of heterogeneous mobile robots[C]// Proceedings of the IEEE/RSJ/GI International Conference on Advanced Robetic Systmes and the Real World. 1994: 776-783

[5] 邵杰,杨静宇. 基于 LCS 的多机器人路径规划控制体系结构[J]. 微电子学与计算机,2010,27(11)

[6] Chakraborty J, Mukhopadhyay S. A robust cooperative multi-robot path-planning in noisy environment[C]// IEEE International Conference on Industrial and Information Systems. 2010: 626-631

[7] Cai Z, Chen B, Wang L, et al. The progress of cooperative technology for heterogeneous multiple mobile robots [J]. CAAI Transcations on Intelligent Systems, 2007, 2(3): 1-7

[8] 刘丽珏. 免疫进化算法及其在多机器人协作中的应用研究[D]. 长沙:中南大学,2008

[9] Sheng J, He G, Guo W, et al. An improved artificial potential field algorithm for virtual human path planning[J]. Entertainment for Education. Digital Techniques and Systems, 2010: 592-601

[10] Sabattini L, Secchi C, Fantuzzi C. Arbitrarily shaped formations of mobile robots: artificial potential fields and coordinate transformation[J]. autonomous robots, 2011, 30(4): 385-397

[11] Parker L E. Path Planning and Motion Coordination in Multiple Mobile Robot Teams[M]. in-chief: Encyclopedia of Complexity and System Science, Springer, The Netherlands, 2009

[12] Wang Mei, Wu Tie-jun. Cooperative co-evolution based distributed path planning of multiple mobile robots[J]. Journal of Zhejiang University-Science A, 2005, 6(7): 697-706

[13] 崔益安,蔡自兴,李满晨. 自组分层式多机器人体系结构[J]. 小型微型计算机系统,2008,29(7):1263-1267

[14] Ding Ying-ying, He Yan, Jiang Jing-ping. Multi-robot cooperation method based on the ant algorithm[C]// Swarm Intelligence Symposium, 2003. 2003: 14

[15] Liu Li-jue. Immune Evolutionary Algorithms and Their APPlications in CooPrative Multi- Robot System[D]. Changsha: Central South University, 2008

[16] Ge S S, Cui Y J. New potential functions for mobile robot path planning[J]. IEEE Transactions on Robotics and Automation, 2000, 16(5): 615-620

[17] Li Qing, Xu Yin-mei, Zhang De-zheng, et al. Global path planning method for mobile robots based on the particle swam algorithm[J]. Journal of University of Science and technology Beijing, 2010, 32(3): 123-128

[18] Yang D, Chen J, Matsumoto N, et al. Multi-robot Path Planning Based on Cooperative Co-evolution and Adaptive CGA[C]// IEEE international conference on Intelligent Agent Technology. 2006: 547-550