第 29 卷 第 4 期
Vol. 29 No. 4

控 制 与 决 策
Control and Decision

2014 年 4 月
Apr. 2014

文章编号: 1001-0920 (2014) 04-0691-05 DOI: 10.13195/j.kzyjc.2012.1769

邻域系统的不确定性度量方法

唐朝辉, 陈玉明

(厦门理工学院 计算机信息与工程学院，福建 厦门 361024)

摘 要: 针对离散型数据系统的不确定性度量方法难以有效解决邻域系统不确定性度量的问题, 引入邻域粗糙集模型, 提出邻域精确度、邻域知识粒度和基于邻域知识粒度的近似精度等邻域系统不确定性度量方法, 进一步从理论上证明其有效性. 实验结果表明, 基于邻域知识粒度的近似精度具有更严格的单调性, 优于邻域近似精度的邻域系统对不确定性度量的效果.

关键词: 不确定性度量；邻域系统；粗糙集；知识粒度

中图分类号: TP18 **文献标志码:** A

Neighborhood system uncertainty measurement approaches

TANG Chao-hui，*CHEN Yu-ming*

(College of Computer and Information Engineering，Xiamen University of Technology，Xiamen 361024，China. Correspondent：TANG Chao-hui，E-mail：chhtang@xmut.edu.cn)

Abstract: Uncertainty measures for the discrete data system can not be effective for the measurement of uncertainty of neighborhood system. Therefore, the neighborhood rough set model is introduced to propose measure approaches such as neighborhood accuracy, neighborhood knowledge granularity as well as neighborhood approximation accuracy based on knowledge granularity. The effectiveness of the approaches are verified theoretically. Experimental results show that approximation accuracy based on knowledge granularity with stricter monotonicity outperforms the neighborhood approximation accuracy for the uncertainty measurement of the neighborhood system.

Key words: uncertainty measurement；neighborhood system；rough set theory；knowledge granulation

0 引 言

由 Pawlak[1-2]提出的粗糙集理论是处理不确定、不精确以及模糊数据的一种有效的工具, 已成功用于特征选择、模式识别、图像处理、知识发现和数据挖掘等诸多领域[3-8].

基于粗糙集理论的不确定性度量是描述系统分类能力的重要依据, 很多学者对此进行了研究. Pawlak[2]提出用精度和粗糙度的概念来度量信息系统的不确定性, 用近似精度和近似粗糙度来衡量决策系统的不确定性. 也有部分学者从其他不同的角度研究了系统的不确定性, 比如信息熵[9-10]、知识粒度[11-13]和近似质量[14-15]等都可以有效地用于粗糙系统的不确定性度量.

传统的粗糙集理论适用于离散型数据系统, 但由于只考虑等价类和等价关系, 不能有效适用于邻域系统. 而邻域系统贴近更多的现实应用场景, 邻域关系比等价关系更为通用[16], 文献 [17] 讨论了近似空间的相关性质. 邻域关系已成功用于不确定数据的属性约简、特征提取以及分类[18-21]. 本文通过深入研究离散型系统的不确定性度量方法, 将其扩展并适用于邻域系统. 通过引入邻域关系、邻域上近似和下近似等概念, 定义了邻域知识粒度、邻域精确度、邻域粗糙度、邻域近似精度、邻域近似粗糙度和基于邻域知识粒度的近似精度等概念, 在理论上证明其在度量邻域系统不确定性上的单调性, 并将其用于邻域系统的不确定性度量.

1 经典粗糙集

定义 1 $I=(U,A,V,f)$ 是一个信息系统. 其中: U 为一个非空实体集合; A 为非空属性集合; V 为所有属性的值域, 即 $V=\bigcup\limits_{a\in A}V_a$, 这里 V_a 表示属性 a 所

收稿日期: 2012-11-27；修回日期: 2013-09-24.

基金项目: 国家青年科学基金项目(61103246)；厦门理工学院人才科研启动项目(YKJ10036R).

作者简介: 唐朝辉(1983−), 男, 讲师, 硕士, 从事机器学习、粗糙集的研究；陈玉明(1977−), 男, 副教授, 博士, 从事粗糙集、粒计算的研究.

有可能取值的集合.

定义 2 给定信息系统 $I=(U,A,V,f)$ 以及属性子集 $B\subseteq A$, B 可以确定一个二元等价关系, 记为

$$\mathrm{IND}(B)=$$
$$\{(x,y)\in U\times U|\forall a\in B,f(x,a)=f(y,a)\}.$$

$\mathrm{IND}(B)$对应了U上的一个划分, 记为

$$U/\mathrm{IND}(B)=\bigotimes\{a\in B|U/\mathrm{IND}(\{a\})\},$$
$$M\bigotimes N=$$
$$\{X\bigcap Y|\forall M\in R,\forall N\in Y,X\bigcap Y\neq\varnothing\}.$$

包含元素 x 的划分记为 $[x]_{\mathrm{IND}(B)}$.

定义 3[1] 给定信息系统 $I=(U,A,V,f)$ 以及等价关系 $\mathrm{IND}(B)$, 对于 $\forall X\subseteq U$, X 的上近似集和下近似集分别计为 $\mathrm{Upper}_B(X)$ 和 $\mathrm{Lower}_B(X)$. 其中

$$\mathrm{Lower}_B(X)=\{x\in U|[x]_{\mathrm{IND}(B)}\in X\},$$
$$\mathrm{Upper}_B(X)=\{x\in U|[x]_{\mathrm{IND}(B)}\bigcap X\neq\varnothing\}.$$

定义 4[1] 假设 M, N 是信息系统 $I=(U,A,V,f)$ 上的等价关系, 则正域、负域可分别定义为

$$\mathrm{POS}_M(N)=\bigcup_{X\in U/\mathrm{IND}(N)}\mathrm{Lower}_M(X),$$
$$\mathrm{NEG}_M(N)=U-\mathrm{POS}_M(N).$$

Pawlak给出了信息系统或决策系统的4种不确定性度量方法.

定义 5[2] 给定信息系统 $I=(U,A,V,f)$, 子域 $X\subseteq U$ 以及属性子集 $P\subseteq A$, 则子域 X 相对于 P 的精确度定义为

$$\mathrm{ac}_P(X)=\frac{|\mathrm{Lower}_P(X)|}{|\mathrm{Upper}_P(X)|},$$

X 相对于 P 的粗糙度定义为 $r_P(X)=1-\mathrm{ac}_P(X)$.

定义 6[2] 给定决策系统 $I=(U,C\bigcup D,V,f)$, 决策属性 D 对论域的划分记为 $U/D=\{D_1,D_2,\cdots,D_m\}$, 对于 $\forall P\subseteq C,U/D$ 相对于 P 的近似精度和近似粗糙度分别定义为

$$\mathrm{acapp}_P(D)=\frac{\sum\limits_{D_i\in U/D}|\mathrm{Lower}_P(D_i)|}{\sum\limits_{D_i\in U/D}|\mathrm{Upper}_P(D_i)|},$$
$$\mathrm{rapp}_P(D)=1-\mathrm{acapp}_P(D).$$

2 邻域系统不确定性度量

经典粗糙集的系统不确定性度量局限于离散型数据的系统, 不能有效适用于邻域系统的不确定性度量. 下面将给出邻域系统不确定性度量的相关概念和性质, 并在理论上加以证明.

2.1 邻域关系与邻域类

在提出邻域系统不确定性度量方法之前, 先引入邻域系统、邻域上近似和邻域下近似等概念, 并阐述部分性质.

定义 7 定义 $I=(U,A,V,f,\delta)$ 为一个邻域信息系统. 其中: U 为一个非空实体集合; A 为非空属性集合; V 为所有属性的值域, 即 $V=\bigcup\limits_{a\in A}V_a$, V_a 表示属性 a 所有可能取值的集合; $\delta\,(0\leqslant\delta\leqslant1)$ 为邻域阈值. 定义 $I=(U,C\bigcup D,V,f,\delta)$ 为邻域决策系统. 其中: C 表示系统的条件属性, D 表示系统的决策属性.

定义 8[17] 给定邻域系统 $I=(U,A,V,f,\delta)$, 对于 $\forall x\in U$, $B\subseteq A$, 定义 x 在 B 上的 δ 邻域

$$n_B^\delta(x)=\{y|x,y\in U,\ D_B(x,y)\leqslant\delta\},$$

其中 $D_B(x,y)=\sum\limits_{a\in B}|a(x)-a(y)|$. 由此易知, 邻域 $n_B^\delta(x)$ 满足

$$x\in n_B^\delta(x),\ y\in n_B^\delta(x)\Leftrightarrow x\in n_B^\delta(y),$$
$$\bigcup_{x\in U}n_B^\delta(x)=U.$$

定义 9[18] 给定邻域系统 $I=(U,A,V,f,\delta)$, $\forall B\subseteq A$, B 决定了邻域阈值 δ 上的邻域关系

$$\mathrm{NR}_\delta(B)=\{(x,y)\in U\times U|D_B(x,y)\leqslant\delta\}.$$

$U/\mathrm{NR}_\delta(B)$ 是 U 上基于 B 的划分, 每个划分称为一个邻域类或邻域知识. 邻域关系满足自反性、对称性, 是一种相似关系, 而非等价关系. 当 $\delta=0$ 时, 邻域关系变为等价关系, 但等价关系不能有效用于连续数据处理, 只适用于离散数据.

定理 1 给定邻域信息系统 $I=(U,A,V,f,\delta)$, $P,Q\subseteq A$ 以及 $x\in U$, 有以下结论: 1) 如果 $P\subseteq Q$, 则 $n_P^\delta(x)\supseteq n_Q^\delta(x)$; 2) 如果 $0\leqslant\alpha\leqslant\beta\leqslant1$, 则 $n_P^\alpha(x)\subseteq n_P^\beta(x)$.

证明: 1) 对于 $\forall x,y$, 如果 $P\subseteq Q$, 则有

$$D_P(x,y)=\sum_{a\in P}|(a(x)-a(y))|\leqslant$$
$$\sum_{a\in P}|(a(x)-a(y))|+\sum_{a\in(Q-P)}|(a(x)-a(y))|=$$
$$D_Q(x,y),$$

即如果 $y\in n_Q^\delta(x)$ 成立, 则 $y\in n_P^\delta(x)$ 成立, 即 $n_P^\delta(x)\supseteq n_Q^\delta(x)$. 由此结论 1) 得证.

2) 同理, 结论 2) 易证. □

定义 10[19] 设 $I=(U,A,V,f,\delta)$ 为一个邻域信息系统, $\forall B\subseteq A$, $X\subseteq U$, 则称

$$\mathrm{Lower}_B^\delta(X)=\bigcup\{x|x\in U/\mathrm{NR}_\delta(B),\ x\in X\}$$

为 X 关于 B 的邻域下近似集, 称

$$\mathrm{Upper}_B^\delta(X)=\bigcup\{x|x\in U/\mathrm{NR}_\delta(B),\ x\bigcap X\neq\varnothing\}$$

为 X 关于 B 的邻域上近似集.

2.2　基于邻域近似精度的不确定度量方法

下面定义邻域精确度、邻域粗糙度、邻域近似精度和邻域近似粗糙度等邻域系统不确定性度量方法的相关概念, 然后证明其相关性质.

定义 11　定义 $I=(U,A,V,f,\delta)$ 为一个邻域信息系统, 给定子域 $X\subseteq U$ 和属性子集 $P\subseteq A$, 子域 X 相对于 P 的邻域精确度定义为

$$\mathrm{ac}_P^\delta(X)=\frac{|\mathrm{Lower}_P^\delta(X)|}{|\mathrm{Upper}_P^\delta(X)|},$$

邻域粗糙度为 $r_P^\delta(X)=1-\mathrm{ac}_P^\delta(X)$.

定理 2　设 $I=(U,A,V,f,\delta)$ 为一个邻域信息系统, 给定 $X\subseteq U$, 且 $P\subseteq Q\subseteq A$, 则有 $\mathrm{ac}_P^\delta(X)\leqslant \mathrm{ac}_Q^\delta(X)$, $r_P^\delta(X)\geqslant r_Q^\delta(X)$.

证明　因为 $P\subseteq Q\subseteq A$, 由定理 1 可知, 对于 $\forall x\in X$, 有 $n_P^\delta(x)\supseteq n_Q^\delta(x)$. 由下近似定义可得

$$\mathrm{Lower}_P^\delta(X)\subseteq \mathrm{Lower}_Q^\delta(X),$$

即

$$|\mathrm{Lower}_P^\delta(X)|\leqslant|\mathrm{Lower}_Q^\delta(X)|.$$

同理可得 $|\mathrm{Upper}_P^\delta(X)|\geqslant|\mathrm{Upper}_Q^\delta(X)|$. 由此易得

$$\frac{|\mathrm{Lower}_P^\delta(X)|}{|\mathrm{Upper}_P^\delta(X)|}\leqslant\frac{|\mathrm{Lower}_Q^\delta(X)|}{|\mathrm{Upper}_Q^\delta(X)|},$$

即 $\mathrm{ac}_P^\delta(X)\leqslant\mathrm{ac}_Q^\delta(X)$, 由此定理 2 得证. □

定理 3　$I=(U,A,V,f,\delta)$ 为一个邻域信息系统, 给定 $X\subseteq U$, $P\subseteq A$, 且 $0\leqslant\alpha\leqslant\beta\leqslant 1$, 则有 $\mathrm{ac}_P^\alpha(X)\leqslant\mathrm{ac}_P^\beta(X)$, $r_P^\alpha(X)\leqslant r_P^\beta(X)$.

参照定理 2 的证明过程即可证明定理 3, 此略.

定义 12　定义邻域决策系统为 $\mathrm{ND}=(U,C\bigcup D,V,f,\delta)$, D 为决策属性, 将论域划分为 $U/D=\{D_1,D_2,\cdots,D_m\}$. 给定 $P\subseteq C$, U/D 相对于 P 的邻域近似精度和邻域近似粗糙度分别定义为

$$\mathrm{acapp}_P^\delta(D)=\frac{\sum\limits_{D_i\in U/D}|\mathrm{Lower}_P^\delta(D_i)|}{\sum\limits_{D_i\in U/D}|\mathrm{Upper}_P^\delta(D_i)|},$$

$$\mathrm{rapp}_P^\delta(D)=1-\mathrm{acapp}_P^\delta(D).$$

邻域近似精度刻画了邻域系统有效知识的完全程度, 值越大, 表明系统包含的有效知识的比例越大, 系统分类能力越强.

定理 4　设 $\mathrm{ND}=(U,C\bigcup D,V,f,\delta)$ 为一个邻域决策系统, 且 $P\subseteq Q\subseteq C$, 则有 $\mathrm{acapp}_P^\delta(U/D)\leqslant\mathrm{acapp}_Q^\delta(U/D)$, $\mathrm{rapp}_P^\delta(U/D)\geqslant\mathrm{rapp}_Q^\delta(U/D)$.

证明　设决策属性 D 将 U 划分为 $U/D=\{D_1, D_2,\cdots,D_m\}$, 由定理 2 的证明过程可知, 对于 $\forall D_i\in U/D$, 有

$$|\mathrm{Lower}_P^\delta(D_i)|\leqslant|\mathrm{Lower}_Q^\delta(D_i)|,$$

$$|\mathrm{Upper}_P^\delta(D_i)|\geqslant|\mathrm{Upper}_Q^\delta(D_i)|,$$

因此有

$$\sum_{D_i\in U/D}|\mathrm{Lower}_P^\delta(D_i)|\leqslant\sum_{D_i\in U/D}|\mathrm{Lower}_Q^\delta(D_i)|,$$

$$\sum_{D_i\in U/D}|\mathrm{Upper}_P^\delta(D_i)|\geqslant\sum_{D_i\in U/D}|\mathrm{Upper}_Q^\delta(D_i)|.$$

由 acapp 的定义可得 $\mathrm{acapp}_P^\delta(U/D)\leqslant\mathrm{acapp}_Q^\delta(U/D)$, 同理 $\mathrm{rapp}_P^\delta(U/D)\geqslant\mathrm{rapp}_Q^\delta(U/D)$. □

定理 5　设 $\mathrm{ND}=(U,C\bigcup D,V,f,\delta)$ 为一个邻域决策系统, $P\subseteq C$, 且 $0\leqslant\alpha\leqslant\beta\leqslant 1$, 则有

$$\mathrm{acapp}_P^\alpha(U/D)\geqslant\mathrm{acapp}_P^\beta(U/D),$$

$$\mathrm{rapp}_P^\alpha(U/D)\leqslant\mathrm{rapp}_P^\beta(U/D).$$

参照定理 4 的证明过程即可证明定理 5, 此略.

例 1　给定一个邻域决策系统, 如表 1 所示. 其中: a_i 为条件属性, Deci 为决策属性. 利用邻域近似精度对其进行不确定性度量的结果如图 1 所示.

表 1　一个邻域决策系统

U	属　性								
	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	Deci
x_1	0.71	0.76	0.44	0.59	0.33	0.43	0.74	0.30	4
x_2	0.74	0.47	0.23	0.38	0.68	0.83	0.32	0.41	2
x_3	0.26	0.48	0.47	0.79	0.41	0.63	0.75	0.52	1
x_4	0.29	0.46	0.51	0.39	0.27	0.22	0.57	0.60	5
x_5	0.90	0.79	0.61	0.35	0.76	0.13	0.39	0.50	3
x_6	0.49	0.58	0.75	0.64	0.59	0.19	0.71	0.13	3
x_7	0.43	0.81	0.8	0.33	0.23	0.38	0.05	0.63	3
x_8	0.21	0.44	0.27	0.61	0.71	0.57	0.55	0.52	2
x_9	0.28	0.29	0.41	0.53	0.45	0.82	0.18	0.36	4
x_{10}	0.56	0.5	0.53	0.28	0.69	0.71	0.40	0.72	3
x_{11}	0.33	0.85	0.41	0.45	0.09	0.63	0.18	0.75	3
x_{12}	0.34	0.3	0.13	0.48	0.41	0.51	0.36	0.54	1
x_{13}	0.68	0.47	0.37	0.29	0.10	0.90	0.88	0.45	5
x_{14}	0.33	0.51	0.59	0.37	0.37	0.38	0.63	0.56	5
x_{15}	0.60	0.62	0.75	0.65	0.49	0.24	0.44	0.48	2

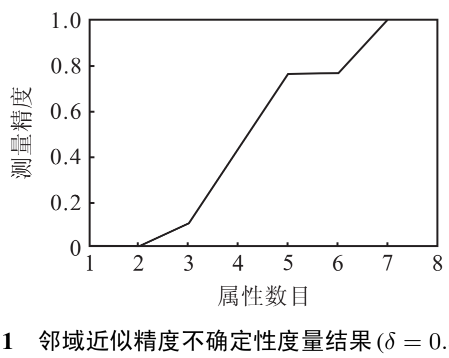

图 1　邻域近似精度不确定性度量结果 ($\delta=0.35$)

由图 1 可见, 当属性数目从 1 增加到 2, 从 5 增加到 6 时, 系统的近似精度值不变, 即系统的分类能力不变, 这样的结果不符合实际情况. 下面将对邻域近似精度的度量方法进行改进.

2.3 基于邻域知识粒度的近似精度

本节通过引入邻域知识粒度的概念来克服邻域近似精度在邻域系统不确定性度量上的不足.

定义 13 设 $\mathrm{ND}=(U,C\bigcup D,V,f,\delta)$ 为一个邻域决策系统, $P\subseteq C$, $x_i\in U$, $n_P^\delta(x_i)$ 为 U 上的邻域关系, 则 P 在 ND 上的邻域知识粒度定义为

$$\mathrm{KG}_\delta(P)=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|n_P^\delta(x_i)|.$$

邻域知识粒度刻画了系统中有效知识的平均粒度大小, 值越大表明系统平均知识的粒度越大, 分类能力越差.

定理 6 $\mathrm{ND}=(U,C\bigcup D,V,f,\delta)$ 为一个邻域决策系统, $P\subseteq C$, 则有 $0\leqslant \mathrm{KG}_\delta(P)\leqslant 1$.

证明 由 $\mathrm{KG}_\delta(P)$ 的定义可知 $\mathrm{KG}_\delta(P)\geqslant 0$. 对于 $\forall x_i\in U$, 有 $|n_P^\delta(x_i)|\leqslant|U|$, 则

$$\mathrm{KG}_\delta(P)\leqslant\frac{1}{|U|^2}\sum_{i=1}^{|U|}|U|=\frac{|U|^2}{|U|^2}=1,$$

由此定理 6 得证. □

定理 7 $\mathrm{ND}=(U,C\bigcup D,V,f,\delta)$ 为一个邻域决策系统, $P\subseteq Q\subseteq C$, 则有 $\mathrm{KG}_\delta(P)\geqslant KG_\delta(Q)$.

证明 已知 $P\subseteq Q$, 对于 $x\in U$, 由定理 1 可知, $|n_P^\delta(x)|\geqslant|n_Q^\delta(x)|$, 则

$$\mathrm{KG}_\delta(P)=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|n_P^\delta(x_i)|\geqslant$$

$$\frac{1}{|U|^2}\sum_{i=1}^{|U|}|n_Q^\delta(x_i)|=\mathrm{KG}_\delta(Q),$$

由此定理 7 得证. □

定理 8 $\mathrm{ND}=(U,C\bigcup D,V,f,\delta)$ 为一个邻域决策系统, $P\subseteq C$, 且 $0\leqslant\alpha\leqslant\beta\leqslant 1$, 则有 $\mathrm{KG}_\alpha(P)\leqslant\mathrm{KG}_\beta(P)$.

参考定理 7 的证明过程即可证明定理 8, 此略.

定义 14 $\mathrm{ND}=(U,C\bigcup D,V,f,\delta)$ 为一个邻域决策系统, $P\subseteq C$, 定义基于邻域知识粒度的近似精度为

$$\mathrm{NKGAA}_\delta(P)=(1-\mathrm{KG}_\alpha(P))\times(\mathrm{acapp}_P^\delta(U/D)+\lambda).$$

其中: λ 为平滑因子, 当 $\mathrm{acapp}_P^\delta(U/D)=0$ 时, 度量结果也可体现邻域知识粒度的贡献.

基于邻域知识粒度的近似精度在描述系统不确定性时, 既考虑了邻域系统中有效知识所占的比重, 也考虑了系统中有效知识的平均粒度的大小, 从而可更严格地描述邻域系统的知识分类能力.

定理 9 设 $\mathrm{ND}=(U,C\bigcup D,V,f,\delta)$ 为一个邻域决策系统, 且 $P\subseteq Q\subseteq C$, 则有 $\mathrm{NKGAA}_\delta(P)\leqslant\mathrm{NKGAA}_\delta(Q)$.

3 实 验

为了验证本文提出的邻域系统不确定性度量方法的有效性, 采用 3 个邻域决策系统进行验证, 分别如表 1～表 3 所示, 每个系统包含若干条件属性 a_i 和一个决策属性 Deci.

3 个邻域系统不确定性度量结果如图 2～图 4 所示. 图中: 实线代表邻域近似精度, 虚线代表基于邻域知识粒度的近似精度. 由图 2～图 4 可知, 当邻域系统不确定性度量所选的条件属性数从 1 增加到 2, 图 2 中的条件属性数目从 5 增加到 6, 以及图 4 中的条件属性数目从 4 增加到 5 时,邻域近似精度未能有效区分系统在属性数目不同时系统分类能力的不同, 但基于邻域知识粒度的近似精度可更精确地刻画系统属性数目不同时系统分类能力的不同.

表 2 一个邻域决策系统

U	属性						
	a_1	a_2	a_3	a_4	a_5	a_6	Deci
x_1	0.21	0.63	0.31	0.60	0.50	0.44	1
x_2	0.65	0.49	0.79	0.62	0.15	0.64	1
x_3	0.48	0.21	0.36	0.42	0.22	0.63	1
x_4	0.51	0.21	0.31	0.79	0.48	0.52	2
x_5	0.46	0.76	0.61	0.47	0.45	0.38	2
x_6	0.51	0.42	0.28	0.68	0.16	0.56	1
x_7	0.37	0.22	0.42	0.41	0.52	0.69	3
x_8	0.35	0.82	0.06	0.41	0.19	0.41	1
x_9	0.31	0.61	0.74	0.54	0.16	0.40	1
x_{10}	0.46	0.59	0.19	0.53	0.26	0.83	3

表 3 一个邻域决策系统

U	属性							
	a_1	a_2	a_3	a_4	a_5	a_6	a_7	Deci
x_1	0.73	0.61	0.63	0.58	0.38	0.53	0.56	3
x_2	0.61	0.14	0.64	0.47	0.12	0.55	0.30	1
x_3	0.49	0.57	0.41	0.72	0.95	0.63	0.25	2
x_4	0.26	0.57	0.45	0.39	0.49	0.47	0.31	4
x_5	0.40	0.75	0.32	0.53	0.28	0.77	0.52	1
x_6	0.75	0.64	0.12	0.67	0.8	0.37	0.62	4
x_7	0.20	0.5	0.65	0.50	0.96	0.55	0.31	2
x_8	0.61	0.41	0.84	0.23	0.74	0.31	0.74	1
x_9	0.34	0.93	0.34	0.27	0.35	0.12	0.46	2
x_{10}	0.64	0.4	0.33	0.42	0.47	0.35	0.61	3
x_{11}	0.61	0.48	0.58	0.80	0.46	0.29	0.49	4
x_{12}	0.66	0.40	0.52	0.68	0.55	0.11	0.53	4
x_{13}	0.58	0.49	0.11	0.23	0.15	0.47	0.46	2

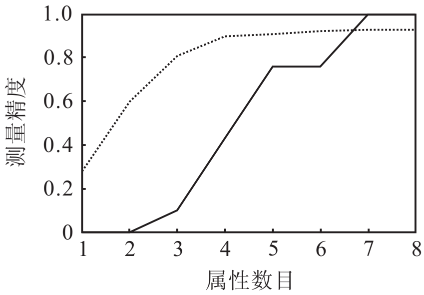

图 2 表 1 系统不确定性度量结果 ($\delta=0.35$)

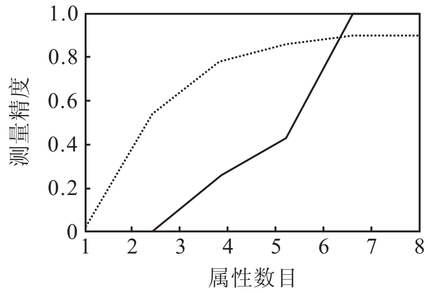

图 3 表 2 系统不确定性度量结果 ($\delta = 0.30$)

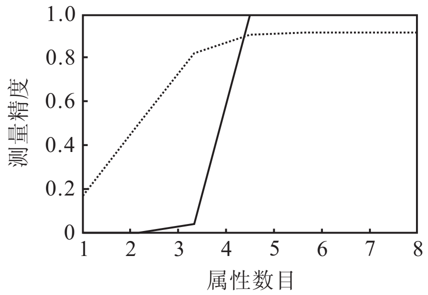

图 4 表 3 系统不确定性度量结果($\delta = 0.30$)

同时, 由图 2~图 4 可知, 当系统条件属性个数为 1 和 2 时, 邻域近似精度度量的结果都为 0, 未能体现少数条件属性对系统分类能力的贡献, 而基于邻域知识粒度的近似精度可有效地克服这种不足.

4 结 论

综上所述, 由于同时考虑了邻域系统中的有效知识的比重以及知识的平均粒度的大小, 基于邻域知识粒度的近似精度度量方法比邻域近似精度方法更能精确地刻画邻域系统的不确定性, 具有更严格的单调性, 从而能更准确地体现系统分类能力随着系统条件属性个数的增加而增加的性质.

参考文献(References)

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

[2] Pawlak Z. Rough sets[M]. Dordrecht: Kluwer Academic Publishers, 1991: 45-64.

[3] Qian Y H, Liang J Y, Pedrycz W, et al. Positive approximation: An accelerator for attribute reduction in rough set theory[J]. Artificial Intelligence, 2010, 174(9/10): 597-618.

[4] Wang D, Miao D Q, Xie C. Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection[J]. Expert Systems with Applications, 2011, 38(11): 14314-14320.

[5] Chen Y M, Miao D Q, Wang R Z. A rough set approach to feature selection based on ant colony optimization[J]. Pattern Recognition Letters, 2010, 31(3): 226-233.

[6] Min F, He H P, Qian Y H, et al. Test-cost-sensitive attribute reduction[J]. Information Science, 2011, 181(22): 4928-4942.

[7] Zhou J, Pedrycz W, Miao D Q. Shadowed sets in the characterization of rough-fuzzy clustering[J]. Pattern Recognition, 2011, 44(8): 1738-1749.

[8] Chen Y M, Miao D Q, Zhang H Z. Neighborhood outlier detection[J]. Expert Systems with Applications, 2010, 37(12): 8745-8749.

[9] Shannon C E. The mathematical theory of communication[J]. Bell System Technical J, 1948, 27(7): 379-423.

[10] Liang J, Shi Z, Li D, et al. Information entropy, rough entropy and knowledge granulation in incomplete information systems[J]. Int J of General System, 2006, 35(6): 641-654.

[11] Bianucci D, Cattaneo G. Information entropy and granulation co-entropy of partitions and coverings: A summary[C]. Trans on Rough Sets. Berlin: Springer, 2009: 15-66.

[12] 苗夺谦, 范世栋. 知识的粒度计算及其应用[J]. 系统工程理论与实践, 2002, 22(1): 48-56.
(Miao D Q, Fan S D. The calculation of knowledge granulation and its application[J]. Systems Engineering-Theory & Practice, 2002, 22(1): 48-56.)

[13] 王国胤, 张清华. 不同知识粒度下粗糙集的不确定性研究[J]. 计算机学报, 2008, 31(9): 1588-1598.
(Wang G Y, Zhang Q H. Uncertainty of rough sets in different knowledge granularities[J]. Chinese J of Computers, 2008, 31(9): 1588-1598.)

[14] Liang J Y, Li R, Qian Y H. Distance: A more comprehensible perspective for measures in rough set theory[J]. Knowledge-Based Systems, 2012, 27(11): 126-136.

[15] Dai J H, Xu Q. Approximations and uncertainty measures in incomplete information systems[J]. Information Sciences, 2012, 198(9): 62-80.

[16] Lin T Y. Neighborhood systems and relational database[C]. Proc of 1988 ACM 16th Annual Computer Science Conf. New York, 1988: 23-25.

[17] Yao Y Y. Relational interpretations of neighborhood operators and rough set approximation operators[J]. Information Sciences, 1998, 111(1-4): 239-259.

[18] Hu Q H, Yu D R, Xie Z X, et al. Fuzzy probabilistic approximation spaces and their information measures[J]. IEEE Trans on Fuzzy Systems, 2006, 14(2): 191-201.

[19] Hu Q H, Yu D R, Xie Z X. Neighborhood classifiers [J]. Experts Systems with Applications, 2008, 34(2): 866-876.

[20] Jensen R, Shen Q. Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches[J]. IEEE Trans of Knowledge and Data Engineering, 2004, 16(12): 1457-1471.

[21] Lin G P, Qian Y H, Li J J. Neighborhood-based multigranulation rough sets[J]. Int J of Approximate Reasoning, 2012, 53(7): 1080-1093.

(责任编辑: 滕 蓉)