第47卷第2期　　　　大连理工大学学报　　　　Vol. 47, No. 2
2007年3月　　Journal of Dalian University of Technology　　Mar. 2007

文章编号: 1000-8608(2007)02-0287-04

一种基于弱模糊相似关系的广义粗糙集

田　宏[1,2], 王秀坤*[1]

(1.大连理工大学 电子与信息工程学院, 辽宁 大连 116024;
2.大连交通大学 软件学院, 辽宁 大连 116028)

摘要: 粗糙集理论是建立在等价关系的基础上发展起来的,但等价关系性质的应用领域是有限的,等价关系不能对现实世界中的元素关系给出客观的描述. 为此提出用模糊相似关系和弱模糊相似关系的概念来代替等价关系,同时引入了弱模糊相似关系的相似度概念和相似类概念,定义了基于弱模糊相似关系的广义粗糙集及标准的广义粗糙集上下近似;研究了两对上下近似算子的性质,讨论了3种粗糙隶属函数的性质并验证了它们的性质.

关键词: 弱模糊相似关系; 相似度; 相似类
中图分类号: TP311　　文献标识码: A

0　引　言

粗糙集理论是一种分析不确定和模糊数据的新的数学理论,最初由Pawlak于1982年提出[1]. 它在许多数据挖掘和知识发现领域发挥了重要的作用. 在机器学习、知识发现、决策分析、专家系统、分类、数据的模式发现、模式识别、数据冗余和许多其他领域,粗糙集理论均提供了一种十分有效的数学方法和工具[2].

经典的粗糙集理论是建立在等价关系基础之上,研究不精确边界集合的理论. 粗糙集是由一对上下近似来描述的,根据两个从论域中给出的精确集合来近似描述[3]. 等价关系的确在现实中比较普遍,但是在实际处理问题时遇到的很多关系都不满足这种严格的等价性. 考虑一个相似关系或相容关系,而不是等价关系,如此的关系表达是一种比较弱的不分明形式,而自反性对于表示任意一种不分明性或相似性的形式似乎都是相当必要的,其他对称性和传递性都是可以放宽的.

本文的主要目的是通过引入弱模糊相似关系和相似度来扩展经典的粗糙集. 本文提出的粗糙集可以被看做是相似粗糙集的推广[4].

1　弱模糊相似关系

假设给定一个有限非空的个体集合 U,考虑 U 上的一个相似关系,它表示相似性的某种形式. 相似关系定义如下:

定义1　一个模糊相似关系映射为 $s: U\times U\to[0,1]$,对任意 $x,y,z\in U$ 满足如下性质

(a) 自反性: $s(x,x)=1$

(b) 对称性: $s(x,y)=s(y,x)$

(c) 传递性: $s(x,z)\geqslant\max\limits_{y\in U}\min[s(x,y),s(y,z)]$

定义2　一个弱模糊相似关系映射为 $S: U\times U\to[0,1]$,对任意 $x,y,z\in U$ 满足如下性质

(a) 自反性: $S(x,x)=1$

(b) 对称性: 如果 $S(x,y)>0$ 那么 $S(y,x)>0$

(c) 传递性: 如果 $S(y,x)\geqslant S(x,y)>0$,并且 $S(z,y)\geqslant S(y,z)>0$,那么 $S(z,x)\geqslant S(x,z)$

定义3　相似度: 设 $_x$ 和 $_y$ 是一个论域 U 中的任意两个对象 x 和 y 分别对应属性集合 A 的属性值的集合. 论域中的 x 与 y 的相似度 $R(x,y)\to[0,1]$ 满足如下性质:

$$R(x,y)=\frac{\sum\limits_{a\in A}{}_x(a)\cap{}_y(a)}{\sum\limits_{a\in A}{}_x(a)}$$

收稿日期: 2005-08-04;　修回日期: 2007-01-19.
基金项目: 辽宁省自然科学基金资助项目(20042114).
作者简介: 田　宏(1968-),女,博士生,副教授,E-mail: th@djtu.edu.cn; 王秀坤*(1945-),女,教授,博士生导师.

引入相似度之后的关系就已经不是传统意义上的相似关系了[5]. 相似度这个概念能客观有效地反映对象之间相似的程度,其反映了对象 x 相似于 y 的程度,x 与 y 相似是有方向的,例如,“儿子像他的父亲”,其中儿子是 x,父亲是 y,它不等价于“父亲像他的儿子”. 相似关系至少在某些情况下不强调对称性,在更多的情况下也不强调传递性.

例 1 已知二进制信息表 1,其中的对象的集合为 $U=\{x_1,\cdots,x_{10}\}$, 8个属性的集合描述为 $A=\{a_1,\cdots,a_8\}$. 对于前 3个对象,有 $x_1=\{a_4,a_5,a_7,a_8\}$, $x_2=\{a_1,a_5\}$, $x_3=\{a_1,a_3,a_5,a_7\}$.

根据相似度的定义有如下结果:

$$R(x_1,x_2)=\frac{1}{4},\ R(x_2,x_1)=\frac{1}{2}$$

$$R(x_1,x_3)=\frac{2}{4},\ R(x_3,x_1)=\frac{2}{4}$$

$$R(x_2,x_3)=\frac{2}{2}=1,\ R(x_3,x_2)=\frac{2}{4}$$

表 1 二进制信息表

Tab. 1 Binary information table

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
x_1	0	0	0	1	1	0	1	1
x_2	1	0	0	0	1	0	0	0
x_3	1	0	1	0	1	0	1	0
x_4	1	0	0	0	0	1	1	0
x_5	0	0	1	0	1	0	1	1
x_6	0	0	0	1	0	0	1	1
x_7	0	1	0	1	1	0	0	1
x_8	1	0	0	1	0	0	1	0
x_9	0	0	1	0	1	1	0	1
x_{10}	1	0	0	1	0	1	0	0

2 广义粗糙集近似

从弱模糊相似关系和相似度出发,基于论域的粗糙集概念被推广.

定义 4 相似类:设 U 是一个非空论域,R 是论域 U 上的弱模糊相似关系. 对于任意元素 $x\in U$, $C_s^{T}(x)$ 和 $C_p^{T}(x)$ 被分别定义为与元素 x 相似的个体 y 的相似类和元素 x 相似于个体 y 的相似类,对于给定相似度 $T\in[0,1]$,有如下公式

$$C_p^{T}(x)=\{y\in U\mid R(y,x)\geqslant T\}$$

$$C_s^{T}(x)=\{y\in U\mid R(x,y)\geqslant T\}$$

根据相似类的定义,可以用相似类代替等价类,能够分别构造论域中的两个覆盖 $\{C_p^{T}(x)\mid x\in U\}$ 和 $\{C_s^{T}(x)\mid x\in U\}$. 通过扩展经典的粗糙集,得到两对广义粗糙集近似.

定义 5 对任意 $X\subseteq U$, 定义两对广义粗糙集近似:

(1) 基于元素的广义粗糙集近似

$$\underline{apr}_e^{T}(X)=\{x\in U\mid C_p^{T}(x)\subseteq X\}$$

$$\overline{apr}_e^{T}(X)=\{x\in U\mid C_p^{T}(x)\cap X\neq\varnothing\}$$

(2) 基于相似类的广义粗糙集近似

$$\underline{apr}_c^{T}(X)=\bigcup\{C_p^{T}(x)\mid C_p^{T}(x)\subseteq X,x\in U\}$$

$$\overline{apr}_c^{T}(X)=\bigcup\{C_p^{T}(x)\mid C_p^{T}(x)\cap X\neq\varnothing,x\in U\}$$

在定义 5(1)中,下近似由那些在 U 中相似类包含于 X 的元素组成,上近似由那些在 U 中相似类与 X 相交的元素组成. 在定义 5(2)中,下近似是所有包含于 X 的相似类的并集,上近似是所有与 X 相交的相似类的并集. 这些相似关系可以表示如下:

$$\underline{apr}_e^{T}(X)\subseteq\underline{apr}_c^{T}(X)\subseteq X\subseteq\overline{apr}_e^{T}(X)\subseteq\overline{apr}_c^{T}(X)$$

上下近似的差被称为 X 的边界区域:

$$BN_e^{T}(X)=\overline{apr}_e^{T}(X)-\underline{apr}_e^{T}(X)$$

$$BN_c^{T}(X)=\overline{apr}_c^{T}(X)-\underline{apr}_c^{T}(X)$$

同理,可以利用覆盖 $\{C_s^{T}(x)\mid x\in U\}$ 定义粗糙集近似.

这对上下近似运算符 $(\underline{apr}_e^{\alpha},\overline{apr}_e^{\alpha})$ 生成两个一元集合算子,被称为粗糙集近似算子[6]. 与其他集合运算符相结合,例如 $\neg$、$\bigcup$ 和 $\bigcap$,得到如下结果:

(pe0) $\underline{apr}_e^{\alpha}(X)=\neg\overline{apr}_e^{\alpha}(\neg X)$

$\overline{apr}_e^{\alpha}(X)=\neg\underline{apr}_e^{\alpha}(\neg X)$

(pe1) $\underline{apr}_e^{\alpha}(X)\subseteq X\subseteq\overline{apr}_e^{\alpha}(X)$

(pe2) $\underline{apr}_e^{\alpha}(\varnothing)=\overline{apr}_e^{\alpha}(\varnothing)=\varnothing$

(pe3) $\underline{apr}_e^{\alpha}(U)=\overline{apr}_e^{\alpha}(U)=U$

(pe4) $\underline{apr}_e^{\alpha}(X\cap Y)=\underline{apr}_e^{\alpha}(X)\cap\underline{apr}_e^{\alpha}(Y)$

$\overline{apr}_e^{\alpha}(X\cap Y)\subseteq\overline{apr}_e^{\alpha}(X)\cap\overline{apr}_e^{\alpha}(Y)$

(pe5) $\underline{apr}_e^{\alpha}(X\cup Y)\supseteq\underline{apr}_e^{\alpha}(X)\cup\underline{apr}_e^{\alpha}(Y)$

$\overline{apr}_e^{\alpha}(X\cup Y)=\overline{apr}_e^{\alpha}(X)\cup\overline{apr}_e^{\alpha}(Y)$

(pe6) $X\neq\varnothing\Rightarrow\overline{apr}_e^{0}(X)=U$

(pe7) $X\subset U\Rightarrow\underline{apr}_e^{0}(X)=\varnothing$

(pe8) $\alpha\leqslant\beta\Rightarrow\overline{apr}_e^{\beta}(X)\subseteq\overline{apr}_e^{\alpha}(X)$

$\alpha\leqslant\beta\Rightarrow\underline{apr}_e^{\alpha}(X)\subseteq\underline{apr}_e^{\beta}(X)$

(pe9) $X\subseteq Y\Rightarrow\underline{apr}_e^{\alpha}(X)\subseteq\underline{apr}_e^{\alpha}(Y)$

$X\subseteq Y\Rightarrow\overline{apr}_e^{\alpha}(X)\subseteq\overline{apr}_e^{\alpha}(Y)$

性质(pe0)说明上下近似是关于集合补运算 $\neg$ 的对偶算子. 性质(pe2)和(pe3)提出了两种边界条件. 性质(pe4)和(pe5)可以分别看做关于集合交与集合并的弱分配率. 当 $\alpha=0$ 时,性质

(pe6) 和 (pe7) 说明非空集合 $X\subset U$的上下近似分别等于 U和 $\varnothing$,性质 (pe8)说明 T的值越大则下近似也越大,而上近似越小. 性质 (pe9) 指出近似算子关于集合包含的单调性.

同理,定义 5(2) 中的上下近似算符 ($\underline{apr}_c^{T}$, $\overline{apr}_c^{T}$) 也生成两个一元集合算子. 满足性质同上.

3　广义粗糙隶属函数

文献 [7] 指出,至少存在两种观点解释粗糙集理论,分别基于算子的观点和基于集合的观点. 在上章中讨论了广义上下近似算子,在本章中给出基于集合观点的粗糙隶属函数的定义.

利用定义 5中论域的覆盖,扩展粗糙隶属函数并得到 3个广义粗糙隶属函数.

定义 **6**　对任意 $X\subseteq U$, $T\in(0,1]$,定义如下 3种粗糙隶属函数:

$$\_{}^{m}_{X}(y)^{T}=\min\left\{\frac{|C_p^{T}(x)\cap X|}{|C_p^{T}(x)|}\,\middle|\, x\in U, y\in C_p^{T}(x)\right\}$$

$$\_{}^{M}_{X}(y)^{T}=\max\left\{\frac{|C_p^{T}(x)\cap X|}{|C_p^{T}(x)|}\,\middle|\, x\in U, y\in C_p^{T}(x)\right\}$$

$$\_{}^{*}_{X}(y)^{T}=\mathrm{avg}\left\{\frac{|C_p^{T}(x)\cap X|}{|C_p^{T}(x)|}\,\middle|\, x\in U, y\in C_p^{T}(x)\right\}$$

它们分别为最小、最大和平均粗糙隶属函数.

广义粗糙隶属函数的定义参考文献 [8]. 文献 [9]提出了对用覆盖定义近似算子的解释. 最大、最小和平均的等式可以被看做在定义粗糙隶属函数中的最乐观、最悲观和平均的观点. 元素 y 的最小粗糙隶属函数由集合 $C_p^{T}(x)$与 X的最小的交集决定,最大的粗糙隶属函数由集合 $C_p^{T}(x)$ 与 X的最大的交集决定,平均粗糙隶属函数取决于所有集合 $C_p^{T}(x)$ 与 X 的交集的平均值,其解释为 $\mathrm{avg}\{0.3, 0.4, 0.5\}=0.4$.

三种粗糙隶属函数的关系由下式表示:

$$\_{}^{m}_{X}(y)^{T}\leqslant \_{}^{*}_{X}(y)^{T}\leqslant \_{}^{M}_{X}(y)^{T}$$

根据 T的取值,可以定义一系列粗糙隶属函数. 对任意 $X, Y\subseteq U$,最大、最小和平均粗糙隶属函数满足如下性质:

(pr0) $\_{}^{m}_{U}(x)^{T}=\_{}^{*}_{U}(x)^{T}=\_{}^{M}_{U}(x)^{T}=1$

(pr1) $\_{}^{m}_{\varnothing}(x)^{T}=\_{}^{*}_{\varnothing}(x)^{T}=\_{}^{M}_{\varnothing}(x)^{T}=0$

(pr2) $[\forall x\in U, y\in R_p^{T}(x)\Leftrightarrow z\in R_p^{T}(x)]\Rightarrow[\_{}^{m}_{X}(y)^{T}=\_{}^{m}_{X}(z)^{T}, \_{}^{*}_{X}(y)^{T}=\_{}^{*}_{X}(z)^{T}, \_{}^{M}_{X}(y)^{T}=\_{}^{M}_{X}(z)^{T}]$

(pr3) $\exists x\in U, y, z\in R_p^{T}(x)\Rightarrow[(\_{}^{m}_{X}(y)^{T}\neq 0\Rightarrow \_{}^{m}_{X}(z)^{T}\neq 0), (\_{}^{m}_{X}(y)^{T}=1\Rightarrow \_{}^{m}_{X}(z)^{T}=1)]$

(pr4) $y\in X\Rightarrow \_{}^{m}_{X}(y)^{T}>0$

(pr5) $\_{}^{M}_{X}(y)^{T}=1\Rightarrow y\in X$

(pr6) $X\subseteq Y\Rightarrow[\_{}^{m}_{X}(y)^{T}\leqslant \_{}^{m}_{Y}(y)^{T}, \_{}^{*}_{X}(y)^{T}\leqslant \_{}^{*}_{Y}(y)^{T}, \_{}^{M}_{X}(y)^{T}\leqslant \_{}^{M}_{Y}(y)^{T}]$

(pr7) $X\neq\varnothing\Rightarrow \_{}^{m}_{X}(x)^{0}=\_{}^{*}_{X}(x)^{0}=\_{}^{M}_{X}(x)^{0}=\frac{|X|}{|U|}=P(X)$

性质 (pr0) 和 (pr1) 说明了边界性质,也就是说,对于 U和 $\varnothing$,对所有元素最大、最小和平均粗糙隶属函数具有相同的值,分别是 1和 0. 性质 (pr2) 和 (pr3) 表明在同一个覆盖中的两个相似元素具有相似的粗糙隶属函数. 性质 (pr4) 和 (pr5) 说明了 X 中元素的隶属函数值的约束性质. 性质 (pr6) 说明了近似算子关于部分包含的单调性. 当 T设为 0时,论域的覆盖为 U,在这种情况下,在 X 中的元素粗糙隶属函数的值等于 X 的概率值,性质 (pr7) 中显示.

4　一个说明例子

用给出的二进制信息表 1来举例说明以上的概念,设 T为 0.6. 根据定义 3、4和 5,可以得到所有元素的相似类如下所示:

$$R_p^{0.6}(x_1)=\{x_1, x_5, x_6, x_7\}$$
$$R_p^{0.6}(x_2)=\{x_2, x_3\}$$
$$R_p^{0.6}(x_3)=\{x_3, x_5\}$$
$$R_p^{0.6}(x_4)=\{x_3, x_4, x_8, x_{10}\}$$
$$R_p^{0.6}(x_5)=\{x_1, x_5, x_9\}$$
$$R_p^{0.6}(x_6)=\{x_1, x_5, x_6, x_7, x_8\}$$
$$R_p^{0.6}(x_7)=\{x_1, x_7\}$$
$$R_p^{0.6}(x_8)=\{x_1, x_3, x_4, x_6, x_8, x_{10}\}$$
$$R_p^{0.6}(x_9)=\{x_5, x_9\}$$
$$R_p^{0.6}(x_{10})=\{x_4, x_{10}\}$$

考虑一个对象集合:

$$X=\{x_1, x_3, x_5, x_7, x_9\}$$

根据定义 6得到 X 的粗糙近似为

$$\underline{apr}_e^{0.6}(X)=\{x_3, x_5, x_7, x_9\}$$
$$\overline{apr}_e^{0.6}(X)=\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$$
$$\underline{apr}_c^{0.6}(X)=\{x_1, x_3, x_5, x_7, x_9\}$$
$$\overline{apr}_c^{0.6}(X)=\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$$

X的粗糙边界为

$$BN_e^{0.6}(X)=\{x_1, x_2, x_4, x_6, x_8\}$$
$$BN_c^{0.6}(X)=\{x_1, x_2, x_4, x_6, x_8, x_{10}\}$$

对于一个元素 x_6,它属于 3个相似类: $R_p^{0.6}(x_1)$、$R_p^{0.6}(x_6)$ 和 $R_p^{0.6}(x_8)$. 有

$$\frac{|R_p^{0.6}(x_1)\cap X|}{|R_p^{0.6}(x_1)|}=\frac{3}{4}$$

$$\frac{|R_p^{0.6}(x_6)\cap X|}{|R_p^{0.6}(x_6)|}=\frac{3}{5}$$

$$\frac{|R_p^{0.6}(x_8)\cap X|}{|R_p^{0.6}(x_8)|}=\frac{2}{6}=\frac{1}{3}$$

由定义 6, x_6 的最小、最大、平均隶属函数值分别给出如下:

$$\_{}^{m}_{X}(x_6)^{0.6}=\min\left\{\frac{3}{4},\frac{3}{5},\frac{1}{3}\right\}=\frac{1}{3}$$

$$\_{}^{M}_{X}(x_6)^{0.6}=\max\left\{\frac{3}{4},\frac{3}{5},\frac{1}{3}\right\}=\frac{3}{4}$$

$$\_{}^{*}_{X}(x_6)^{0.6}=\mathrm{avg}\left\{\frac{3}{4},\frac{3}{5},\frac{1}{3}\right\}=\frac{101}{180}$$

5 结 语

一般而言,元素之间的关系的传递性是不必要的. 对于元素之间的非等价关系,在文献 [10]中也可以被看做一种特殊的弱模糊相似关系.

本文引入了弱模糊相似关系的概念. 相似度被用来构造和解释论域的覆盖. 从相似类引入覆盖得到了标准的广义粗糙集近似. 研究了两对上下近似算子,讨论了最大、最小和平均 3种粗糙隶属函数及其性质.

参考文献:

[1] PAWLAK Z. Rough sets [J]. **Int J Inf and Comput Sci**, 1982, **11**(5): 341-356

[2] 张文修, 吴伟志. 粗糙集合理论介绍和研究综述 [J]. 模糊系统与数学, 2000, **14**(4): 1-12

[3] PAWLAK Z. Rough sets, rough relations and rough functions [J]. **Fundament Inf**, 1996, **27**(2-3): 8-23

[4] 刘文奇, 吴从 . 相似关系粗糙集理论与相似关系信息系统 [J]. 模糊系统与数学, 2002, **16**(3): 50-58

[5] SLOWINSKI R, VANDERPOOTEN D. Similarity relation as a basis for rough approximations [C]// WANG P. **Advances in Machine Intelligence & Soft Computing**. Raleigh NC: Bookwrights, 1997: 17-33

[6] YAO Y Y. A comparative study of fuzzy set and rough sets [J]. **Int J Inf Sci**, 1998, **109**: 227-242

[7] YAO Y Y. Two views of the theory of rough sets in finite universe [J]. **Int J of Approximate Reasoning**, 1996, **15**: 291-317

[8] PAWLAK Z, SKOWRON A. Rough membership functions [M]// **Fuzzy Logic for the Management of Uncertainty**. New York: Wiley, 1994: 251-271

[9] YAO Y Y, ZHANG J P. Interpreting fuzzy membership functions in the theory of rough sets [C] // **Proceedings of RSCTC'00**. Berlin: Springer-Verlag, 2000: 50-57

[10] INTAN R, MUKAIDONO M, YAO Y Y. Generalization of rough sets with T-coverings of the universe induced by conditional probability relations [J]. **Bull Int Rough Set Soc**, 2001, **5**(1/2): 20-22

Generalization of rough sets based on weak fuzzy similarity relations

TIAN Hong[1,2], WANG Xiu-kun*[1]

(1.School of Electr. and Inf. Eng., Dalian Univ. of Technol., Dalian 116024, China;
2.Software Inst., Dalian Jiaotong Univ., Dalian 116028, China)

Abstract Rough set theory is developed based on the notion of equivalence relation, but the property of equivalence has limited its application fields, which may not provide a realistic description of real-world relationships between elements. The notion of weak fuzzy similarity relations, a generalization of fuzzy similarity relations, is used to represent indiscernibility of elements. At the same time, the notions of similarity degree and similarity class are presented. All basic concepts or rough set theory are extended. Upper and lower approximations of generalized rough set are defined by using similarity class of the universe induced by a weak fuzzy similarity relation. Two pairs of upper and lower approximate arithmetic operators are discussed. Three types of rough membership functions are defined and their properties are examined.

Key words weak fuzzy similarity relation; similarity degree; similarity class