一种基于模糊粗糙集知识获取方法

王基一[1]　顾沈明[2]

（浙江师范大学计算机科学与工程学院　浙江金华321004）[1]　（浙江海洋学院信息学院　浙江舟山316004）[2]

摘　要　本文介绍了粗糙集和模糊粗糙集的上下近似。并且利用模糊粗糙上下近似算子，论述了在不完备模糊信息系统中知识获取的一种方法。应用这种方法能够让隐藏在不完备模糊信息系统中的知识，以决策规则的形式表示出来。最后给出了一种实现算法和实例。

关键词　粗糙集，模糊粗糙集，信息系统，决策规则

A Method of Knowledge Acquisition Based on Fuzzy Rough Set

WANG Ji-Yi[1]　GU Shen-Ming[2]

(College of Computer Science and Engineer, Zhejiang Normal University, Jinhua321004)[1]

(College of Information, Zhejiang Ocean University, Zhoushan 316004)[2]

Abstract　This paper introduces the lower and upper approximations operator of rough sets and fuzzy sets. By using fuzzy rough lower and upper approximation operators, this paper provides a method of knowledge acquisition is incomplete fuzzy information systems. The knowledge hidden in the incomplete fuzzy information system can be expressed in the form of decision rules using the method. The algorithm of implementation is also discussed, and an example is given.

Keywords　Rough sets, Fuzzy rough sets, Information systems, Decision rule

粗糙集理论是波兰数学家Z. Pawlak提出的一种数学理论[1,2]，它为分析不完整、不确定性数据提供了新的思路。近年来，该理论被广泛应用于专家系统、模式识别、机器学习等领域，是一个新兴的研究热点。

应用粗糙集理论，可以把隐藏在信息系统中的知识，以规则的形式挖掘出来[3]，由于在实践中碰到许多信息系统是不完备的，本文采用模糊集的思想，研究了基于模糊粗糙集的知识获取方法。

1　粗糙集[1]

设U非空有限集，称为论域，R是U上二元等价关系，称(U,R)为Pawlak近似空间。R把论域U分成若干个等价类，记$[x]$为对象x所在的等价类。对于任意的$X\subseteq U$，X不一定能用R的等价类来描述，于是用下近似$\underline{R}(X)$和上近似$\overline{R}(X)$来描述：

$\underline{R}(X)=\{x\in U\mid [x]\subseteq X\}$,

$\overline{R}(X)=\{x\in U\mid [x] I\ X\neq\phi\}$

$(\underline{R}(X),\overline{R}(X))$称为$X$关于近似空间$(U,R)$的粗糙集。

2　模糊粗糙集[4,5]

设U和W为有限非空论域，R是从U到W的模糊关系，三元组(U,W,R)称为一般的模糊近似空间。任意的$X\in 2^W$，X关于近似空间(U,W,R)的下近似$\underline{R}(X)$和上近似$\overline{R}(X)$是U的模糊集，它们隶属函数定义为：对于任意的$x\in U$，

$\underline{R}(X)(x)=\bigwedge_{y\in W}[(1-R(x,y))\vee X(y)]=\min\{1-R(x,y)\mid y\notin X\},x\in U$

$\overline{R}(X)(x)=\bigvee_{y\in W}[R(x,y)\wedge X(y)]=\max\{R(x,y)\mid y\in X\},x\in U$

$(\underline{R}(X),\overline{R}(X))$称一般的模糊粗糙集。

3　信息系统

$S=(U,AT)$称是一个信息系统，U为有限非空对象集，称为论域或对象空间，AT为有限非空属性集；对于每一个属性$a\in AT$，任意的对象$x\in U$，有一个映射$a:x\rightarrow a(x)$，即对象x在属性a上的取值。若$a(x)$是唯一确定的，则称S是完备的信息系统。

但是许多信息系统中，一些属性值是不确定的，如有缺省值或知道部分值，此时对象x在属性a上的取值有多个可能，即$a(x)\subseteq V_a$，其中V_a是所有对象在属性a上的值域，则称S是不完备的信息系统[6,7]。

在不完备的信息系统中，如果对每一个$x\in U$和$a\in AT$，$a(x)$是模糊子集，则称S是不完备的模糊信息系统。对每一个$a\in AT$，可以在U上定义一个模糊关系R_a，它的隶属函数是：

$$R_a(x,y)=\overline{F}_a(a(x))(y)=\bigvee_{v\in V_a}(a(x)(v)\wedge a(y)(v)),x,y\in U$$

因此，一个属性子集$A\subseteq AT$可决定一个U上的模糊关系$R_A=\operatorname*{I}_{a\in A}R_a$，称为模糊相似关系。我们把对象$x$在属性子集$A$上的模糊相似邻域记作$F_A(x)$，

$F_A(x)(y)=R_A(x,y),\quad y\in U$

$S=(U,AT)$是不完备模糊信息系统，给定$X\subseteq U$和$A\subseteq AT$，X的下近似$\underline{F_A}(X)$和上近似$\overline{F_A}(X)$是U的模糊子集，它们隶属函数为：

$$\underline{F_A}(X)(x)=\bigwedge_{y\in U}(X(y)\vee(1-R_A(x,y)))=\min\{1-R_A(x,y)\mid y\notin X\},x\in U$$

$\overline{F_A}(X)(x)=\bigvee_{y\in U}(X(y)\wedge R_A(x,y))=\max\{R_A(x,y)\mid y\in X\},x\in U$

在不完备模糊信息系统中，增加一个属性$\{d\}$，$d\notin AT$，其中d是完备属性，即$d(x)\in V_d$，$\forall x\in U$。则$DT=(U,AT\cup\{d\})$称为决策表，其中AT称为条件属性集，d称为决策属性。

隐藏在决策表中的知识，可以由以下规则的形式找出来：

$r:t\xrightarrow{c_A(x)}s$

其中：$t=\wedge(a,\bigvee_{v\in a_0(x)}v)$，$a\in A\subseteq AT$；$s=d(d,d(x))$，$x\in U$

注：$a_0(x)=\{x\in U\mid a(x)>0\}$，$c_A(x)$是规则的可信度，它的定义如下：

$$c_A(x)=\frac{\sum_{y\in[x]_d}F_A(x)(y)}{\sum_{y\in U}F_A(x)(y)}$$

公式中$[x]_d=\{u\in U\mid d(u)=d(x)\}$，$x\in U$。

当$c_A(x)=1$时，称规则$r:t\to s$是确定性的规则；当$0<c_A(x)<1$时，称规则$r:t\to s$是可能性的规则。

在输出规则之前进行约简。对于属性子集$A\subseteq AT$，给定$x\in U$，若：$F_A(x)=F_{AT}(x)$且$F_B(x)\neq F_A(x)$，$\forall B\subset A$，则称A是针对x的约简，记作$red(x)$。对于属性子集$A\subseteq AT$，若：$R_A=R_{AT}$且$R_B\neq R_A$，$\forall B\subset A$，则称A是系统属性集的约简。

4 知识获取算法

输入：不完全模糊信息系统决策表。假设系统决策表为$DT=(U,AT\cup\{d\})$，$U=\{x_1,x_2,\cdots,x_N\}$，N是对象的个数，$AT=\{a_1,a_2,\cdots,a_M\}$，$M$是属性集中属性的个数。

输出：确定性的规则和可能性的规则。

Step1：计算每一个属性a的R_a，

$$R_a=\begin{bmatrix}R_a(x_1,x_1)&R_a(x_1,x_2)&\cdots&R_a(x_1,x_M)\\R_a(x_2,x_1)&R_a(x_2,x_2)&\cdots&R_a(x_2,x_M)\\\cdots&\cdots&\cdots&\cdots\\R_a(x_M,x_1)&R_a(x_M,x_2)&\cdots&R_a(x_M,x_M)\end{bmatrix}$$

显然这是一个$M\times M$实对称矩阵，且对角线元素为1，所以在计算时只需计算矩阵的下（或上）三角元素即可。

上述步骤完成后，我们得到了所有属性a的R_a（共计m个方阵）；

Step2：计算$R_{AT}=\bigcap_{a\in AT}R_a$；

Step3：把所有对象根据属性d分类，得到$[x]_d$

Step4：计算$red(x)$

Step5：输出$c_{AT}(x)=1$的确定性规则

Step6：输出$0<c_{AT}(x)<1$的可能性规则

5 一个实例

设一个不完备模糊信息系统S的决策表如表1所示。

其中 $U=\{x_1,x_2,x_3,x_4,x_5,x_6\}$，

$AT=\{a_1,a_2,a_3\}$，$V_{a_1}=\{r_1,r_2,r_3\}$，

$V_{a_2}=\{s_1,s_2,s_3\}$，$V_{a_3}=\{t_1,t_2,t_3\}$

对每一个属性a计算R_a。下列矩阵为属性a_1的R_{a_1}：

$$R_{a_1}=\begin{bmatrix}1&0.3&0&1&1&1\\0.3&1&1&0.3&0.3&1\\0&1&1&0&0.1&1\\1&0.3&0&1&0.4&1\\1&0.3&0.1&0.4&1&1\\1&1&1&1&1&1\end{bmatrix}$$

表1 一个信息系统S的决策表

U	a_1	a_2	a_3	d
x_1	$1/r_2+1/r_3$	$1/s_3$	$1/t_2$	Y
x_2	$1/r_1+0.3/r_2$	$1/s_2+0.4/s_3$	$1/t_1+0.5/t_2$	Y
x_3	$1/r_1$	$1/s_2$	$1/t_1+0.1/t_3$	Y
x_4	$1/r_2$	$1/s_1+0.6/s_2$	$0.7/t_1+1/t_3$	N
x_5	$0.1/r_1+0.4/r_2+1/r_3$	$1/s_1+1/s_2$	$1/t_1+1/t_3$	Y
x_6	$1/r_1+1/r_2+1/r_3$	$1/S_3$	$1/t_2+0.2/t_3$	N

共有三个属性，所以 m=2、3时，分别计算R_{a_2}和R_{a_3}

计算$R_{AT}=\bigcap_{a\in AT}R_a$，即$R_{a_1}$、$R_{a_2}$和$R_{a_3}$的交，结果如下：

$$R_{AT}=\begin{bmatrix}1&0.3&0&0&0&1\\0.3&1&1&0.3&0.3&0.4\\0&1&1&0&0.1&0\\0&0.3&0&1&0.4&0\\0&0.3&0.1&0.4&1&0\\1&0.4&0&0&0&1\end{bmatrix}$$

根据属性d，把所有对象分类，得到等价类$\{x_1,x_2,x_3,x_5\}$和$\{x_4,x_6\}$

计算$c_{AT}(x)$，结果如下：

$C_{AT}(x_1)=0.57$，$c_{AT}(x_2)=0.79$，$c_{AT}(x_3)=1$，$c_{AT}(x_4)=0.59$，$c_{AT}(x_5)=0.78$，$c_{AT}(x_6)=0.42$

计算$red(x)$，可得到的一种结果：

$red(x)=\{a_1,a_2\}$，$x\neq x_6$

$red(x_6)=\{a_2\}$

输出$c_{AT}(x)=1$的确定性规则：

规则1：IF $a_1=r_1$ AND $a_2=s_2$ THEN $d=Y$；规则的可信度为1（规则的支持集是$\{x_3\}$）。

输出$c_{AT}(x)<1$的可能性规则：

规则1：IF $a_1\neq r_1$ AND $a_2=s_3$ THEN $d=Y$；规则的可信度为0.57（规则的支持集是$\{x_1\}$）。

规则2：IF $a_1=r_1$ OR r_2 AND $a_2=S_2$ OR S_3 THEN $d=Y$；规则的可信度为0.79（规则的支持集是$\{x_2\}$）。

规则3：IF $a_1=r_2$ AND $a_2=S_1$ OR S_2 THEN $d=N$；规则的可信度为0.59（规则的支持集是$\{x_4\}$）。

规则4：IF $a_1=r_1$ OR r_2 OR r_3 AND $a_2\neq S_3$ THEN $d=Y$；规则的可信度为0.78（规则的支持集是$\{x_5\}$）。

规则5：IF $a_2=S_3$ THEN $d=N$；规则的可信度为0.42（规则的支持集是$\{x_6\}$）。

结论 利用模糊粗糙集上下近似算子，本文给出了一种在不完备模糊信息系统中知识获取的模糊粗糙集方法，它能够把隐藏在不完备模糊信息系统中的知识，以决策规则的形式表现出来。实践证明本文提出的模糊粗糙集方法是完全可行的。

参考文献

1 Pawlak Z. Rough sets. International Journal of Computer and Information Science, 1982, 11(5): 341～356

2 张文修，吴伟志，梁吉业，等. 粗糙集理论与方法. 北京：科学出版社，2001

3 Pawlak Z. Drawing conclusions from data—the rough set way. International Journal of Intelligent Systems, 2001, 16: 3～11

4 Dubois D, Prade H. Rough fuzzy sets and fuzzy rough sets. International Journal of General System, 1990, 17: 191～208

5 Bodjanova S. Approximation of a fuzzy concepts in decision making. Fuzzy Sets and Systems, 1997, 85: 23～29

6 Kryszkiewicz M. Rough set approach to incomplete information systems. Information Sciences, 1998, 112: 39～49

7 Kryszkiewicz M. Rules in incomplete information systems. Information Sciences, 1999, 113: 271～292