计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.190500174

基于边界域的邻域知识距离度量模型

杨 洁[1,2] 王国胤[1] 李 帅[1]
1 重庆邮电大学计算智能重庆市重点实验室 重庆 400065
2 遵义师范学院物理与电子科学学院 贵州 遵义 563002
(530966074@qq.com)

摘 要 粗糙集的不确定性度量在知识获取中扮演着非常重要的角色。在邻域粗糙集理论中,当前不确定性度量方面的研究工作主要专注于度量单个知识空间的不确定性及其随粒度变化的单调性规律,其仍存在以下缺点:1)邻域粗糙集不确定性来自于邻域粒中属于目标概念的元素和不属于目标概念的元素,当前的方法没有同时考虑每个邻域信息粒的这两部分;2)不能反映不同知识空间对目标概念刻画能力的差异性;3)由于当前的知识距离包含了粒度划分的信息,已有方法在一些应用场合下不够准确,例如属性约简中的知识启发式搜索及其粒度选择。对此,文中首先构建了一种更加直观准确的邻域粗糙集的不确定性度量方法——邻域熵,并证明了不确定性度量随着粒度的细化具有单调性;为了反映不同邻域信息粒对目标概念刻画能力的差异性,提出了一种带近似描述能力的邻域粒距离,称为相对邻域粒距离,并介绍了它的相关性质;针对分层递阶的多粒度知识空间中的粒度选择问题,建立了基于边界域的邻域知识距离度量模型,该知识距离可以反映不同邻域知识空间对目标概念的刻画能力的差异性。

关键词:不确定性度量;邻域粗糙集;相对邻域粒距离;知识距离;邻域熵
中图法分类号 TP311

Neighborhood Knowledge Distance Measure Model Based on Boundary Regions

YANG Jie[1,2], WANG Guo-yin[1] and LI Shuai[1]
1 Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Post and Telecommunications, Chongqing 400065, China
2 School of Physics and Electronic Science, Zunyi Normal University, Zunyi, Guizhou 563002, China

Abstract Uncertainty measure of rough sets plays an important role in knowledge acquisition. In neighborhood rough sets, the current researches on uncertainty measure mainly focus on measuring the uncertainty of a single knowledge space and its monotonicity with the changing granularities. However, there are still some shortcomings. Firstly, the uncertainty of neighborhood rough set comes from elements belonging to target concept and elements not belonging to target concept in neighborhood granules, but current researches do not consider the two parts of each neighborhood information granule at the same time. Secondly, the difference between different knowledge spaces for describing the target concept is hard to reflect. Thirdly, the current knowledge distance measures are too fine, which contains granularity information and is inaccurate in some applications, i.e. heuristic search in attribute reduction. Therefore, based on the granularity measure of neighborhood information granules, this paper constructed the neighborhood entropy which is monotonic with the granularity being finer. In order to reflect the difference between different neighborhood information granule for describing the target concept, this paper proposed a neighborhood granule distance with approximate description ability, which is called relative neighborhood granule distance (RNGD). Then, several important properties were presented. The neighborhood knowledge distance based on boundary regions was established based on the RNGD, which can reflect the difference between different neighborhood knowledge spaces for describing the target concept. Finally, the validity of neighborhood knowledge distance based on decision regions was verified by experiments.

Keywords Uncertainty measure, Neighborhood rough sets, Relative neighborhood granule distance, Knowledge distance, Neighborhood entropy

---

到稿日期:2019-05-30 返修日期:2019-07-22 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(61572091,61472056);贵州省高层次创新人才项目(遵市科合人才[2018]15);贵州省教育厅科技人才成长项目(黔教合 KY(2018)318)
This work was supported by the National Science Foundation of China (61572091, 61472056), High level Innovative Talents Project of Guizhou Province and Science([2018]15) and technology talent growth project of Guizhou Province(KY(2018)318).
通信作者:王国胤(wanggy@cqupt.edu.cn)

1 引言

粒计算是一种模拟人类多粒度观察和处理问题的计算范式[1-5]。当前在粒计算的框架下有不同的粒化模型，如模糊集[6]、粗糙集[7]、商空间[8]、云模型[9]，这些模型从不同的视角采用不同的机制对数据进行粒化，实现了不确定性知识的转化和表示。粗糙集是一种利用当前信息粒来处理不确定性信息的有效工具，即采用上、下近似集对不确定性目标概念进行近似描述。为了处理连续数据，Lin[10]首次提出了邻域系统(NS)的概念。基于邻域系统，Yao[11-12]提出了邻域粗糙集的概念。当前，有许多研究者围绕邻域粗糙集展开了相关研究并取得了许多成果。为了处理信息系统中的异构数据，Hu等[13-14]首次将粗糙集与分类器相结合，构建了基于邻域系统的分类器。通过将邻域关系引入决策粗糙集，Li等[15]提出了基于邻域关系的决策粗糙集。Yang等[16]介绍了一种处理不完备信息系统的邻域粗糙集理论框架。为了解决当前大数据中数据标签获取代价高的问题，Wang等[17]首次提出了局部邻域粗糙集模型。除此之外，邻域粗糙集已被应用于许多领域，包括属性约简[18-19]、图像分类[20]、脑电波信号处理[21]、光学信号处理[22]及其他领域[23-24]。

在粗糙集理论中，不确定性度量在知识获取中发挥着非常重要的作用。对于邻域粗糙集而言，当前不确定性度量方面的研究成果非常有限，主要包括邻域精度、信息含量、邻域熵和基于熵的粗糙度等邻域系统的不确定性度量模型[25-28]。但是，这些度量模型仅考虑知识粒度或边界域大小，忽略了边界域的结构信息。再者，两个邻域知识空间具有相同的不确定性，并不意味着它们等价，当前的不确定性度量模型很难刻画它们之间的差异性。为了解决这个问题，Qian等[29-31]首次提出了知识距离来刻画不同知识空间之间的差异性。基于Qian的研究，当前已有一些学者在知识距离方面进行了相关研究[32-35]。例如，Chen等[33]提出了基于邻域粒距离的KNN分类器，为从粒计算的角度设计分类器提供了一种新思路。但是，以上的多数知识距离仅能反映不同知识空间之间的差异性，不能刻画它们对目标概念近似能力的差异性。再者，在一些应用场合中，当前的知识距离过于精细，从不确定性度量的角度而言，无法反映知识空间的不确定性差异。为了同时解决以上两个问题，本文建立了基于决策域的邻域知识距离度量模型，该知识距离可以反映不同邻域知识空间对目标概念的刻画能力的差异性。

本文第2节简要介绍了相关的基本概念，包括粗糙集、邻域粗糙集、知识粒度等；第3节提出了一种邻域粗糙集的不确定性度量方法——邻域熵；第4节介绍了相对邻域粒距离及其相关性质，并在此基础上提出了基于边界域的邻域知识空间距离；最后总结全文并展望未来。

2 相关定义

在介绍粗糙模糊集的不确定性度量之前，首先回顾一些基本概念。

定义1(粗糙集)[7] 假设$K=(U,\mathbb{R})$是一个知识基，其中$R\in\mathbb{R}$，$X\subseteq U$，那么X的上、下近似集定义为：

$$\overline{R}(X)=\{x\in U|[x]_R\cap X\neq\varnothing\} \tag{1}$$

$$\underline{R}(X)=\{x\in U|[x]_R\subseteq X\} \tag{2}$$

其中，$[x]_R$代表由等价关系U/R诱导的等价类，即$U/R=\{[x]_R\}=\{[x]_1,[x]_2,\cdots,[x]_m\}$。

本文中，一个划分空间U/R通常也叫做一个知识空间或粒度空间。简单而言，为了防止混淆，假设$[x]_R\triangleq[x]$。如果$\overline{R}(X)=\underline{R}(X)$，则$X$是一个可定义集，否则$X$是一个粗糙集。在粗糙集中，论域$U$通常被划分为正域、负域和边界域，其定义分别为：

$$POS_R(X)=\underline{R}(X) \tag{3}$$

$$NEG_R(X)=U-\overline{R}(X) \tag{4}$$

$$BND_R(X)=\overline{R}(X)-\underline{R}(X) \tag{5}$$

定义2(邻域粗糙集)[13-14] 假设$K=(U,\mathbb{R})$是一个知识基，其中$R\in\mathbb{R}$，$X\subseteq U$，那么X的上、下近似集定义为：

$$\overline{NR}(X)=\{x|\delta_R(x)\cap X\neq\varnothing,x\in U\} \tag{6}$$

$$\underline{NR}(X)=\{x|\delta_R(x)\subseteq X,x\in U\} \tag{7}$$

其中，$\delta_R(x)=\{x_j|d(x,x_j)\leqslant\delta\}$；$d(\cdot,\cdot)$表示距离度量，如欧氏距离；$\delta$表示邻域半径。

定义3(邻域粒距离)[29-31] 假设U是一个有限论域，$X\subseteq U$是一个目标集合。$\forall s,t\in U$，$\delta(s)$和$\delta(t)$是两个邻域粒，它们之间的邻域粒距离可定义为：

$$d(\delta(s),\delta(t))=\frac{|\delta(s)\oplus\delta(t)|}{|U|} \tag{8}$$

其中，$|\delta(s)\oplus\delta(t)|=|\delta(s)\bigcup\delta(t)|-|\delta(s)\bigcap\delta(t)|$。

由定义3可知，邻域粒距离可以实现信息粒的相似性度量。

例1 假设$U=\{x_1,x_2,x_3,x_4,x_5\}$，$X=\{x_2,x_3,x_5\}$是一个目标集合。$\delta(x_1)=\{x_1,x_2,x_3\}$和$\delta(x_2)=\{x_2,x_3,x_4,x_5\}$是两个邻域粒。

$$d(\delta(x_1),\delta(x_2))=\frac{|\delta(x_1)\oplus\delta(x_2)|}{|U|}=\frac{3}{5}$$

由此可见，$d(\delta(x_1),\delta(x_2))$的值仅与$\delta(x_1)$和$\delta(x_2)$两个邻域信息粒的大小有关，与目标概念$X$无关，因此$d(\delta(x_1),\delta(x_2))$不能反映$\delta(x_1)$和$\delta(x_2)$对目标概念近似能力的差异性度量。在本文中，$d(\delta(s),\delta(t))$被称为绝对邻域粒距离。

定义4(邻域知识距离)[29-31] 假设$K=(U,\mathbb{R})$是一个知识基，$P,Q\in\mathbb{R}$，$X\subseteq U$是一个目标集合。$K_P=\{\delta_P(x_i)|x_i\in U\}$和$K_Q=\{\delta_Q(x_i)|x_i\in U\}$是$K$上的两个知识空间，那么$K_P$和$K_Q$之间的基于边界域的邻域知识距离可定义为：

$$D(K_P,K_Q)=\frac{1}{|U|}\sum_{x\in U}d(\delta_P(x_i),\delta_Q(x_i)) \tag{9}$$

例2 假设$K=(U,\mathbb{R})$是一个知识基，$P,Q\in\mathbb{R}$，$X=\{x_2,x_3,x_4\}$是一个目标集合。$K_P=\{\{x_1,x_2,x_3\},\{x_1,x_2,x_5\},\{x_3,x_4\},\{x_3,x_4\},\{x_1,x_2,x_5\}\}$和$K_Q=\{\{x_1,x_2\},\{x_1,x_2\},\{x_3\},\{x_4\},\{x_5\}\}$是$K$上的两个知识空间。

$$D(K_P,K_Q)=\frac{1}{5}\sum_{x\in U}\left(\frac{1}{5}+\frac{1}{5}+\frac{1}{5}+\frac{1}{5}+\frac{2}{5}\right)=\frac{6}{25}$$

由此可见，$D(K_P,K_Q)$的值仅与 K_P 和 K_Q 两个邻域知识空间的粒度划分有关，与目标概念 X 无关，因此 $D(K_P,K_Q)$不能反映 K_P 和 K_Q 对目标概念近似能力的差异性度量。

3　邻域熵度量模型

粗糙集的不确定性分别来自于边界域中属于目标概念 X 的元素和边界域中不属于目标概念 X 的元素，对于邻域粗糙集来说，其不确定性度量模型应该同时考虑每个邻域信息粒的这两部分。因此，基于文献[35]提出的不确定性度量模型，本文提出了一种适用于邻域粗糙集的不确定性度量模型——邻域熵，其定义如下。

定义 5　假设 $K=(U,\mathbb{R})$是一个知识基，$P\in\mathbb{R}$，$X\subseteq U$是一个目标集合。$K_P=\{\delta_P(x_i)\mid x_i\in U\}$是 K 上的一个知识空间，那么 K_P 的邻域熵可定义为：

$$NE_{K_P}=-\frac{1}{|U|^2\ln 2}\sum_{x_i\in U}|\delta_P(x_i)|[\mu_{\delta_P(x_i)}\ln\mu_{\delta_P(x_i)}+(1-\mu_{\delta_P(x_i)})\ln(1-\mu_{\delta_P(x_i)})] \tag{10}$$

在定义 5 中，$\mu_{\delta_P(x_i)}$代表邻域粒 $\delta_P(x_i)$对于目标概念 X 的隶属度。从概率统计的角度来说，$\mu_{\delta_P(x_i)}$表示属于目标概念 X 的概率，$1-\mu_{\delta_P(x_i)}$代表邻域粒 $\delta_P(x_i)$不属于目标概念 X 的概率。式(10)分别由 $\mu_{\delta_P(x_i)}\ln\mu_{\delta_P(x_i)}$和$(1-\mu_{\delta_P(x_i)})\ln(1-\mu_{\delta_P(x_i)})$两部分信息熵构成，其中，前者刻画属于 X 的元素包含的不确定性，后者刻画不属于 X 的元素包含的不确定性，同时考虑这两部分才能更加准确地反映邻域粗糙集的不确定性。

定理 1　假设 $K=(U,\mathbb{R})$是一个知识基，$P,Q\in\mathbb{R}$，$X\subseteq U$是一个目标集合。$K_P=\{\delta_P(x_i)\mid x_i\in U\}$和 $K_Q=\{\delta_Q(x_i)\mid x_i\in U\}$是 K 上的两个知识空间，且 $K_Q\prec K_P$，则 $NE_{K_P}>NE_{K_Q}$。

证明：由于 $K_Q\prec K_P$，$\forall x_i\in U$，$\delta_Q(x_i)\subseteq\delta_P(x_i)$，为了简化证明，本文设 K_P 中只有 $\delta_P(s)$随着知识空间的细化而发生变化，因此 $\delta_Q(s)\subseteq\delta_P(s)$；$K_P$ 中其他邻域粒不发生变化。

(1)当 $\delta_P(s)\cap X=\varnothing$ 时，对于任意的 $x(x\in\delta_P(s))$，$|\delta_P^*(s)|=0$，故 $\mu_{\delta_P(s)}=0$；由于 $\delta_Q(x_i)\subseteq\delta_P(x_i)$，故 $\delta_Q(s)\cap X=\varnothing$，可得 $\mu_{\delta_Q(s)}=0$，因此 $NE_{K_Q}=NE_{K_P}$。

(2)当 $\delta_P(s)\subseteq X$ 时，对于任意的 $x(x\in\delta_P(s))$，$|\delta_P^*(s)|=|\delta_P(s)|$，故 $\mu_{\delta_P(s)}=1$；由于 $\delta_Q(x_i)\subseteq\delta_P(x_i)$，因此 $|\delta_Q^*(s)|=|\delta_Q(s)|$，可得 $\mu_{\delta_Q(s)}=1$，故 $NE_{K_Q}=NE_{K_P}$。

(3)$\delta_P(s)\cap X\neq\varnothing$，且 $\delta_P(s)\cap X\neq\delta_P(s)$时，设 $|\delta_P(s)\cap X|=a$，$|\delta_P(s)|-|\delta_P(s)\cap X|=b$，则 $|\delta_P(s)|[\mu_{\delta_P(s)}\ln\mu_{\delta_P(s)}+(1-\mu_{\delta_P(s)})\ln(1-\mu_{\delta_P(s)})]=-a\ln\frac{a}{a+b}-b\ln\frac{b}{a+b}>0$。

情形 1　若 $\delta_Q(s)\cap X=\varnothing$，则 $NE_{K_Q}=0$，故 $NE_{K_P}>NE_{K_Q}$。

情形 2　若 $\delta_Q(s)\subseteq X$，则 $NE_{K_Q}=0$，故 $NE_{K_P}>NE_{K_Q}$。

情形 3　若 $\delta_Q(s)\cap X\neq\varnothing$ 且 $\delta_Q(s)\cap X\neq\delta_Q(s)$，令 $|\delta_Q(s)\cap X|=a_1>0$，$|\delta_Q(s)|-|\delta_Q(s)\cap X|=b_1>0$，则 $-|\delta_Q(s)|[\mu_{\delta_Q(s)}\ln\mu_{\delta_Q(s)}+(1-\mu_{\delta_Q(s)})\ln(1-\mu_{\delta_Q(s)})]=-a_1\ln\frac{a_1}{a_1+b_1}-b_1\ln\frac{b_1}{a_1+b_1}$。

令函数 $f(a,b)=-\frac{a}{a+b}\ln\frac{a}{a+b}-\frac{b}{a+b}\ln\frac{b}{a+b}$，因为$\frac{\partial f}{\partial b}=\ln\frac{a+b}{b}>0$，$\frac{\partial f}{\partial a}=\ln\frac{a+b}{a}>0$，所以 $f(a,b)$是关于 a,b 的增函数，因此$-a\ln\frac{a}{a+b}-b\ln\frac{b}{a+b}>-a_1\ln\frac{a_1}{a_1+b_1}-b_1\ln\frac{b_1}{a_1+b_1}$，故 $NE_{K_P}>NE_{K_Q}$。

(续例 2)$NE_{K_P}=-\frac{1}{25\ln 2}[9(\frac{2}{3}\ln\frac{2}{3}+\frac{1}{3}\ln\frac{1}{3})]=0.33$，$NE_{K_Q}=-\frac{1}{25\ln 2}[4(\frac{1}{2}\ln\frac{1}{2}+\frac{1}{2}\ln\frac{1}{2})]=0.16$，因此 $NE_{K_P}>NE_{K_Q}$。

通过以上分析可知，相对于传统的邻域粗糙集不确定性度量，邻域熵包含了属于 X 的元素和不属于 X 的元素两方面的不确定性，更加直观准确；再者，由定理 1 可知，随着知识空间的细化，邻域粗糙集的边界域上的信息粒不断发生细分，邻域熵将严格递减，这符合人们对不确定性问题的认知规律。

4　基于边界域的邻域知识距离

4.1　相对邻域粒距离

定义 4 中的邻域信息粒距离实现了不同信息粒之间的差异性度量，但是从粗糙集的角度出发，当考虑目标概念时，邻域信息粒距离无法刻画信息粒之间对目标概念近似描述能力的差异性度量。为了解决这个问题，本文提出了一种带近似描述能力的邻域信息粒距离。

文献[35]提出了一种模糊信息粒距离，即假设 $G_{\widetilde{P}}(s)$与 $G_{\widetilde{Q}}(s)$分别是两个由模糊二元关系$\widetilde{P}$和 $\widetilde{Q}$ 产生的模糊信息粒，那么它们的差异性可以通过如下公式进行刻画：

$$\widetilde{d}(G_{\widetilde{P}}(s),G_{\widetilde{Q}}(s))=\frac{|G_{\widetilde{P}}(s)\cup G_{\widetilde{Q}}(s)|-|G_{\widetilde{P}}(s)\cap G_{\widetilde{Q}}(s)|}{|U|} \tag{11}$$

本文将式(11)具体化到邻域粗糙集中，提出了带近似描述能力的相对邻域粒公式。本文中，为了方便描述，用$\delta^*(s)$表示 $\delta(s)$对应的带近似描述能力的邻域粒，称其为相对邻域粒，即考虑了邻域粒 $\delta(s)$对目标概念的刻画能力。例如，假设 $U=\{x_1,x_2,x_3,x_4,x_5\}$，$X=\{x_2,x_3,x_5\}$是一个目标集合，$\delta(x_1)=\{x_1,x_2,x_3\}$，则 $\delta^*(x_1)=\frac{0}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}$。

定义 6(相对邻域粒距离)　假设 U 是一个有限论域，$X\subseteq U$是一个目标集合，对于 $\forall s,t\in U$，$\delta(s)$和 $\delta(t)$是两个邻域粒，它们之间的邻域粒距离可定义为：

$$\Delta(\delta(s),\delta(t))=\frac{|\delta^*(s)\cup\delta^*(t)|-|\delta^*(s)\cap\delta^*(t)|}{|U|} \tag{12}$$

其中，$\delta^*(s)$和 $\delta^*(t)$分别表示邻域粒 $\delta(s)$和 $\delta(t)$分别近似描述目标概念 X 时对应的模糊邻域粒，本文称 $\delta^*(s)$和 $\delta^*(t)$为 $\delta(s)$和 $\delta(t)$对应的相对邻域粒，且 $|\delta^*(s)\cup\delta^*(t)|=\sum_{x\in U}\mu_{\delta^*(s)\cup\delta^*(t)}(x)$，$|\delta^*(s)\cap\delta^*(t)|=\sum_{x\in U}\mu_{\delta^*(s)\cap\delta^*(t)}(x)$。

(续例 1)由条件可知，$X=\{x_2,x_3,x_5\}$，因此，$\delta^*(x_1)=$

$\frac{0}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}$，$\delta^*(x_2)=\frac{1}{x_2}+\frac{1}{x_3}+\frac{0}{x_4}+\frac{1}{x_5}$，则 $\Delta(\delta(x_1),\delta(x_2))=\frac{3-2}{5}=\frac{1}{5}$。

例 1 中，当 $X=\frac{0.2}{x_1}+\frac{0.8}{x_2}+\frac{1}{x_3}+\frac{0.3}{x_4}+\frac{0.9}{x_5}$ 时，$\delta^*(x_1)=\frac{0.2}{x_1}+\frac{0.8}{x_2}+\frac{1}{x_3}$，$\delta^*(x_2)=\frac{0.8}{x_2}+\frac{1}{x_3}+\frac{0.3}{x_4}+\frac{0.9}{x_5}$，则 $\Delta(\delta(x_1),\delta(x_2))=\frac{3.2-1.8}{5}=0.28$。

因此，式(12)同样适用于目标概念为模糊集时的情形。

性质 1 假设 U 是一个有限论域，$X\subseteq U$ 是一个目标集合，对于 $\forall s\in U$，$\delta(s)$ 是一个邻域粒，则 $\mu_{\delta(s)}=\frac{\sum\limits_{x\in\delta(s)}\Delta(\delta(s),x)}{\sum\limits_{x\in\delta(s)}d(\delta(s),x)}$。

证明：由于 $\mu_{\delta(s)}=\frac{|\delta^*(s)|}{|\delta(s)|}$，因此 $\frac{\sum\limits_{x\in\delta(s)}\Delta(\delta(s),x)}{\sum\limits_{x\in\delta(s)}d(\delta(s),x)}=\frac{\sum\limits_{x\in\delta(s)}\mu(x)-\mu(x)}{|\delta(s)||\delta(s)-1|}=\frac{|\delta^*(s)||\delta(s)-1|}{|\delta(s)||\delta(s)-1|}=\frac{|\delta^*(s)|}{|\delta(s)|}$。由性质 1 可知，邻域粒 $\delta(s)$ 的均值隶属度可以从邻域距离的角度进行刻画，即 $\Delta(\delta^*(s),x)$ 越大或 $d(\delta(s),x)$ 越小，则 $\mu_{\delta(s)}$ 越大，反之亦然。

引理 1 假设 U 是一个有限论域，则 $\Delta(\cdot,\cdot)$ 是 U 上的一个距离。

定理 2 假设 U 是一个有限论域，$X\subseteq U$ 是一个目标集合，对于 $\forall s,t\in U$，$\delta(s)$ 和 $\delta(t)$ 是两个邻域粒，则以下性质成立：

(1)如果 $\delta(s),\delta(t)\in POS(X)$，$\Delta(\delta(s),\delta(t))=d(\delta(s),\delta(t))$；

(2)如果 $\delta(s),\delta(t)\in NEG(X)$，$\Delta(\delta(s),\delta(t))=0$；

(3)如果 $\delta(s),\delta(t)\in BND(X)$，$\Delta(\delta(s),\delta(t))\leqslant d(\delta(r),\delta(s))$。

证明：由于 $\delta(s),\delta(t)\in POS(X)$，$|\delta^*(s)|=|\delta(s)|$ 且 $|\delta^*(t)|=|\delta(t)|$，因此 $\Delta(\delta(s),\delta(t))=d(\delta(s),\delta(t))$。

由于 $\delta(s),\delta(t)\in NEG(X)$，$|\delta^*(s)|=|\delta^*(t)|=0$，因此 $\Delta(\delta(s),\delta(t))=0$。

令 $\delta(s)=A$，$\delta(t)=B$，则 $\delta^*(s)=A\cap X$，$\delta^*(t)=B\cap X$，只需证明 $|A\cup B|-|A\cap B|\geqslant|(A\cap X)\cup(B\cap X)|-|(A\cap X)\cap(B\cap X)|$。由于 $A\cap B\subseteq A\cup B$，$(A\cap X)\cap(B\cap X)\subseteq(A\cap X)\cup(B\cap X)$，因此证明 $(A\cap X)\cup(B\cap X)-(A\cap X)\cap(B\cap X)\subseteq(A\cup B)-(A\cap B)$ 即可。

由于：

$$\begin{aligned}&(A\cap X)\cup(B\cap X)-(A\cap X)\cap(B\cap X)\\&=(A\cup B)\cap X-(A\cap B\cap X)\\&=(A\cup B)\cap X\cap(\neg A\cup\neg B\cup\neg X)\\&=(A\cup B)\cap[(\neg A\cap X)\cup(\neg B\cap X)]\\&=(A\cup B)\cap(\neg A\cup\neg B)\cap X)\\&=[(A\cup B)-(A\cap B)]\cap X\subseteq(A\cup B)-(A\cap B)\end{aligned}$$

因此，$\Delta(\delta(s),\delta(t))\leqslant d(\delta(r),\delta(s))$。

定理 3 假设 U 是一个有限论域，$X\subseteq U$ 是一个目标集合，对于 $\forall r,s,t\in U$，$\delta(r)$，$\delta(s)$ 和 $\delta(t)$ 是 3 个邻域粒，如果 $\delta(t)\subseteq\delta(s)\subseteq\delta(r)$，则 $\Delta(\delta(r),\delta(t))=\Delta(\delta(r),\delta(s))+\Delta(\delta(s),\delta(t))$。

证明：如果 $\delta(t)\subseteq\delta(s)\subseteq\delta(r)$，可得：

$|\delta^*(r)\cup\delta^*(s)|=|\delta^*(r)|$，$|\delta^*(s)\cup\delta^*(t)|=|\delta^*(s)|$，$|\delta^*(r)\cup\delta^*(t)|=|\delta^*(r)|$

$|\delta^*(r)\cap\delta^*(s)|=|\delta^*(s)|$，$|\delta^*(s)\cap\delta^*(t)|=|\delta^*(t)|$，$|\delta^*(r)\cap\delta^*(t)|=|\delta^*(t)|$

则：

$$\begin{aligned}&\Delta(\delta(r),\delta(s))+\Delta(\delta(s),\delta(t))\\&=\frac{|\delta^*(r)\cup\delta^*(s)|-|\delta^*(r)\cap\delta^*(s)|}{U}+\\&\quad\frac{|\delta^*(s)\cup\delta^*(t)|-|\delta^*(s)\cap\delta^*(t)|}{U}\\&=\frac{|\delta^*(r)|-|\delta^*(t)|}{U}\\&=\Delta(\delta(r),\delta(t))\end{aligned}$$

因此，$\Delta(\delta(r),\delta(t))=\Delta(\delta(r),\delta(s))+\Delta(\delta(s),\delta(t))$。

4.2 邻域知识距离

为了更准确地反映邻域知识空间之间的差异性，在相对邻域粒距离的基础上，本节进一步建立了基于决策的邻域知识距离。

例 3 假设 $K=(U,\mathbb{R})$ 是一个知识基，$P,Q\in\mathbb{R}$，$X=\{x_2,x_3,x_4\}$ 是一个目标集合；$K_P=\{\{x_1,x_2,x_5\},\{x_3,x_4\},\{x_3,x_4\},\{x_1,x_2,x_5\}\}$ 和 $K_Q=\{\{x_1,x_2,x_5\},\{x_1,x_2,x_5\},\{x_3\},\{x_4\},\{x_1,x_2,x_5\}\}$ 是 K 上的两个知识空间。由式(10)可得：$NE_{K_P}=NE_{K_Q}=0.33$。由式(9)可得：$D(K_P,K_Q)=\frac{2}{25}$。由此可见，$D(K_P,K_Q)$ 的值仅与 K_P 和 K_Q 两个知识空间的划分有关，与目标概念 X 无关。因此，$D(K_P,K_Q)$ 不能反映 K_P 和 K_Q 对目标概念近似能力的差异性度量。

再者，由于 $K_Q\prec K_P$，因此 $BND_P(X)=BND_Q(X)=\{x_1,x_2,x_5\}$，$POS_P(X)=POS_Q(X)=\{x_3,x_4\}$，且 $NE_{K_P}=NE_{K_Q}$。从粒度选择的角度来说，K_P 和 K_Q 两个知识空间的重要程度是相同的。但是，$D(K_P,K_Q)=0.08\neq0$，从不确定性的角度不能准确地反映 K_P 和 K_Q 两个知识空间之间具有的差异性。

定义 7(基于边界域的邻域知识距离) 假设 $K=(U,\mathbb{R})$ 是一个知识基，$P,Q\in\mathbb{R}$，$X\subseteq U$ 是一个目标集合；$K_P=\{\delta_P(x_i)\mid x_i\in U\}$ 和 $K_Q=\{\delta_Q(x_i)\mid x_i\in U\}$ 是 K 上的两个知识空间，那么 K_P 和 K_Q 之间基于边界域的邻域知识距离可定义为：

$$ND_{\mathrm{BND}}(K_P,K_Q)=\frac{1}{|U|}\sum_{x_i\in\{BND_P(X)\cup BND_Q(X)\}}\Delta(\delta_P(x_i),\delta_Q(x_i)) \tag{13}$$

(续例 3)由于 $BND_P(X)=BND_Q(X)=\{x_1,x_2,x_5\}$，由式(6)可得：

$$ND_{\mathrm{BND}}(K_P,K_Q)=\frac{1}{5}\sum_{x_i\in\{x_1,x_1,x_5\}}\Delta(\delta_P(x_i),\delta_Q(x_i))=0$$

由以上例子可知，根据基于边界域的邻域知识距离的刻画结果，K_P 和 K_Q 两个知识空间的重要程度是相同的。基于边界域的邻域知识距离仅集中体现两个邻域知识空间对应的边界域结构信息的差异性，与模糊粒结构距离不同，其忽略了边界域的粒度划分信息，而这恰好是在粒度选择和属性约简中所需要的。众所周知，属性约简过程是在一个分层递阶的多粒度知识空间中进行渐进式知识空间选择的过程，从经典粗糙集的角度来说，由于边界域随着知识空间的细化不断减少，边界域中的元素分配到正域或负域，使得正域随着知识空间的细化而单调递增。因此，在分层递阶的多粒度知识空间中，通常只需要比较不同知识空间边界域的差异性。

（续例 2）由条件可知 $K_Q \prec K_P$，$BND_P(X)=\{x_1,x_2,x_5\}$，$BND_Q(X)=\{x_1,x_2\}$，$POS_P(X)=POS_Q(X)=\{x_3,x_4\}$，则 $ND_{\mathrm{BND}}(K_P,K_Q)=\frac{1}{5}\sum_{x_i\in\{x_1,x_2,x_5\}}\Delta(\delta_P(x_i),\delta_Q(x_i))=\frac{2}{25}$。

定理 4　假设 $K=(U,\mathbb{R})$ 是一个知识基，$P,Q,R\in\mathbb{R}$，X 是一个目标集合；$K_P=\{\delta_P(x_i)\mid x_i\in U\}$，$K_Q=\{\delta_Q(x_i)\mid x_i\in U\}$ 和 $K_R=\{\delta_R(x_i)\mid x_i\in U\}$ 是 K 上的 3 个知识空间，且 $K_R\prec K_Q\prec K_P$，则 $ND_{\mathrm{BND}}(K_P,K_R)=ND_{\mathrm{BND}}(K_P,K_Q)+ND_{\mathrm{BND}}(K_Q,K_R)$。

证明：由于 $K_R\prec K_Q\prec K_P$，对于 $\forall x_i\in U$，$\delta_R(x_i)\subseteq\delta_Q(x_i)\subseteq\delta_P(x_i)$，由定理 3 可得：

$$\begin{aligned}
&ND_{\mathrm{BND}}(K_P,K_Q)+ND_{\mathrm{BND}}(K_Q,K_R)\\
&=\frac{1}{|U|}\sum_{x_i\in\{BND_P(X)\cup BND_Q(X)\}}\Delta(\delta_P(x_i),\delta_Q(x_i))+\\
&\quad\frac{1}{|U|}\sum_{x_i\in\{BND_Q(X)\cup BND_R(X)\}}\Delta(\delta_Q(x_i),\delta_R(x_i))\\
&=\frac{1}{|U|}\sum_{x_i\in BND_P(X)}(\Delta(\delta_P(x_i),\delta_Q(x_i))+\Delta(\delta_Q(x_i),\\
&\quad\delta_R(x_i)))\\
&=\frac{1}{|U|}\sum_{x_i\in BND_P(X)}\Delta(\delta_P(x_i),\delta_R(x_i))\\
&=ND_{\mathrm{BND}}(K_P,K_R)
\end{aligned}$$

例 4　假设 $K=(U,\mathbb{R})$ 是一个知识基，$P,Q\in\mathbb{R}$，$X=\{x_2,x_3,x_4\}$ 是一个目标集合；$K_P=\{\{x_1,x_2,x_3\},\{x_1,x_2,x_5\},\{x_3,x_4\},\{x_3,x_4\},\{x_1,x_2,x_5\}\}$，$K_Q=\{\{x_1,x_2\},\{x_1,x_2\},\{x_3\},\{x_4\},\{x_5\}\}$ 和 $K_R=\{\{x_1\},\{x_1,x_2\},\{x_3\},\{x_4\},\{x_5\}\}$ 是 K 上的 3 个知识空间。

由条件可知，$BND_P(X)=\{x_1,x_2,x_5\}$，$BND_Q(X)=\{x_1,x_2\}$，$BND_R(X)=\{x_2\}$，因此：

$$ND_{\mathrm{BND}}(K_P,K_Q)=\frac{1}{5}\sum_{x\in\{x_1,x_2,x_5\}}(\frac{1}{5}+0+\frac{1}{5})=\frac{2}{25}$$

$$ND_{\mathrm{BND}}(K_Q,K_R)=\frac{1}{5}\sum_{x\in\{x_1,x_2\}}(\frac{1}{5}+0)=\frac{1}{25}$$

$$ND_{\mathrm{BND}}(K_P,K_R)=\frac{1}{5}\sum_{x\in\{x_1,x_2,x_5\}}(\frac{2}{5}+0+\frac{1}{5})=\frac{3}{25}$$

故 $ND_{\mathrm{BND}}(K_P,K_R)=ND_{\mathrm{BND}}(K_P,K_Q)+ND_{\mathrm{BND}}(K_Q,K_R)$。

结束语　针对邻域粗糙集理论中不确定性度量方面存在的一些问题，本文分别做了以下 3 个方面的工作：1）构建了一种更加直观准确的邻域粗糙集的不确定性度量方法——邻域熵；2）提出了一种带近似描述能力的邻域粒距离，其可以反映不同邻域信息粒对目标概念刻画能力的差异性；3）针对分层递阶的多粒度知识空间中的粒度选择问题，建立了基于边界域的邻域知识距离度量模型，该模型可以反映不同邻域知识空间对目标概念刻画能力的差异性。这些工作能很好地促进粗糙集不确定性理论的发展，在下一步的研究工作中，我们将利用本文的邻域知识距离设计出更有启发性的属性重要度表达式，以期取得更好的属性约简效果。

参 考 文 献

[1] PEDRYCZ W, ALHMOUZ R, MORFEQ A, et al. The design of free structure granular mappings: the use of the principle of justifiable granularity [J]. IEEE Transactions on Cybernetics, 2013, 43(6): 2105-2113.

[2] PEDRYCZ W, SKOWRON A, KREINOVICH V. Handbook of granular computing [M]. Wiley-Interscience, 2008: 719-740.

[3] YAO J T, VASILAKOS A V, PEDRYCZ W. Granular Computing: Perspectives and Challenges[J]. IEEE Transactions on Cybernetics, 2013, 43(6): 1977-1989.

[4] YAO Y Y. Perspectives of granular computing[C]// IEEE International Conference on Granular Computing. IEEE, 2005: 85-90.

[5] WANG G Y, YANG J, XU J. Granular computing: from granularity optimization to multi-granularity joint problem solving [J]. Granular Computing, 2017, 2(3): 1-16.

[6] ZADEH L A. Fuzzy sets[J]. Information and Control, 1965, 8(3): 338-353.

[7] PAWLAK Z. Rough sets[J]. International Journal of Computer Information Sciences, 1982, 11(5): 341-356.

[8] 张钹，张铃. 问题求解理论及应用[M]. 北京：清华大学出版社，1990.

[9] LI D Y, MENG H J. Membership and membership cloud generator[J]. Computer Research and Development, 1995(6): 15-20.

[10] LIN T Y. Neighborhood systems and relational databases[C]// Proceedings of the 1988 ACM Sixteenth Annual Conference on Computer Science. ACM, 1988: 725.

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

[12] YAO Y Y. Granular computing using neighborhood systems, advances in soft computing: engineering design and manufacturing [C]// The 3rd On-line World Conference on Soft Computing. London: Springer, 1999: 539-553.

[13] HU Q H, YU D, XIE Z. Neighborhood classifiers [J]. Expert Systems with Applications, 2008, 34(2): 866-876.

[14] HU Q H, YU D, LIU J, et al. Neighborhood rough set based heterogeneous feature subset selection [J]. Information Sciences, 2008, 178(18): 3577-3594.

[15] LI W, HUANG Z, JIA X, et al. Neighborhood based decision-theoretic rough set models[J]. International Journal of Approximate Reasoning, 2016, 69: 1-17.

[16] YANG X, MING Z, DOU H, et al. Neighborhood systems-based rough sets in incomplete information system[J]. Knowledge-Based Systems, 2011, 24(6): 858-867.
[17] WANG Q, QIAN Y H, LIANG X Y, et al. Local neighborhood rough set[J]. Knowledge-Based Systems, 2018, 153: 53-64.
[18] YONG L, HUANG W, JIANG Y, et al. Quick attribute reduct algorithm for neighborhood rough set model[J]. Information Sciences, 2014, 271(7): 65-81.
[19] CHEN Y M, ZENG Z, LU J. Neighborhood rough set reduction with fish swarm algorithm[J]. Soft Computing, 2016, 21(23): 1-12.
[20] YING Y, PEDRYCZ W, MIAO D. Neighborhood rough sets based multi-label classification for automatic image annotation [J]. International Journal of Approximate Reasoning, 2013, 54(9): 1373-1387.
[21] KUMAR S U, INBARANI H H. Neighborhood rough set based ECG signal classification for diagnosis of cardiac diseases[J]. Soft Computing, 2016, 21(16): 4721-4733.
[22] XIE H, TAN K, WANG L, et al. Hyperspectral band selection based on a variable precision neighborhood rough set[J]. Applied Optics, 2016, 55(3): 462.
[23] ZHONG Y, ZHANG X, SHAN F. Hybrid data-driven outlier detection based on neighborhood information entropy and its developmental measures[J]. Expert Systems with Applications, 2018, 112: 243-257.
[24] MENG J, JING Z, RUI L, et al. Gene selection using rough set based on neighborhood for the analysis of plant stress response [J]. Applied Soft Computing, 2014, 25(C): 51-63.
[25] CHEN Y M, WU K, CHEN X, et al. An entropy-based uncertainty measurement approach in neighborhood systems[J]. Information Sciences, 2014, 279: 239-250.
[26] CHEN Y M, XUE Y, MA Y, et al. Measures of uncertainty for neighborhood rough sets[J]. Knowledge-Based Systems, 2017, 120: 226-235.
[27] TANG Z H, CHEN Y M. Uncertainty measurement methods for neighborhood systems[J]. Control and Decision, 2014, 29(4): 691-695.
[28] ZHENG T, ZHU L. Uncertainty measures of Neighborhood System-based rough sets[J]. Knowledge-Based Systems, 2015, 86: 57-65.
[29] QIAN Y H, LIANG J Y, DANG C Y. Knowledge structure, knowledge granulation and knowledge distance in a knowledge base[J]. International Journal of Approximate Reasoning, 2009, 50: 174-188.
[30] LIANG J Y, RU L, QIAN Y H. Distance: A more comprehensible perspective for measures in rough set theory[J]. Knowledge-Based Systems, 2012, 27(3): 126-136.
[31] QIAN Y H, CHENG H, WANG J, et al. Grouping granular structures in human granulation intelligence[J]. Information Sciences, 2017, 382: 150-169.
[32] YANG X B, QIAN Y H, YANG J Y. On characterizing hierarchies of granulation structures via distances[J]. Fundamenta Informaticae, 2013, 123(3): 365-380.
[33] CHEN Y M, QIN N, LI W, et al. Granule structures, distances and measures in neighborhood systems[J]. Knowledge-Based Systems, 2018, 165: 268-281.
[34] YANG J, WANG G Y, ZHANG Q H. Knowledge Distance Measure in Multigranulation Spaces of Fuzzy Equivalence Relations[J]. Information Sciences, 2018, 448-449: 18-35.
[35] QIAN Y H, LI Y, LIANG J Y, et al. Fuzzy Granular Structure Distance[J]. IEEE Transactions on Fuzzy Systems, 2015, 23(6): 2245-2259.
[36] WANG G Y, ZHANG Q H. Uncertainty of Rough Sets in Different Knowledge Granularities[J]. Chinese Journal of Computers, 2008, 31(9): 1588-1598.

YANG Jie, born in 1987, Ph.D, associate professor. His research interests include data mining, machine learning, three-way decision and rough set.

WANG Guo-yin, born in 1970, Ph. D, professor. His research interests include data mining, machine learning, granular computing and rough set.