第16卷 第12期
2006年12月

计算机技术与发展
COMPUTER TECHNOLOGY AND DEVELOPMENT

Vol.16 No.12
Dec. 2006

基于新集对理论下的粗糙集模型

张海东,舒 兰
(电子科技大学,四川 成都 610054)

摘 要:为了用集对分析方法进一步刻画不完备信息系统,文中把新集对分析理论与粗糙集理论结合在一起,提出了一种新的集对粗糙集模型,从而拓宽了集对分析方法的应用;定义了一种不完备信息系统的上、下近似算子,得到了一些相关的性质。最后通过一个简单的例子说明了上述方法的可行性。

关键词:不完备信息系统;集对分析;粗糙集

中图分类号:TP301;TP311 **文献标识码**:A **文章编号**:1673-629X(2006)12-0004-03

The Rough Set Model Based on a New Idea of Set Pair Theory

ZHANG Hai-dong, SHU Lan
(University of Electronic Science and Technology of China, Chengdu 610054, China)

Abstract: In order to describe the incomplete information system, the rough set model based on a new idea of set pair analysis is proposed by introducing a new idea of set pair theory into rough sets in this paper, which extends the application of set pair analysis. The upper approximation and the lower approximation operators are defined in incomplete information system. Furthermore, some related properties are discussed. Finally, by an example, it is verified that this method is more feasible.

Key words: incomplete information system; set pair analysis; rough set

0 前 言

文献[1]提出了集对分析之后,近年来的研究和应用日趋广泛。集对论实质上是一种研究不确定性的数学方法,其核心思想是把确定不确定视作一个系统。虽然如此,集对论在认识和理论上仍有许多值得研究的问题:

(1)“集对”A 和 B 中元素的数目不一定相等;

(2)“集对分析”不应是集合 A 和 B 中相应元素的直接对比,而应在元素序偶中进行对比。因此,对于在“集对”的数学本质、“联系度”的数学结构和“集对分析”的严格数学法则等方面,有必要重新定义和创立。

1 集对、联系度的新定义

定义 1[2] 设有两个非空有限集合:

$X = \{x: \forall x \in X, X \neq \emptyset\}$,

$Y = \{y: \forall y \in Y, Y \neq \emptyset\}$,

其基数分别为 $\overline{X} = m$ 和 $\overline{Y} = n$,则称

$H(X,Y) = X \times Y = \{(x,y): \forall x \in X \ \& \ y \in Y\}$

的非空有限序偶集为由 X 与 Y 构成的“集对直集”,其基数为 $\overline{H} = N = mn$。

定义 2[2] 有问题 W:“X 和 Y 两集合间有无关系 R?”,若:

(1)“$x \in X$ 与 $y \in Y$ 有关系 R”,记为 xRy,称为“x 与 y 在问题 W 下具有同一性”,称序偶子集:$H_R(X,Y) = \{(x,y): \forall x \in X \ \& \ \forall y \in Y, xRy\}$ 为集合 X 与 Y 在问题 W 下的同一性序偶集,若 $\overline{H}_R = S$ 为 $H_R(X,Y)$ 的基数,则 $\frac{\overline{H}_R}{\overline{H}} = \frac{S}{N}$ 称为集合 X 与 Y 在问题 W 下的同一度,简记为 a。

(2)“$x \in X$ 与 $y \in Y$ 无关系 R”,记为 $x\not R y$,称为“x 与 y 在问题下 W 具有对立性”,称序偶子集:$H_{\not R}(x,y) = \{(x,y): \forall x \in X \ \& \ \forall y \in Y, x\not R y\}$ 为集合 X 与 Y 在问题 W 下的对立性序偶集,若 $\overline{H}_{\not R} = P$ 为 $H_{\not R}(X,Y)$ 的基数,则 $\frac{\overline{H}_{\not R}}{\overline{H}} = \frac{P}{N}$ 称为集合 X 与 Y 在问题 W 下的对立度,简记为 b。

(3)“$x \in X$ 与 $y \in Y$ 不确定有无关系 R”,记为 $x\tilde{R}y$,称为“x 与 y 在问题 W 下具有不确定性”,称序偶子集:$H_{\tilde{R}}(x,y) = \{(x,y): \forall x \in X \ \& \ \forall y \in Y, x\tilde{R}y\}$ 为集合 X 与 Y 在问题 W 下的不确定性序偶集,若 $\overline{H}_{\tilde{R}} = F$ 为 $H_{\tilde{R}}(X,Y)$ 的基数,则 $\frac{\overline{H}_{\tilde{R}}}{\overline{H}} = \frac{F}{N}$ 称为集合 X 与 Y 在问题 W 下的不确定度,简记为 c。

引入“不确定性标记 i”和“对立性标记 j”,则表达式

收稿日期:2006-03-20

基金项目:电子科技大学青年科技基金

作者简介:张海东(1980-),男,甘肃人,硕士研究生,从事模糊集和粗糙集理论的研究;舒 兰,教授,博士生导师,从事模糊数学理论及应用、模糊信息处理技术、模糊模式识别、粗糙理论及应用等领域的研究。

$$u(x,y)=\frac{S}{N}+\frac{F}{N}i+\frac{P}{N}j \tag{1}$$

称为同异反联系度,式(1)可简写为:

$$u(x,y)=a+bi+cj$$

其中:$a,b,c\in[0,1]$ 为实数,$a+b+c=1$。

2 基于集对联系度的粗糙集模型

对于信息系统[3] $S=(U,AT)$,其中 U 为非空论域,AT 是属性的非空有限集合。对于 $\forall a\in AT,x\in U$,$f(x,a)\in V_a$,其中 f 是一个信息函数,V_a 是 a 的值域。

令 $A\subseteq AT$,定义相似关系[3-5] 如下:

$\mathrm{SIM}(A)=\{(x,y)\in U\times U:\forall a\in A,f(x,a)=f(y,a)$ 或 $f(x,a)=*$ 或 $f(y,a)=*\}$

定义3 在一个不完备信息系统 $S=(U,AT)$ 中,称 x 与 a 或 y 与 a 有关系 R,记为 xRa 或 yRa,若 $f(x,a)=f(y,a)$ and $f(x,a)\neq *$ and $f(y,a)\neq *$;称 x 与 a 或 y 与 a 无关系 R,记为 $x\not R a$ 或 $y\not R a$,若 $f(x,a)\neq f(y,a)$ and $f(x,a)\neq *$ and $f(y,a)\neq *$;称 x 与 a 或 y 与 a 不确定有无关系 R,记为 $x\tilde{R}a$ 或 $y\tilde{R}a$,若 $f(x,a)=*$ or $f(y,a)=*$,其中 $\forall x\in U,y\in U,a\in AT$。

对于 $A\subseteq AT$,$H(U,A)=U\times A=\{(x,a):\forall x\in X,\forall a\in A\}$,

其基数为 $|H(U,A)|=|U|\cdot|A|=n,\forall x\in U,a\in A$,记

$$u_A(x,y)=\frac{Ns}{n}+\frac{Nf}{n}i+\frac{Np}{n}j \tag{2}$$

其中 $|U|=N,\forall x\in U,y\in U$,$s$ 为集合 U 与 A 中元素有关系 R 时的基对个数,即

$s=|\{(y,a)\in U\times A:f(x,a)=f(y,a)$ and $f(x,a)\neq *$ and $f(y,a)\neq *\}|$

p 为集合 U 与 A 中元素无关系 R 时的基对个数,即

$p=|\{(y,a)\in U\times A:f(x,a)\neq f(y,a)$ and $f(x,a)\neq *$ and $f(y,a)\neq *\}|$

f 为集合 U 与 A 中元素不确定有无关系 R 时的基对个数,即

$f=|\{(y,a)\in U\times A:f(x,a)=*$ or $f(y,a)=*\}|$

简记式(2)为 $u_A(x,y)=a+bi+cj$

其中 $a=\frac{Ns}{n},b=\frac{Nf}{n},c=\frac{Np}{n}$,特别地,记

$u_A(x,y)=u(x,y)$

当 $A\subseteq AT,0\le\alpha\le1$,定义如下集对 α 相似关系[5]:

$$\mathrm{SIM}(A)_\alpha=\{(x,y)\in U\times U:u_A(x,y)=a+bi+cj,a+b\ge\alpha\}$$

定义4 对于信息系统 $S=(U,AT),x\in U,A\subseteq AT$,定义 x 的 $A-\alpha$ 邻域如下:

$S_A^\alpha(x)=\{y:(x,y)\in\mathrm{SIM}(A)_\alpha\}$

它表示在 A 下与对象 x 的对立度不大于 $1-\alpha$ 的所有对象的集合。

定义5 设有信息系统 $S=(U,AT),x\in U,A\subseteq AT,\alpha>0$,定义 X 的 $A-\alpha$ 集对型下、上近似如下:

$\underline{R}_A^\alpha(X)=\{x:S_A^\alpha(x)\subseteq X\}$,

$\bar{R}_A^\alpha(X)=\{x:S_A^\alpha(x)\cap X\neq\emptyset\}$

当 $S=(U,AT)$ 为完备信息系统时,

$\underline{R}_A^1(X)=\underline{R}(X),\bar{R}_A^1(X)=\bar{R}(X)$

即 Pawlak 上、下近似是集对型上、下近似的特殊情况。

定理1 设信息系统 $S=(U,AT),x\in U,A\subseteq AT,\alpha>0$,则 X 的 $A-\alpha$ 集对型下、上近似具有以下性质:

(1) $\underline{R}_A^\alpha(X)=\sim\bar{R}_A^\alpha(\sim X),\bar{R}_A^\alpha(X)=\sim\underline{R}_A^\alpha(\sim X)$;

(2) $0\le\alpha_1<\alpha_2\le1,\underline{R}_A^{\alpha_1}(X)\subseteq\underline{R}_A^{\alpha_2}(X)$,

$\bar{R}_A^{\alpha_1}(X)\supseteq\bar{R}_A^{\alpha_2}(X)$;

(3) 若 $X\subseteq Y,\underline{R}_A^\alpha(X)\subseteq\underline{R}_A^\alpha(Y)$,

$\bar{R}_A^\alpha(X)\subseteq\bar{R}_A^\alpha(Y)$;

(4) $\underline{R}_A^\alpha(X\cap Y)=\underline{R}_A^\alpha(X)\cap\underline{R}_A^\alpha(Y)$,

$\bar{R}_A^\alpha(X\cup Y)=\bar{R}_A^\alpha(X)\cup\bar{R}_A^\alpha(Y)$;

(5) $\underline{R}_A^\alpha(X\cup Y)\supseteq\underline{R}_A^\alpha(X)\cup\underline{R}_A^\alpha(Y)$,

$\bar{R}_A^\alpha(X\cap Y)\subseteq\bar{R}_A^\alpha(X)\cap\bar{R}_A^\alpha(Y)$;

证明:只证下近似相关的性质,上近似与此类似。

① 根据定义5,$\sim\bar{R}_A^\alpha(\sim X)=$

$\sim\{x:S_A^\alpha(x)\cap(\sim X)\neq\emptyset\}=$

$\{x:S_A^\alpha(x)\cap(\sim X)=\emptyset\}=$

$\{x:S_A^\alpha(x)\subseteq X\}=\underline{R}_A^\alpha(X)$

② 如果 $\alpha_1\le\alpha_2$,则 $\forall x\in U,S_A^{\alpha_1}(x)\supseteq S_A^{\alpha_2}(x)$。$\forall x\in\underline{R}_A^{\alpha_1}(X)$,所以 $S_A^{\alpha_1}(x)\subseteq X$,故 $S_A^{\alpha_2}(x)\subseteq X$,从而 $x\in\underline{R}_A^{\alpha_2}(X)$,故 $\underline{R}_A^{\alpha_1}(X)\subseteq\underline{R}_A^{\alpha_2}(X)$。

③ 如果 $X\subseteq Y,\forall x\in\underline{R}_A^\alpha(X)$,使得 $S_A^\alpha(x)\subseteq X\subseteq Y$,从而 $x\in\underline{R}_A^\alpha(Y)$。

④ 根据定义5,$\underline{R}_A^\alpha(X\cap Y)=$

$\{x:S_A^\alpha(x)\subseteq X\cap Y\}=$

$\{x:S_A^\alpha(x)\subseteq X\}\cap\{x:S_A^\alpha(x)\subseteq Y\}=$

$\underline{R}_A^\alpha(X)\cap\underline{R}_A^\alpha(Y)$

⑤ 根据性质(3)直接可得。

证毕。

另外,$\underline{R}^\alpha(X)\subseteq X\subseteq\bar{R}^\alpha(X)$,

$\underline{R}^\alpha(\underline{R}^\alpha(X))\subseteq\underline{R}^\alpha(X)\subseteq\bar{R}^\alpha(\underline{R}^\alpha(X))\subseteq X\subseteq\underline{R}^\alpha(\bar{R}^\alpha(X))\bar{R}^\alpha(X)\subseteq\bar{R}^\alpha(\bar{R}^\alpha(X))$ 这两个式子也成立。

3 实　例

这里给出一个不完备信息系统的实际例子,并计算出给 X 定的上、下近似集,从而验证其相关的性质,来说明这种定义方法的可行性。

设不完备信息系统 $S=(U,AT)$ 如表1所示,有 $U=\{1,2,3,4,5,6,7,8,9,10\}$,$AT=\{A,B,C,D,E\}$,$\forall a\in AT,x\in U,f(x,a)=0$,或 $f(x,a)=1$ 或 $f(x,a)=*$。

表1 不完备信息系统(U,AT)

U \ AT	A	B	C	D	E
1	0	1	1	0	1
2	1	0	0	*	0
3	0	1	*	0	1
4	*	0	0	1	1
5	1	0	1	0	0
6	0	*	0	1	0
7	0	1	0	*	0
8	1	*	0	0	0
9	1	1	0	0	1
10	0	0	1	0	1

取 $\alpha = 0.6$,有:

$S_A^{\alpha}(1) = \{1,3,7,10\}, S_A^{\alpha}(2) = \{2,4,5,6,8,9\}$,

$S_A^{\alpha}(3) = \{1,3,4,7,8,10\}, S_A^{\alpha}(4) = \{2,3,4,6,8\}$

$S_A^{\alpha}(5) = \{2,5,8\}, S_A^{\alpha}(6) = \{2,4,6,7,8\}$,

$S_A^{\alpha}(7) = \{1,3,6,7,8,9\}, S_A^{\alpha}(8) = \{2,3,4,5,6,7,8,9\}$,

$S_A^{\alpha}(9) = \{2,7,8,9\}, S_A^{\alpha}(10) = \{1,3,10\}$,

设 $X = \{2,4,5,6,7,8,9\}$, $\underline{R}^{\alpha}(X) = \{2,5,6,9\}$, $\bar{R}^{\alpha}(X) = \{1,2,3,4,5,6,7,8,9\}$

显然,$\underline{R}^{\alpha}(X) \subseteq X \subseteq \bar{R}^{\alpha}(X)$ 成立。

$\underline{R}^{\alpha}(\underline{R}^{\alpha}(X)) = \emptyset, \bar{R}^{\alpha}(\underline{R}^{\alpha}(X)) = \{2,4,5,6,7,8,9\}$

$\underline{R}^{\alpha}(\bar{R}^{\alpha}(X)) = \{2,4,5,6,7,8,9\}$

$\bar{R}^{\alpha}(\bar{R}^{\alpha}(X)) = \{1,2,3,4,5,6,7,8,9,10\}$

显然,$\underline{R}^{\alpha}(\underline{R}^{\alpha}(X)) \subseteq \underline{R}^{\alpha}(X) \subseteq \bar{R}^{\alpha}(\underline{R}^{\alpha}(X)) \subseteq X \subseteq \underline{R}^{\alpha}(\bar{R}^{\alpha}(X) \subseteq \bar{R}^{\alpha}(\bar{R}^{\alpha}(X))$ 成立。

4 结 论

集对分析自1989年提出来以后,在许多领域得到了应用,但仍有许多认识上和理论上的问题值得研究[1,2]。文中在文献[2]基础之上,考虑用集对分析方法进一步刻画不完备信息系统,把此理论与粗糙集理论有机地结合在一起,提出了一种新的集对粗糙集理论,定义了一种不完备信息系统的上、下近似运算,得到了一些性质,并且通过一个简单的例子说明了上述方法的可行性,在粗糙集用于研究不完备信息系统方面做了一定推广。

参考文献:

[1] 赵克勤.集对分析及其初步应用[M].杭州:浙江科学出版社,2000.

[2] 张 鹏,王光远.新集对论[J].哈尔滨建筑大学学报,2000,33(3):1-5.

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

[4] 黄 兵,周献中.基于集对分析的不完备信息系统粗糙集模型[J].计算机科学,2002,29(9):1-3.

[5] 黄 兵,钟 斌.改进集对粗集模型[J].计算机工程与应用,2004,40(2):82-84.

(上接第3页)

过滤Agent进行主题的过滤、个性化过滤;最后将得到的结构返回给用户Agent,并显示出来。信息过滤系统Agent的协同工作的流程如图3所示。

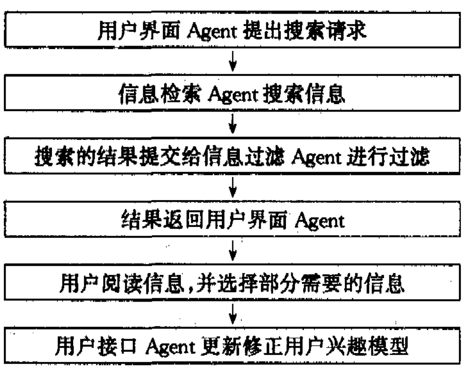

图3 信息过滤系统Agent的操作流程图

5 小 结

文中提出了一种基于Agent的个性化信息过滤系统结构,适用于一般万维网上的智能信息搜索系统。与传统搜索引擎相比,该结构下的信息搜索可以帮助使用者根据本人的兴趣和偏爱获得较高的匹配。同时,该结构下系统的实现面临着一些难点问题,突出表现为用户兴趣更新问题、Agent间的通信问题、概念的标准化定义问题和多系统的集成问题,这些研究的进展将为多Agent系统的进一步实用化奠定基础。

参考文献:

[1] Mladenic D. Personal WebWatcher: design and implementation [EB/OL]. 1996. http://ranger. uta. edu/alp/ix/readings/mladenic96personalWeWatcher. pdf.

[2] 冯 翱,刘 斌,卢增翔,路海明,等.Open Bookmark—基于Agent的信息过滤系统[J].清华大学学报:自然科学版,2001,41(3):85-88.

[3] 费洪晓,巩艳玲,谢文彪,等.基于混合学习策略的多Agent信息过滤系统[J].计算机应用,2006,26(2):267-269.

[4] 李 俊,张灵玲,周文辉,等.一个智能用户接口Agent的设计与实现[J].软件学报,1999,10(8):23-27.

[5] 白丽君.基于智能Agent的用户兴趣发现和更新[J].计算机工程,2003,21(2):70-72.

[6] 张国印,陈 先,皮 鹏.基于词频统计的个性化信息过滤技术[J].哈尔滨工业大学学报,2003,24(1):63-67.