基于变精度上近似与程度下近似的双量化边界及其算法

张贤勇

(四川师范大学数学与软件科学学院 成都 610068) (同济大学计算机科学与技术系 上海 201804)

摘 要 近似空间中,精度与程度结合形成的双量化是一个创新课题。利用笛卡尔积进行量化信息合成,基于变精度上近似与程度下近似探讨双量化边界及其算法。首先,基于上述两个近似,自然地构建了双量化扩张粗糙集模型,定义了双量化扩张边界。接着,分析了该边界的双量化语义,得到了该边界的精确刻画与数学性质;为计算该边界,提出了近似集算法与信息粒算法,进行了算法分析与算法比较,得到了信息粒算法具有更优的算法空间复杂性的重要结论。最后,应用一个医疗实例对该边界及其算法进行了说明。该边界扩张了经典Pawlak边界,并对局部不确定性进行了双量化的完备与精细刻画,这对双量化的不确定性分析与应用具有重要意义。

关键词 粗糙集,粒计算,不确定性,双量化,边界

中图法分类号 TP18 **文献标识码** A

Double-quantitative Boundary and its Algorithms Based on Variable Precision Upper Approximation and Grade Lower Approximation

ZHANG Xian-yong

(College of Mathematics and Software Science, Sichuan Normal University, Chengdu 610068, China)

(Department of Computer Science and Technology, Tongji University, Shanghai 201804, China)

Abstract The double-quantification with precision and grade acts as a novel project in the approximate space. By the Cartesian-Product combination of quantitative information, this paper aimed to explore a double-quantitative boundary and its algorithms based on the variable precision upper approximation and grade lower approximation. First, a double-quantitative expansion-model was naturally constructed by the two approximations, and the double-quantitative expansion-boundary was correspondingly defined. Then, the double-quantitative semantics was analyzed for the boundary, and its precise description and mathematical properties were obtained, in order to calculate the boundary. The approximation-set algorithm and information-granule algorithm were proposed, analyzed and compared. The information-granule algorithm has more advantages on the space complexity. Finally, a medical example was provided to illustrate the boundary and its algorithms. The boundary expands the Pawlak-boundary, and makes the complete and fine double-quantitative descriptions for partial uncertainty, thus, it has a great significance for the uncertainty analyses and applications with respect to the double-quantification.

Keywords Rough set, Granular computing, Uncertainty, Double-quantification, Boundary

1 引言

粗糙集理论是一种研究不确定性的数学理论,是一种关于数据分析与知识推理的工具。其能够有效地分析不精确、不一致、不完整的数据信息,从中发现隐含知识,揭示潜在规律。当前,粗糙集理论在知识获取、机器学习、规则生成、决策分析、智能控制等领域获得了广泛的应用。经典模型(Pawlak-Model)[1]缺乏近似空间的定量信息,具有一定的局限性,如其不能有效地处理噪声数据,也不能完全地挖掘边界中的潜在知识。因此,Pawlak-Model需要被改进,量化扩张模型则具有重要价值。概率是刻画不确定性的重要工具,其被引入形成了概率粗糙集模型。该模型具有概率信息的可测性、模型的泛性与弹性、对噪声的不敏感性等突出的优点[2]。进而,概率粗糙集模型得到了广泛的发展,其包括了诸多具体的模型,如变精度粗糙集模型(VPRS-Model)[3]、决策粗糙集模型(DTRS-Model)[4]等。同时,程度粗糙集模型(GRS-Model)[5]也是一种重要的量化扩张模型。

通过引入量化参数并允许一定程度的错误分类率存在,VPRS-Model和GRS-Model完全扩张了Pawlak-Model,并且在很大程度上解决了信息系统的数据噪音问题,这对数据采集以及数据分析具有重要意义。目前,这两类模型具有许多研究成果,如文献[6-8]研究了VPRS-Model的知识约简与规则提取,文献[9]应用VPRS-Model于心理学,文献[10-12]则研究了GRS-Model的构建与性质。VPRS-Model和GRS-

到稿日期:2012-11-03 返修日期:2013-02-25 本文受国家自然科学基金(61203285,61273304,60970061,11071178),中国博士后科学基金(2012M520930),四川省教育厅重点项目(12ZA138)资助。

张贤勇(1978—),男,博士后,副教授,CCF会员,主要研究方向为粗糙集、粒计算、数据挖掘,E-mail:xianyongzh@sina.com。

Model 主要利用精度与程度对近似空间进行相对量化与绝对量化。精度量化与程度量化分别刻画相对比率与绝对数目，两者不等价，具有紧密的、互补的、辩证的关系。进而，精度与程度结合形成的双量化具有完备性优势，对近似空间的刻画更精确，但计算复杂性与单量化同线性级别。双量化能够对近似空间进行全面刻画，强化多级容错特性，推进已有模型的发展和应用，因此双量化的研究是一项具有重要创新性与深远意义的工作。此外，基于基本量化扩张特性，VPRS-Model 和 GRS-Model 具有很强的相似性、可比性、结合性，如变精度近似算子与程度近似算子的相似性，因此双量化研究具有科学的可行性。文献[13]强化了程度绝对量化的重要性，并阐述了精度、程度的互补性、完备性，进而比较研究了 VPRS-Model与 GRS-Model，得到了两者的关系和转化；该文为双量化研究奠定了重要基础。

粗糙集理论把那些无法确认的个体都归属于边界，而边界定义为上近似与下近似的差；进而边界具有确定的数学公式描述，不确定性区域可以被客观地计算。粗糙集理论的不确定性正是一种基于边界的不确定性，所以边界成为了粗糙集理论的一个极其重要的概念。文献[14,15]研究了基于边界的属性约简、特征提取。Pawlak-Model 中，Pawlak 边界停留在宏观定性的最大绝对层面，这很大地限制了粗糙集理论对不确定性的刻画与应用，因此边界的深入量化刻画具有重要意义。在粗糙集模型的量化扩张过程中，具有阈值容忍性的边界相对缩小，在阈值意义下不确定性相对降低。这种边界的量化研究，事实上量化刻画了 Pawlak 边界中的特定局部，进而通过阈值的系统层次，可以形成 Pawlak 边界中的层次量化结构。正是在此背景下，本文利用笛卡尔积进行量化信息合成，基于变精度上近似与程度下近似组建双量化扩张粗糙集模型，构造 Pawlak 边界内的局部双量化边界，并进行双量化研究。该双量化边界完全扩张了 Pawlak 边界，具有具体的精确刻画与特定的双量化语义。本文还提出了该双量化边界的两个计算算法，进行了算法分析与算法比较，也研究了其数学性质。

2 基于变精度上近似与程度下近似的边界

U 为非空有限论域，R 为其上的等价关系，称为知识，(U,R)称为近似空间，元素 x 所在等价类$[x]_R$ 称为信息粒。2^U 为 U 的幂集，集合 $A\in 2^U$，$c([x]_R,A)=1-|[x]_R\cap A|/|[x]_R|$称为$[x]_R$ 关于 A 的相对错误分类率，令 $p([x]_R,A)=1-c([x]_R,A)$。Pawlak 近似、变精度近似与程度近似分别为：

$$\overline{R}A=\bigcup\{[x]_R:[x]_R\cap A\neq\phi\}$$

$$\underline{R}A=\bigcup\{[x]_R:[x]_R\subseteq A\}$$

$$\overline{R}_\beta A=\bigcup\{[x]_R:c([x]_R,A)<1-\beta\}$$

$$\underline{R}_\beta A=\bigcup\{[x]_R:c([x]_R,A)\leqslant\beta\}$$

$$\overline{R}_k A=\bigcup\{[x]_R:|[x]_R\cap A|>k\}$$

$$\underline{R}_k A=\bigcup\{[x]_R:|[x]_R|-|[x]_R\cap A|\leqslant k\}$$

其中，β 称为分类误差，k 为非负整数，β、k 也称为阈值。简单地，$1-\beta$ 称为精度，k 称为程度；深刻地，精度与程度被根本地联系于量化：$p([x]_R,A)$、$c([x]_R,A)$和$|[x]_R\cap A|$、$|[x]_R|-|[x]_R\cap A|$[13]。一般情况下 $\beta\in[0,0.5)$；为了理论的完备性，本文采用文献[13]中扩张的范围，即 $\beta\in[0,1]$。近似集可以提取出近似算子，粗糙集模型可通过它们来描述，如 VPRS-Model 即为$(U,\overline{R}_\beta,\underline{R}_\beta)$。显然，基于近似的阈值扩张性，VPRS-Model 和 GRS-Model 量化扩张了 Pawlak-Model。

定义 1 $\overline{R}_\beta$ 为变精度上近似算子，$\underline{R}_k$ 为程度下近似算子，即

$$\overline{R}_\beta:2^U\to 2^U,\ \underline{R}_k:2^U\to 2^U,\ \forall A\in 2^U,\ \overline{R}_\beta(A)=\overline{R}_\beta A,\ \underline{R}_k(A)=\underline{R}_k A$$

它们共同决定的模型称为变精度上近似与程度下近似粗糙集模型，记为$(U,\overline{R}_\beta,\underline{R}_k)$。其中，

$$\forall A\in 2^U,\ bnR_{\beta k}(A)=bnR_{\beta k}A=\overline{R}_\beta A-\underline{R}_k A$$

$bnR_{\beta k}$称为该模型的边界算子，$bnR_{\beta k}A$ 称为集合 A 在该模型中的边界。

基于变精度上近似与程度下近似，定义 1 自然地构建了模型$(U,\overline{R}_\beta,\underline{R}_k)$。这一模型蕴含着精度与程度的双量化，即为一个双量化模型。更深刻地，基于上下近似的二维独立性，这里的双量化其实是利用笛卡尔积合成了近似空间的精度、程度量化信息。具体地，该模型具有关联于上下近似的双量化语义：模型$(U,\overline{R}_\beta,\underline{R}_k)$的上近似描述“关于集合 A 的错误分类率小于 $1-\beta$”的信息粒，下近似描述“不属于集合 A 的元素个数最多 k 个”的信息粒。基于笛卡尔积对于信息集成的完备性与可还原性等优点，模型$(U,\overline{R}_\beta,\underline{R}_k)$以特定角度很好地双量化描述了近似空间，具有理论价值与应用前景。其中的集合边界 $bnR_{\beta k}A$ 是基于边界的常规定义自然产生的，也利用笛卡尔积来蕴藏着重要的精度程度双量化信息。

命题 1 $\beta=0$、$k=0$ 时，模型$(U,\overline{R}_\beta,\underline{R}_k)$退化为 Pawlak-Model$(U,\overline{R},\underline{R})$，且 $bnR_{\beta k}A=\overline{R}A-\underline{R}A=bnRA$。

证明：由 $\beta=0$ 时 $\overline{R}_\beta A=\overline{R}A$、$k=0$ 时，$\underline{R}_k A=\underline{R}A$，易证。

命题 1 表明，基于变精度近似与程度近似对经典 Pawlak 近似的扩张性，模型$(U,\overline{R}_\beta,\underline{R}_k)$完全扩张了 Pawlak-Model $(U,\overline{R},\underline{R})$，其中的边界 $bnR_{\beta k}A$ 也完全扩张了经典 Pawlak 边界 $bnRA$。换言之，模型$(U,\overline{R}_\beta,\underline{R}_k)$即为一个双量化扩张粗糙集模型，$bnR_{\beta k}A$ 即为一个双量化扩张边界。

命题 2 $\forall\beta,k,\ bnR_{\beta k}A\subseteq bnR_{\overline{0}\underline{0}}A=bnRA$。

证明：由 $\forall\beta,k$ 有 $\overline{R}_\beta A\subseteq\overline{R}A$、$\underline{R}_k A\supseteq\underline{R}A$，及边界定义，可证。

命题 2 表明，一般情况下，边界 $bnR_{\beta k}A$ 比 Pawlak 边界 $bnRA$ 缩小了，即说明上述双量化扩张是往边界变小的方向施行的。边界的大小决定着不确定性的大小，边界绝对缩小则意味着不确定性减弱，确定性增强。因此，命题 2 表明，在阈值容忍性情况下，边界 $bnRA$ 相对缩小，不确定性相对降低；这正是模型$(U,\overline{R}_\beta,\underline{R}_k)$与边界 $bnR_{\beta k}A$ 进行了双量化扩张后的良性结果。一般情况下，边界 $bnR_{\beta k}A$ 为 Pawlak 边界的子集，故边界 $bnR_{\beta k}A$ 能够双量化刻画 Pawlak 边界中的特定局部，即部分不确定性区域得到了深刻的描述。同时，若阈值 β、k 进行不同层次的取值，则边界 $bnR_{\beta k}A$ 将在 Pawlak 边界内形成深刻的粒层次结构，进而不确定性区域将得到更深的刻

画；从粒计算角度来讲，这一结果奠定了层次问题描述与求解的基础。综上，边界 $bnR_{\underline{\beta}}^{\underline{k}}A$ 具有重要研究价值。

命题 3 $bnR_{\underline{\beta}}^{\underline{k}}A=\cup\{[x]_R:c([x]_R,A)<1-\beta,|[x]_R|-|[x]_R\cap A|>k\}$。

从精度与程度出发，命题 3 清晰地呈现了边界 $bnR_{\underline{\beta}}^{\underline{k}}A$ 的双量化基本形态。边界 $bnR_{\underline{\beta}}^{\underline{k}}A$ 本质上施行了精度与程度的复合刻画，并且这种双量化主要是基于笛卡尔积的，因此边界 $bnR_{\underline{\beta}}^{\underline{k}}A$ 对于精度程度双量化信息具有完备性。$bnR_{\underline{\beta}}^{\underline{k}}A$ 描述的是“关于集合 A 的错误分类率小于 $1-\beta$，但不属于集合 A 的元素个数多于 k 个”的信息粒。可见，边界 $bnR_{\underline{\beta}}^{\underline{k}}A$ 具有特定的双量化语义，进而具有双量化错误描述特征与容错机制。同时，结合命题 2，命题 3 本质上给出了 Pawlak 边界的子集或层次结构的精细双量化刻画。因此，边界 $bnR_{\underline{\beta}}^{\underline{k}}A$ 对双量化的不确定性分析与应用具有重要意义。

命题 4 $bnR_{\underline{\beta}}^{\underline{k}}A=\cup\{[x]_R:|[x]_R\cap A|>\beta|[x]_R|,|[x]_R\cap A|<|[x]_R|-k\}$。

基于命题 3、命题 4 进一步给出了边界 $bnR_{\underline{\beta}}^{\underline{k}}A$ 的一个计算公式。其中只涉及 $\beta|[x]_R|$、$|[x]_R|-k$ 两个量；下面，将通过讨论这两个量的关系，得到由阈值 β、k 决定的边界 $bnR_{\underline{\beta}}^{\underline{k}}A$ 的微观精确刻画。

命题 5 1）$\beta=1$ 时，$bnR_{\underline{\beta}}^{\underline{k}}A=\phi$；

2）$\beta\in[0,1)$ 时，$bnR_{\underline{\beta}}^{\underline{k}}A=\cup\{[x]_R:|[x]_R|>\frac{k}{(1-\beta)},\beta|[x]_R|<|[x]_R\cap A|<|[x]_R|-k\}$，其中 $|[x]_R|>\frac{k}{(1-\beta)}$ 为 $[x]_R\subseteq bnR_{\underline{\beta}}^{\underline{k}}A$ 的必要条件。

证明：1）$\beta=1$ 时，$\overline{R}_\beta A=\phi$，因此 $bnR_{\underline{\beta}}^{\underline{k}}A=\phi$；

2）$\beta|[x]_R|<|[x]_R|-k$ 时，即 $|[x]_R|>\frac{k}{(1-\beta)}$，$[x]_R\subseteq bnR_{\underline{\beta}}^{\underline{k}}A$ 可能成立；

$\beta|[x]_R|\geq|[x]_R|-k$ 时，即 $|[x]_R|\leq\frac{k}{(1-\beta)}$，$[x]_R\not\subset bnR_{\underline{\beta}}^{\underline{k}}A$。

所以 $bnR_{\underline{\beta}}^{\underline{k}}A=\cup\{[x]_R:|[x]_R|>\frac{k}{(1-\beta)},\beta|[x]_R|<|[x]_R\cap A|<|[x]_R|-k\}$。

命题 5 进一步对命题 4 得到的公式进行细化，从而得到了边界 $bnR_{\underline{\beta}}^{\underline{k}}A$ 的精确刻画公式。基于近似空间的基本核心信息：$|[x]_R|$、$|[x]_R\cap A|$，这种精确刻画公式处于描述结构的最底层，具有信息粒最直接的信息，因此对边界 $bnR_{\underline{\beta}}^{\underline{k}}A$ 的本质刻画与优化计算具有重要意义。双量化对单量化进行了提升，计算复杂性线性地提高；而双量化的计算优化，主要拟降低计算复杂性，故具有实用价值。综上，命题 5 为边界 $bnR_{\underline{\beta}}^{\underline{k}}A$ 的计算，特别是优化计算，奠定了基础。

3 边界的算法

命题 5 表明，$\beta=1$ 是一种特殊情况，此时边界 $bnR_{\underline{\beta}}^{\underline{k}}A$ 为空集。下面主要在 $\beta\in[0,1)$ 的一般情形下，探索计算边界 $bnR_{\underline{\beta}}^{\underline{k}}A$ 的两种不同算法，并进行算法分析与算法比较，给出优化计算的相关结论。

算法 1（近似集算法）

输入：$|[x]_R|$、$|[x]_R\cap A|$、阈值 β，k。

Step 1 计算变精度上近似 $\overline{R}_\beta A$ 和程度下近似 $\underline{R}_k A$；

Step 2 由集合的差（即定义 1），计算边界 $bnR_{\underline{\beta}}^{\underline{k}}A$。

输出：边界 $bnR_{\underline{\beta}}^{\underline{k}}A$。

算法 2（信息粒算法）

输入：$|[x]_R|$、$|[x]_R\cap A|$、阈值 β，k。

Step 1 由命题 5，判别基数在 $(\frac{k}{(1-\beta)},+\infty)$ 内的每个信息粒是否包含于边界 $bnR_{\underline{\beta}}^{\underline{k}}A$ 内；

Step 2 把包含于边界 $bnR_{\underline{\beta}}^{\underline{k}}A$ 内的信息粒合成集合。

输出：边界 $bnR_{\underline{\beta}}^{\underline{k}}A$。

近似集算法的计算过程非常清晰。在信息粒算法中，为了算法的确定性，需要细化算法的计算过程；主要是在比较大小时，需要确定参数的先后次序。在信息粒算法的第 1 步中，当 $|[x]_R|>\frac{k}{(1-\beta)}$ 时，$|[x]_R\cap A|$ 与参数比较的次序规定为：先 $\beta|[x]_R|$ 后 $|[x]_R|-k$。

这两个算法的核心部分是：判断每个信息粒是否包含于特定集合。每个信息粒都需要 2 个输入数据：$|[x]_R|$、$|[x]_R\cap A|$，设信息粒共有 n 个，则需要 $2n$ 个输入数据。下面，选取与 $|[x]_R|$、$|[x]_R\cap A|$、n 相关的运算（包括除、减、比较大小）作为基本计算，以信息粒的规模 n 作为实例特征，来分析与比较这两个算法。

在 VPRS-Model 中，为简化计算，主要使用公式 $\overline{R}_\beta A=\cup\{[x]_R:p([x]_R,A)>\beta\}$。近似集算法中，要判断每个信息粒是否包含于 $\overline{R}_\beta A$ 和 $\underline{R}_k A$ 内，需要计算 2 次，比较大小 2 次，需要辅助变量 2 个：$p([x]_R,A)$、$|[x]_R|-|[x]_R\cap A|$，其余是 1 步集合差运算。所以，近似集算法的时间、空间复杂性分别为：$T(n)=4n$、$S(n)=2n$，且结果是稳定不变的。

信息粒算法中，每个信息粒的基数要先与 $k/(1-\beta)$ 比较大小 1 次。1）若 $|[x]_R|\leq k/(1-\beta)$，则不需要比较大小和辅助变量，有结论 $[x]_R\not\subset bnR_{\underline{\beta}}^{\underline{k}}A$；2）若 $|[x]_R|>k/(1-\beta)$，则 $|[x]_R\cap A|$ 最多需要与 $\beta|[x]_R|$ 和 $|[x]_R|-k$ 比较大小 2 次，最多需要计算 2 次，需要 2 个辅助变量：$\beta|[x]_R|$、$|[x]_R|-k$。根据信息粒的分布及是否包含于 $bnR_{\underline{\beta}}^{\underline{k}}A$ 的归属来看，信息粒总共有 4 种情形。表 1 给出了信息粒的分布及其基本计算次数、辅助变量个数的详细情况。在最坏的情况下，信息粒算法的时间、空间复杂性分别为：$T(n)=5n$、$S(n)=2n$。

表 1 信息粒算法中信息粒的分布及其基本计算次数、辅助变量个数

情形	第 1 轮比较大小：$\|[x]_R\|$	第 2 轮比较大小：$\|[x]_R\cap A\|$	$\subseteq bnR_{\underline{\beta}}^{\underline{k}}A$	基本计算次数	辅助变量个数
(1)	$\leq k/(1-\beta)$	—	否	1	0
(2)	$>k/(1-\beta)$	$\leq\beta\|[x]_R\|$	否	3	1
(3)	$>k/(1-\beta)$	($\beta\|[x]_R\|$，$\|[x]_R\|-k$)	是	5	2
(4)	$>k/(1-\beta)$	$\geq\|[x]_R\|-k$	否	5	2

在最坏的情况下，两个算法的时间复杂性与空间复杂性的渐进分析相同：$T(n)=\Theta(n)$、$S(n)=\Theta(n)$，但信息粒算法比近似集算法更具有算法空间优势。近似集算法的时间和空间复杂性是固定不变的。但在信息粒算法中，情形(1)可以降低算法的时间复杂性和空间复杂性，情形(2)可以降低算法的

空间复杂性。显然，近似集算法的空间复杂性是信息粒算法空间复杂性的上界。同时，信息粒算法的辅助变量 $\beta|[x]_R|$、$|[x]_R|-k$ 优于近似集算法的辅助变量 $p([x]_R,A)$、$|[x]_R|-|[x]_R\cap A|$。每个信息粒均满足条件 $|[x]_R|\leqslant k/(1-\beta)$ 是信息粒算法的最好情况；此时，信息粒算法的时间复杂性和空间复杂性分别为：$T(n)=n$、$S(n)=c$。

近似集算法以变精度上近似和程度下近似的基本概念为中心，利用定义进行宏观计算，直接、简洁、常规、自然；信息粒算法则立足于信息粒这一原子微粒，利用边界 $bnR_{\leqslant}A$ 的精确刻画进行微观计算，更创新、更根本。以上算法分析表明，信息粒算法比近似集算法具有更优的空间复杂性。原因在于，命题 5 得到了边界 $bnR_{\leqslant}A$ 的精确刻画，得到了 $[x]_R\subseteq bnR_{\leqslant}A$ 的必要条件 $|[x]_R|>k/(1-\beta)$；同时通过算法分析可知，在情形(1)和(2)的信息粒分布情况下，算法空间复杂性会大大降低。因而，信息粒算法计算时，先由参数 β、k 分化出 $(0,k/(1-\beta)]$、$(k/(1-\beta),+\infty)$ 两个区间，再根据信息粒基数与上述区间的关系，进一步计算出信息粒是否包含于 $bnR_{\leqslant}A$，进而求得 $bnR_{\leqslant}A$。而近似集算法必须先确定信息粒的变精度上近似和程度下近似的两种集合归属，然后再用集合差运算转换。显然，信息粒算法适用于 k 偏大或 β 偏大的情形。在海量数据处理中，可能选用信息粒算法计算更快捷。此时可通过参数 k、β 的大小，$|[x]_R|$、$|[x]_R\cap A|$ 的排序及它们间关系的定位来决策是否采用信息粒算法。

4 医疗案例

这里，我们采用文献[13]中的医疗案例来说明边界 $bnR_{\leqslant}A$ 及其算法。信息系统 $S=(U,T,V,f)$，U 由 36 个观测病人组成，$T=\{r_1,r_2,r_3\}$，r_1、r_2、r_3 分别表示“发烧”、“头痛”和“感冒”，$V_{r_1}=\{0,1,2\}$、$V_{r_2}=\{0,1,2\}$、$V_{r_3}=\{0,1\}$，$V_{r_3}=\{0,1\}$ 中的 1、0 分别表示“有感冒”、“无感冒”。基于观测病人检测数据，表 2 给出了观测病人类别的统计数据，其中 R 为“发烧”、“头痛”决定的等价关系，$[x]_m=(i,j)(m=1,2,\cdots,9)$ 为相应的观测病人等价类别，A 表示“感冒病人”集合。

表 2 观测病人类别的统计数据

$[x]_m$：(i,j)	$[x]_m$ 元素	$\lvert[x]_m\rvert$	$[x]_m\cap A$ 元素	$\lvert[x]_m\cap A\rvert$	$c([x]_m,A)$	$\lvert[x]_m\rvert-\lvert[x]_m\cap A\rvert$
$[x]_1$：(0,0)	1、7、13、19、22、30、35	7	—	0	1	7
$[x]_2$：(0,1)	15、32	2	15	1	1/2	1
$[x]_3$：(0,2)	3、17、25	3	3	1	2/3	2
$[x]_4$：(1,0)	5	1	5	1	0	0
$[x]_5$：(1,1)	2、10、16、27、34	5	10、34	2	3/5	3
$[x]_6$：(1,2)	8、11、20、24、31	5	11、20、24	3	2/5	2
$[x]_7$：(2,0)	12、21、28、36	4	21、28	2	1/2	2
$[x]_8$：(2,1)	4、14、18、23、29、33	6	14、18、29、33	4	1/3	2
$[x]_9$：(2,2)	6、9、26	3	6、9、26	3	0	0

下面在 $\beta=0.35$、$k=2$ 的阈值条件下，分别用近似集算法与信息粒算法来计算边界 $bnR_{\leqslant}A$。

近似集算法　1)$\beta=0.35$ 时，$\overline{R}_\beta A=[x]_2\cup[x]_4\cup[x]_5\cup[x]_6\cup[x]_7\cup[x]_8\cup[x]_9$，$k=2$ 时，$\underline{R}_k A=[x]_2\cup[x]_3\cup[x]_4\cup[x]_6\cup[x]_7\cup[x]_8\cup[x]_9$；2)$bnR_{\overline{0.352}}A=[x]_5$。

信息粒算法　1)$\beta=0.35$、$k=2$，$\frac{k}{(1-\beta)}=3.08$，$[x]_2$、$[x]_3$、$[x]_4$、$[x]_9$ 满足条件 $|[x]_R|\leqslant 3.08$，其余信息粒的基数全属于区间 $(3.08,+\infty)$。故 $[x]_2$，$[x]_3$，$[x]_4$，$[x]_9\not\subset bnR_{\overline{0.352}}A$，其余信息粒全为怀疑粒。经计算有 $[x]_1$，$[x]_6$，$[x]_7$，$[x]_8\not\subset bnR_{\overline{0.352}}A$、$[x]_5\subseteq bnR_{\overline{0.352}}A$。详细计算参见表 3；2)进而 $bnR_{\overline{0.352}}A=[x]_5$。

根据“发烧”、“头痛”属性，论域划分为 9 种观测病人类别。上近似 $\overline{R}A=[x]_2\cup[x]_3\cup[x]_4\cup[x]_5\cup[x]_6\cup[x]_7\cup[x]_8\cup[x]_9$ 与下近似 $\underline{R}A=[x]_4\cup[x]_9$ 分别表示“可能和一定纳入感冒病人集合”的观测病人类别。而 $bnR_{\overline{0.352}}A=[x]_5$ 表示“关于感冒病人集合的错误分类率小于 0.65，但不属于感冒病人集合的观测病人个数多于 2 个”的观测病人类别。

显然，在 Pawlak-Model 中，Pawlak 边界 $bnRA=[x]_2\cup[x]_3\cup[x]_5\cup[x]_6\cup[x]_7\cup[x]_8$，$bnR_{\overline{0.352}}A\subseteq bnRA$。由本医疗案例可见，Pawlak 边界中的特定局部：边界 $bnR_{\leqslant}A$，利用精度和程度进行了双量化刻画，具有实际的双量化语义与双量化容错机制，对不确定性的深入刻画与量化应用具有实际意义。本医疗案例中信息粒数目为 9 个，近似集算法的时间、空间复杂性为 $T(9)=36$、$S(9)=18$；而信息粒算法的时间、空间复杂性为 $T(9)=27$、$S(9)=9$，信息粒算法的分析参见表 3。可见，信息粒算法具有更优的算法空间复杂性优势，也具有一定的算法时间复杂性优势。

表 3 医疗案例的信息粒算法计算与分析

病人类别	情形	第 1 轮比较大小：$\lvert[x]_m\rvert$	第 2 轮比较大小：$\lvert[x]_m\cap A\rvert$	$\subseteq bnR_{\overline{0.352}}A$	基本计算次数	辅助变量个数
$[x]_1$	(2)	>3.08	$\leqslant 0.35\lvert[x]_1\rvert$	否	3	1
$[x]_2$	(1)	$\leqslant 3.08$	—	否	1	0
$[x]_3$	(1)	$\leqslant 3.08$	—	否	1	0
$[x]_4$	(1)	$\leqslant 3.08$	—	否	1	0
$[x]_5$	(3)	>3.08	$(0.35\lvert[x]_5\rvert,\lvert[x]_5\rvert-2)$	是	5	2
$[x]_6$	(4)	>3.08	$\geqslant\lvert[x]_6\rvert-2$	否	5	2
$[x]_7$	(4)	>3.08	$\geqslant\lvert[x]_7\rvert-2$	否	5	2
$[x]_8$	(4)	>3.08	$\geqslant\lvert[x]_8\rvert-2$	否	5	2
$[x]_9$	(1)	$\leqslant 3.08$	—	否	1	0

5 性质

命题 6　1)$bnR_{\leqslant}\phi=\phi$，$bnR_{\leqslant}U=\phi$；

2)$A\subseteq B$ 时，$bnR_{\leqslant}A\subseteq\overline{R}_\beta A$，$\overline{R}_\beta B$、$bnR_{\leqslant}B\subseteq\sim\underline{R}_k B$，$\sim\underline{R}_k A$；

3)$bnR_{\leqslant}(A\cup B)\subseteq\sim\underline{R}_k A$，$\sim\underline{R}_k B$；

4)$bnR_{\leqslant}(A\cap B)\subseteq\overline{R}_\beta A$，$\overline{R}_\beta B$；

5)$bnR_{\leqslant}(\sim A)=\overline{R}_k A-\underline{R}_\beta A$；

6)$\beta\geqslant\alpha$、$k\geqslant l$ 时，$bnR_{\leqslant}A\subseteq bnR_{\leqslant}A$。

命题 7　$bnR_{\leqslant}(bnR_{\leqslant}A)=\phi$。

证明：记 $bnR_{\leqslant}A=B$，故需证 $bnR_{\leqslant}B=\phi$。

1)$\beta=1$ 时，$bnR_{\beta}A=\phi$，所以 $bnR_{\beta}(bnR_{\beta}A)=bnR_{\beta}A=\phi$；

2)$B=\phi$ 时，$bnR_{\beta}B=\phi=B$；

3)现 $\beta\in[0,1)$、$B\neq\phi$，若 $\exists[x]_R\subseteq bnR_{\beta}B$，则 $|[x]_R\cap B|>\beta|[x]_R|$，$|[x]_R\cap B|<|[x]_R|-k$。

由 $|[x]_R\cap B|>\beta|[x]_R|\geqslant 0$ 有 $[x]_R\subseteq B$，则 $|[x]_R\cap B|=|[x]_R|<|[x]_R|-k$，这是一个矛盾式，所以 $bnR_{\beta}B=\phi$。

推论 1 $bnR(bnRA)=\phi$。

命题 6 研究了边界 $bnR_{\beta}A$ 关于集合系统的一般性质，命题 7 则得到了其幂作用性质。基于双量化扩张性(命题 1)与命题 7 及推论 1 得到了 Pawlak 边界的幂作用结果。

结束语 本文利用笛卡尔积，结合变精度上近似与程度下近似组建了新的双量化扩张粗糙集模型，类似可以做其它平行推广。其中提出的边界，对经典 Pawlak 边界而言，也具有双量化扩张性，因此具有具体的双量化语义，同时对局部不确定性进行了双量化的完备与精细刻画，这对双量化的不确定性分析与应用具有意义。本文研究了该边界的算法，还需要对近似空间或粗糙集理论的双量化扩张与双量化不确定分析进行进一步深入与系统的研究。

参 考 文 献

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

[2] Yao Y Y. The superiority of three-way decision in probabilistic rough set models [J]. Information Sciences, 2011, 181: 1080-1096

[3] Ziarko W. Variable precision rough set model [J]. Journal of Computer and System Sciences, 1993, 46(1): 39-59

[4] Yao Y Y, Wong S K M, Lingras P. A decision-theoretic rough set model [C]// The 5th International Symposium on Methodologies for Intelligent Systems. North-Holland, New York, 1990: 17-25

[5] Yao Y Y, Lin T Y. Generalization of rough sets using modal logics [J]. Intelligent Automation and Soft Computing, 1996, 2(2): 103-120

[6] Inuiguchi M, Yoshioka Y, Kusunoki Y. Variable-precision dominance-based rough set approach and attribute reduction [J]. International Journal of Approximate Reasoning, 2009, 50(8): 1199-1214

[7] Wang J Y, Zhou J. Research of reduct features in the variable precision rough set model [J]. Neurocomputing, 2009, 72: 2643-2648

[8] Mi J S, Wu W Z, Zhang W X. Approaches to knowledge reduction based on variable precision rough set model [J]. Information Sciences, 2004, 159(3/4): 255-272

[9] Yanto I T R, Vitasari P, Herawan T, et al. Applying variable precision rough set model for clustering student suffering study's anxiety [J]. Expert Systems with Applications, 2012, 39(1): 452-459

[10] Yao Y Y, Lin T Y. Graded rough set approximations based on nested neighborhood systems[C]// Proceedings of 5th European Congress on intelligent techniques and Soft computing, EUFIT'97. Verlag Mainz, Aachen, 1997: 196-200

[11] Liu C H, Miao D Q, Zhang N, et al. Graded rough set model based on two universes and its properties [J]. Knowledge-Based Systems, 2012, 33: 65-72

[12] Xu W H, Liu S H, Wang Q R, et al. The first type of graded rough set based on rough membership function[C]// 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery(FSKD). Yantai, China, 2010: 1922-1926

[13] Zhang X Y, Mo Z W, Xiong F, et al. Comparative study of variable precision rough set model and graded rough set model [J]. International Journal of Approximate Reasoning, 2012, 53(1): 104-116

[14] Parthalain N M, Shen Q, Jensen R. A distance measure approach to exploring the rough set boundary region for attribute reduction [J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(3): 305-317

[15] Parthalain N M, Shen Q. Exploring the boundary region of tolerance rough sets for feature selection [J]. Pattern Recognition, 2009, 42(5): 655-667

(上接第 200 页)

步预测。

参 考 文 献

[1] 张树京，齐立心. 时间序列分析简明教程[M]. 北京：北方交通大学大学出版社，2003

[2] Vapnik V N. The nature of statistical learning theory[M]. New York: Springer- Verlag, 1995

[3] 陈磊，张土乔. 基于最小二乘支持向量机的时用水量预测模型[J]. 哈尔滨工业大学学报，2006，38(9)：1528-1529

[4] 王宇红，黄德先，高东杰，等. 基于 LS-SVM 的非线性预测控制技术[J]. 控制与决策，2004，19(4)：384

[5] 王汝言，唐季超，吴大鹏，等. WSN 中基于 GM-LSSVM 的数据融合方法[J]. 计算机工程与设计，2012，33(9)：3372

[6] 陈卫民，陈志刚. 基于 PSR-LSSVM 的网络流量预测[J]. 计算机科学，2012，39(7)：92-95

[7] 陈伟利，范玉刚，吴建德，等. 基于 LSSVM/PID 复合逆系统的预测控制[J]. 计算机科学，2012，39(8)：239-241

[8] 易丹辉. 数据分析与 Eviews 应用[M]. 北京：中国统计出版社，2002：46-92

[9] 黄显峰，邵东国，阳书敏. 降雨时间序列分解预测模型及应用[J]. 中国农村水利水电，2007(9)：6-7

[10] 钱光兴，崔东文. 盘龙河流域水文气象要素变化趋势分析研究[J]. 广东水利水电，2011，4(4)：33

[11] 刘思峰，谢乃明. 灰色系统理论及其应用(第 4 版)[M]. 北京：科学出版社，2008

[12] 耿秋燕，梁毅刚. 基于灰色自适应粒子群 LSSVM 的铁路货运量预测[J]. 西南交通大学学报，2012，47(1)：145-146

[13] 张淑宁，王福利，尤富强，等. 基于鲁邦学习的最小二乘支持向量机及其应用[J]. 控制与决策，2010，25(8)：1170