基于知识粒度的粗糙集的不确定性度量

解 滨[1,2] 李磊军[1] 米据生[1]
（河北师范大学数学与信息科学学院 石家庄 050016）[1] （河北师范大学信息技术学院 石家庄 050016）[2]

摘 要 粗糙集的不确定性与其所在近似空间知识粒度的大小密切相关。提出了近似空间中集合的相对知识粒度的概念。基于相对知识粒度的粗糙集的粗糙性度量既刻画了近似空间对粗糙集不确定性的影响，又去除了负域的干扰。从边界熵的角度提出了一种粗糙集的模糊性度量。随着近似空间知识粒的细分，粗糙集的粗糙度与模糊度均单调递减。

关键词 粗糙集，知识粒度，粗糙度，模糊度，不确定性
中图法分类号 TP18 **文献标识码** A

Uncertainty Measures of Rough Sets Based on Knowledge Granularities

XIE Bin[1,2] LI Lei-jun[1] MI Ju-sheng[1]
(College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang 050016, China)[1]
(College of Information Technology, Hebei Normal University, Shijiazhuang 050016, China)[2]

Abstract Uncertainty of rough sets has close relatives with the knowledge granularities of the approximation space. A concept of relative knowledge granularities was proposed. The roughness of a rough set based on relative knowledge granularities not only reflects the action of the approximation space, but also gets rid of the effect of the negative region of the rough set. A new kind of fuzziness of rough sets based on boundary entropy was designed. Both of the roughness and fuzziness are monotonously decreasing with the refining of knowledge granularities in approximation spaces.

Keywords Rough sets, Knowledge granularities, Roughness, Fuzziness, Uncertainty

1 引言

粗糙集理论是为处理模糊和不确定性问题而提出的一种很有效的数学方法。近年来，关于粗糙集理论中不确定性问题的研究越来越受到关注[1-5]。近似空间知识颗粒的大小直接影响着粗糙集的不确定程度。直观理解的话，知识粒越大信息含量越少，不确定性就越大；知识粒越小信息含量越大，不确定性就越小。Pawlak 提出的精确度和粗糙度[6]可以很直观地反映出一个粗糙集的不确定性，但当包含于一个集合正域中的知识粒被细分时，粗糙集的粗糙度将不再变化。为克服这个问题，有研究者提出了粗糙信息熵及粗糙熵等概念[7,8]，把近似空间本身的不确定性也考虑进来。如梁吉业[8]等定义了一种粗糙熵，将集合 X 的粗糙度与近似空间中的知识粒度做积。这种粗糙熵是随着近似空间知识颗粒的细分而严格递减的，在一定程度上弥补了原粗糙度的不足。但是如果对集合 X 的负域的知识颗粒进行细分，粗糙度将不变，而粗糙熵却严格递减，这说明与集合 X 无关的知识粒的变化也会导致 X 的粗糙熵发生变化，这不符合人们对不确定性问题的直观认识。为此，本文提出集合相对知识粒度的概念，通过相对知识粒度与 Pawlak 粗糙度共同对集合 X 的粗糙性刻画，既兼顾了近似空间的不确定性和集合本身的不确定性，又去除了负域的影响。模糊集[9]自 1965 年由 Zadeh 首次提出以来，被广泛用于不确定问题的研究。很多学者将模糊集理论应用于粗糙集的不确定度量[10-13]。本文从边界熵的角度将粗糙集的模糊度加以推广，构造一种新的模糊性度量。这一度量不仅在知识粒度细化的过程中单调递减，同时具有许多符合人们认知规律的性质。

2 基本概念

设 $K=(U,R)$ 是一个近似空间，其中 $U=\{x_1,x_2,\cdots,x_n\}$ 是一个非空有限集，称为论域；R 是 U 上的一个等价关系或不可辨识关系，$\frac{U}{R}=\{R_1,R_2,\cdots,R_s\}$ 是 U 的一个划分。一个信息系统是四元组 $S=(U,A,V,f)$，其中 $A=\{a_1,a_2,\cdots,a_m\}$ 是属性集，$V=\bigcup_{a\in A}V_a$，V_a 是属性 a 的值域，$f_a:U\rightarrow V_a$ 是信息函数。任意 $B\subseteq A$ 都对应不可辨识关系 $IND(B)=\{(x,y)\in U\times U\mid \forall a\in B(f_a(x)=f_a(y))\}$，其产生的划分简记为 U/B[2]。

给定 U 的一个划分 U/R 和 U 的一个子集 $X\subseteq U$，则 X 的下近似和上近似定义如下。

$\underline{R}X=\bigcup\{R_i\in U/R\mid R_i\subseteq X\}$

$\overline{R}X=\bigcup\{R_i\in U/R\mid R_i\cap X\neq\emptyset\}$

到稿日期：2009-10-20 返修日期：2009-12-30 本文受国家自然科学基金(60773174,60963006)，河北省自然科学基金(F2009000316)资助。
解 滨(1976—)，男，博士生，副教授，主要研究方向为粗糙集、模糊集、近似推理等，E-mail：xiebin_hebtu@126.com；**李磊军**(1985—)，男，硕士生，主要研究方向为近似推理、粗糙集理论等；**米据生**(1966—)，男，教授，博士生导师，主要研究方向为粗糙集、随机集、概念格、近似推理等。

集合 $POS_R(X)=\underline{R}X$ 称为 X 的 R 正域；$NEG_R(X)=U-\overline{R}X$ 称为 X 的 R 负域；$BN_R(X)=\overline{R}X-\underline{R}X$ 称为 X 的 R 边界域。称 X 为 R 可定义集当且仅当 $\underline{R}X=\overline{R}X$，$X$ 为 R 粗糙集当且仅当 $\underline{R}X\neq\overline{R}X$[14]。

集合的不确定性是由于存在边界域而引起的，集合的边界域越大，其精确性越低。Pawlak[6]定义了集合 X 的 R 近似精度为

$$\alpha_R(X)=\frac{|\underline{R}X|}{|\overline{R}X|}$$

X 的 R 粗糙度为

$$\rho_R(X)=1-\alpha_R(X)=1-\frac{|\underline{R}X|}{|\overline{R}X|}=\frac{|BN_R(X)|}{|\overline{R}X|}$$

式中，$X\neq\emptyset$，$|X|$ 表示集合 X 的基数。精度 $\alpha_R(X)$ 反映了我们了解集合 X 的知识的完全程度，而粗糙度 $\rho_R(X)$ 则表达了集合 X 的知识的不完全程度。显然，对于每一个 R 和 $X\subseteq U$，有 $0\leqslant\alpha_R(X)\leqslant 1$，$0\leqslant\rho_R(X)\leqslant 1$。

设 $U/P=\{P_1,P_2,\cdots,P_m\}$ 和 $U/Q=\{Q_1,Q_2,\cdots,Q_l\}$ 是 U 的两个划分，若 $\forall P_i\in P$，$\exists Q_j\in Q$ 使得 $P_i\subseteq Q_j$，则称划分 U/P 比划分 U/Q 细，记为 $U/P\preceq U/Q$。若 $U/P\preceq U/Q$ 且 $U/P\neq U/Q$，则称划分 U/P 比划分 U/Q 严格细，记为 $U/P\prec U/Q$[8]。

例 1　设 $U=\{x_1,x_2,\cdots,x_{10}\}$，$U$ 的一个子集 $X=\{x_1,x_2,x_3,x_4,x_5\}$。给定 U 的如下划分：

$U/R_1=\{\{x_1,x_2,x_3,x_4\},\{x_5,x_6\},\{x_7,x_8,x_9,x_{10}\}\}$，$U/R_2=\{\{x_1\},\{x_2\},\{x_3,x_4\},\{x_5,x_6\},\{x_7,x_8,x_9,x_{10}\}\}$。易见集合 X 的 R_2 正域中的知识粒要比 X 的 R_1 正域中的知识粒严格细，而集合 X 对于 R_1 和 R_2 有相同的精度和粗糙度。这与人们的认知直觉不相符，其原因是这种度量没有反映出近似空间自身的不确定性。

为了克服上述问题，梁吉业等定义了粗糙熵。

定义 1[8]　设 $K=(U,R)$ 为一近似空间，$U/R=\{R_1,R_2,\cdots,R_m\}$ 为 U 的划分，则 R 的知识粒度定义为

$$GK(R)=\frac{1}{|U|^2}\sum_{i=1}^{m}|R_i|^2$$

从定义可以看出，$GK(R)$ 是对知识粒大小的一种平均度量。

定义 2[8]　设 $S=(U,A,V,f)$ 是一个信息系统，$\emptyset\subset X\subseteq U$，$R\subseteq A$，则 X 的 R 粗糙性定义为

$$Roughness_R(X)=\rho_R(X)GK(R)$$

在该度量下，例 1 中的集合 X 关于 R_1 和 R_2 的粗糙性为

$$Roughness_{R_1}(X)=\rho_{R_1}(X)GK(R_1)=(1-\frac{4}{6})\frac{1}{10^2}(4^2+2^2+4^2)=0.12$$

$$Roughness_{R_2}(X)=\rho_{R_2}(X)GK(R_2)=(1-\frac{4}{6})\frac{1}{10^2}(1^2+1^2+2^2+2^2+4^2)=0.09$$

显然，$Roughness_{R_1}(X)>Roughness_{R_2}(X)$，这是由于 U/R_2 比 U/R_1 严格细。

例 2　下面考虑例 1 中 U 的另两种划分。

$U/R_3=\{\{x_1,x_2,x_3,x_4\},\{x_5,x_6\},\{x_7\},\{x_8\},\{x_9,x_{10}\}\}$，$U/R_4=\{\{x_1,x_2,x_3,x_4\},\{x_5,x_6,x_7\},\{x_8\},\{x_9\},\{x_{10}\}\}$。

经计算，$Roughness_{R_3}(X)=0.09$，$Roughness_{R_4}(X)=0.12$，与例 1的结果比较，易见 $Roughness_{R_3}(X)=Roughness_{R_2}(X)$，$Roughness_{R_1}(X)=Roughness_{R_4}(X)$。

集合 X 的 R_2 正域中的知识粒比 X 的 R_3 正域中的知识粒严格细，这一点在上述度量中没有反映出来。X 的 R_1 正域中的知识粒与 X 的 R_4 正域中的知识粒相同而边界域的知识粒不相同，直观上 X 在近似空间 (U,R_4) 下的不确定性要大于在 (U,R_1) 下的不确定性，但是反映出的粗糙性却一样，这也是不符合人们的认知规律的。其原因在于度量中与 X 无关的负域对不确定性也产生了影响。

3　近似空间下集合的相对知识粒度

经分析发现，对于集合 X，粗糙集的不确定度量应该是由集合本身的 R 粗糙度和近似空间 (U,R) 的不确定性两方面决定的。对于近似空间引起的不确定部分主要来自集合 X 的 R 正域和边界域知识粒的大小，而负域中的知识粒的变化与集合 X 是无关的。

为了能更合理地反映粗糙集的不确定性，下面给出集合 X 的相对知识粒度的定义。

定义 3　设 $U=\{x_1,x_2,\cdots,x_n\}$ 是论域，R 是 U 上的不可辨识关系，划分 $U/R=\{R_1,R_2,\cdots,R_m\}$。对 $X\subseteq U$，X 的 R 相对知识粒度定义为

$$GK_R(X)=\frac{1}{|U|^2}\sum_{R_i\subseteq\overline{R}X}|R_i|^2$$

称 $GK_R^{POS}(X)=\frac{1}{|U|^2}\sum_{R_i\subseteq X}|R_i|^2$ 为 X 的 R 正域知识粒度，$GK_R^{BN}(X)=\frac{1}{|U|^2}\sum_{R_i\subseteq BN_RX}|R_i|^2$ 为 X 的 R 边界域知识粒度。易见 $GK_R(X)=GK_R^{POS}(X)+GK_R^{BN}(X)$。这说明 X 的 R 相对知识粒度是由 X 的 R 正域知识粒度和 X 的 R 边界域知识粒度两部分决定的，X 的 R 负域知识颗粒的大小对其不产生影响。

当 X 为精确集时，$GK_R^{BN}(X)=0$，$GK_R(X)=GK_R^{POS}(X)$。

从定义 3 可以看出 $\frac{1}{|U|}\leqslant GK_R(X)\leqslant 1$，这是由于 R_i 满足 $1\leqslant|R_i|\leqslant|U|$。对于给定的近似空间，$GK_R(X)$ 也不是定值，而是随 X 的变化而变化，因此具有相对性。

下面定义一种粗糙集的不确定性度量为

$$E_R(X)=\rho_R(X)GK_R(X)\tag{1}$$

该度量由两部分组成：一是集合 X 的 R 粗糙度；二是近似空间 (U,R) 对集合产生的不确定性，该不确定性是集合 X 的 R 正域知识粒度与 R 边界域知识粒度共同作用的结果。

例 1 与例 2 中集合 X 在各个近似空间下的不确定性由式(1)计算得出，$E_{R_1}(X)=0.07$，$E_{R_2}(X)=0.03$，$E_{R_3}(X)=0.07$，$E_{R_4}(X)=0.11$。不难发现度量 $E_R(X)$ 的计算结果去除了负域的影响，因此更符合人们的认知规律。

定理 1　设 P,Q 为 U 上的两个不可辨识关系，对于 $X\subseteq U$，若 $U/P\preceq U/Q$，则 $E_P(X)\leqslant E_Q(X)$。

证明：设 $U/P=\{P_1,P_2,\cdots,P_p\}$，$U/Q=\{Q_1,Q_2,\cdots,Q_q\}$。因为 $U/P\preceq U/Q$，则有 $\overline{P}X\subseteq\overline{Q}X$，$\underline{P}X\supseteq\underline{Q}X$ 及 $p\geqslant q$，并且存在指标集 $\{1,2,\cdots,p\}$ 的一个划分 $\{C_1,C_2,\cdots,C_q\}$ 满足 $Q_j=\bigcup_{i\in C_j}P_i$，$j=1,2,\cdots,q$，所以

$$GK_Q(X)=\frac{1}{|U|^2}\sum_{j(Q_j\subseteq\overline{Q}X)}|Q_j|^2=\frac{1}{|U|^2}\sum_{j(Q_j\subseteq\overline{Q}X)}|\bigcup_{i\in C_j}P_i|^2$$

$$=\frac{1}{|U|^2}\sum_{j(Q_j\subseteq \bar{Q}X)}(\sum_{i\in C_j}|P_i|)^2\geqslant\frac{1}{|U|^2}\sum_{i(P_i\subseteq \bar{Q}X)}|P_i|^2$$

又因为 $GK_{\bar{P}}(X)=\frac{1}{|U|^2}\sum_{i(P_i\subseteq \bar{P}X)}|P_i|^2$ 及 $\bar{P}X\subseteq\bar{Q}X$，有 $GK_{\bar{P}}(X)\leqslant\frac{1}{|U|^2}\sum_{i(P_i\subseteq \bar{Q}X)}|P_i|^2$，所以 $GK_{\bar{P}}(X)\leqslant GK_{\bar{Q}}(X)$。

由 $\bar{P}X\subseteq\bar{Q}X$，$\underline{P}X\supseteq\underline{Q}X$，易知 $|\bar{P}X|\leqslant|\bar{Q}X|$，$|\underline{P}X|\geqslant|\underline{Q}X|$，故 $\alpha_P(X)=\frac{|\underline{P}X|}{|\bar{P}X|}\geqslant\frac{|\underline{Q}X|}{|\bar{Q}X|}=\alpha_Q(X)$，则 $\rho_P(X)=1-\alpha_P(X)\leqslant1-\alpha_Q(X)=\rho_Q(X)$。

综上，$\rho_P(X)GK_{\bar{P}}(X)\leqslant\rho_Q(X)GK_{\bar{Q}}(X)$，因此，$E_P(X)\leqslant E_Q(X)$。证毕。

对于任意给定一个信息系统 $S=(U,A,V,f)$，属性集 A 的幂集 $P(A)$ 与包含关系 $\subseteq$ 构成一个格 $\langle P(A),\subseteq\rangle$。在其对应的 Hasse 图中，从 $\emptyset$ 到 A 的任一路径均构成一条属性链。同一属性链上的任意两个属性子集 $B_i\subset B_j\subseteq A$ 均满足 $U/B_j\leq U/B_i$[2]。由定理 1 易得如下推论。

推论 1 设信息系统 $S=(U,A,V,f)$，对于 $X\subseteq U$，任意一条属性链 $\emptyset=B_0\subset B_1\subset\cdots\subset B_k=A$，均有 $E_{B_k}(X)\leqslant E_{B_{k-1}}(X)\leqslant\cdots\leqslant E_{B_1}(X)\leqslant E_{B_0}(X)$。

4 基于知识粒度的粗糙集的模糊度

对粗糙集的模糊性研究主要通过两个方面入手：一是在一个近似空间下将一个粗糙集转化为一个模糊集；二是利用模糊集的不确定性测度研究粗糙集的模糊性。

定义 4[14] 设 (U,R) 是 Pawlak 近似空间，对象子集 $X\subseteq U$，则对任意的 $x\in U$，x 属于集合 X 的粗糙隶属度为

$$R(X)(x)=\frac{|X\cap[x]_R|}{|[x]_R|}\tag{2}$$

由式(2)不难发现，对于经典集 X，粗糙隶属函数 $R(X)$ 与 X 在近似空间 (U,R) 下的粗糙集是密切相关的。当 $x\in POS_R(X)$ 时，$R(X)(x)=1$；当 $x\in NEG_R(X)$ 时，$R(X)(x)=0$；当 $x\in BN_R(X)$ 时，$0<R(X)(x)<1$。

定义 5[12] 设 U 是有限论域，记 $F((U)$ 为 U 上的模糊集合全体，$\forall A\in F(U)$，A 的模糊度定义为

$$E(A)=\frac{4}{|U|}\sum_{x\in U}A(x)(1-A(x))\tag{3}$$

定义 4 中的 $R(X)$ 为 U 上的一个模糊集合，记 $ER(X)$ 为 $R(X)$ 的模糊度。由式(3)计算其模糊度时，只有当 $x\in BN_R(X)$ 时 $R(X)(x)(1-R(X)(x))\neq0$，即 X 的 R 正域和负域中元素的隶属度对于 $R(X)$ 的模糊度不产生影响。也就是说，$R(X)$ 的模糊性完全来自于边界，但是式(3)定义的模糊度量并没有完全反映出边界知识粒的大小。为了解决这一问题，我们定义边界熵

$$E_R^{BN}(X)=\sum_{R_i\subseteq BN_R(X)}\frac{|R_i|}{|U|}(1-\frac{|R_i|}{|U|})$$

用

$$F_R(X)=ER(X)(1-E_R^{BN}(X))\tag{4}$$

表示集合 X 在近似空间 (U,R) 中粗糙集的模糊性度量。这一度量由集合本身的模糊性和边界知识粒度引起的模糊性两部分组成，其结果可更好地反映出不同知识粒度对集合的模糊性的影响。

文献[14]用拓扑特征刻画了集合 X 在近似空间 (U,R) 中粗糙集的不确定性：

(1)如果 $\underline{R}X\neq\emptyset$，$\bar{R}X\neq U$，则称 X 是粗糙可定义的；

(2)如果 $\underline{R}X=\emptyset$，$\bar{R}X\neq U$，则称 X 是内不可定义的；

(3)如果 $\underline{R}X\neq\emptyset$，$\bar{R}X=U$，则称 X 是外不可定义的；

(4)如果 $\underline{R}X=\emptyset$，$\bar{R}X=U$，则称 X 是全不可定义的。

命题 1 由式(4)定义的模糊度满足如下性质：

(1) $F_R(X)=0$ 当且仅当 X 为经典可定义集合；

(2)对任意 $X\subseteq U$，有 $F_R(X)=F_R(\sim X)$；

(3)在信息系统 $S=(U,A,V,f)$ 中，$B_1\subseteq A$，$B_2\subseteq A$，$X\subseteq U$，如果 $U/B_2\prec U/B_1$，则 $F_{B_2}(X)\leqslant F_{B_1}(X)$；

(4)在信息系统 $S=(U,A,V,f)$ 中，$R\subseteq A$，$X\subseteq U$，若 $F_R(X)=1$，则 X 关于 R 是全不可定义的，并且 $U/R=\{U\}$；

(5)在信息系统 $S=(U,A,V,f)$ 中，$B_1\subseteq B_2\subseteq A$，$X\subseteq U$，若 $F_{B_1}(X)=F_{B_2}(X)$，则 $\rho_{B_1}(X)=\rho_{B_2}(X)$。

证明：性质(1)、(2)由定义易证。

(3)设 $U/B_1=\{P_1,P_2,\cdots,P_l\}$，$U/B_2=\{Q_1,Q_2,\cdots,Q_m\}$，显然有 $l<m$。因 $U/B_2\prec U/B_1$，故 U/B_2 对 $U/B_1=\{P_1,P_2,\cdots,P_l\}$ 中至少一个元素进行细分。为简化证明，我们不妨设 U/B_1 中只有 P_1 被细分成两个部分(与细分为多个部分的证明类似)，$P_1=Q_i\cup Q_j$ $(Q_i,Q_j\in U/B_2)$，U/B_1 的其它元素不变(其它情况可以依据此种情况类似证明)。下面分情况讨论：

1)当 $P_1\cap X=\emptyset$ 时，对于任意的 $x(x\in P_1)$，$B_1(X)(x)=\frac{|P_1\cap X|}{|P_1|}=0$。因为 $P_1=Q_i\cup Q_j$ $(Q_i\cap Q_j=\emptyset)$，所以对于任意的 $x\in Q_i\cup Q_j$，$B_2(X)(x)=\frac{|Q_i\cap X|}{|Q_i|}=\frac{|Q_j\cap X|}{|Q_j|}=0$。因此，$EB_1(X)=EB_2(X)$。此时，$P_1\nsubseteq BN_{B_1}(X)$，$Q_i\cup Q_j\nsubseteq BN_{B_2}(X)$，故 $E_{B_1}^{BN}(X)=E_{B_2}^{BN}(X)$。

2)当 $P_1\subseteq X$ 时，对于任意的 $x\in P_1$，$B_1(X)(x)=\frac{|P_1\cap X|}{|P_1|}=1$。对于任意的 $x\in Q_i\cup Q_j$，$B_2(X)(x)=\frac{|Q_i\cap X|}{|Q_i|}=\frac{|Q_j\cap X|}{|Q_j|}=1$。因此，$EB_1(X)=EB_2(X)$。此时，$P_1\nsubseteq BN_{B_1}(X)$，$Q_i\cup Q_j\nsubseteq BN_{B_2}(X)$，故 $E_{B_1}^{BN}(X)=E_{B_2}^{BN}(X)$。

3)当 $P_1\cap X\neq\emptyset$，且 $P_1\cap X\neq P_1$ 时，可分以下 3 种情况讨论：

①如果 $Q_i\cap X=\emptyset$，设 $|P_1\cap X|=a$，$|P_1|-|P_1\cap X|=b$，则 $\frac{4}{|U|}\sum_{x\in P_1}B_1(X)(x)(1-B_1(X)(x))=\frac{4}{|U|}a(1-\frac{a}{a+b})$。这里 $|Q_j\cap X|=a$，$|Q_j|-|Q_j\cap X|=t<b$，则 $\frac{4}{|U|}\sum_{x\in Q_i\cup Q_j}B_2(X)(x)(1-B_2(X)(x))=\frac{4}{|U|}a(1-\frac{a}{a+t})$。而 $\frac{4}{|U|}a(1-\frac{a}{a+b})-\frac{4}{|U|}a(1-\frac{a}{a+t})=\frac{4}{|U|}\frac{a^2(b-t)}{(a+b)(a+t)}>0$。

②如果 $Q_i\subseteq X$，则 $|Q_j\cap X|=s<a$，$|Q_j|-|Q_j\cap X|=b$，则 $\frac{4}{|U|}\sum_{x\in Q_i\cup Q_j}B_2(X)(x)(1-B_2(X)(x))=\frac{4}{|U|}s(1-\frac{s}{s+b})$。而 $\frac{4}{|U|}a(1-\frac{a}{a+b})-\frac{4}{|U|}s(1-\frac{s}{s+b})=\frac{4}{|U|}\frac{(a-s)b^2}{(a+b)(s+b)}>0$。

③如果 $Q_i\cap X\neq\emptyset$，$Q_i\cap X\neq Q_i$，$Q_j\cap X\neq\emptyset$，$Q_j\cap X\neq Q_j$。令

$|Q_i \cap X|=h$，$|Q_j \cap X|=k$，$|Q_i|-|Q_i \cap X|=m$，$|Q_j|-|Q_j \cap X|=n$。此时，$h+k=a$，$m+n=b$。

令 $h+m=h_1$，$k+n=k_1$，则 $a+b=h_1+k_1$，而 $\frac{4}{|U|}\sum_{x\in P_1}B_1(X)(x)(1-B_1(X)(x))=\frac{4}{|U|}(h+k)(1-\frac{h+k}{h_1+k_1})$，$\frac{4}{|U|}\sum_{x\in Q_i\cup Q_j}B_2(X)(x)(1-B_2(X)(x))=\frac{4}{|U|}h(1-\frac{h}{h_1})+\frac{4}{|U|}k(1-\frac{k}{k_1})$。因此，$\frac{4}{|U|}\sum_{x\in P_1}B_1(X)(x)(1-B_1(X)(x))-\frac{4}{|U|}\sum_{x\in Q_i\cup Q_j}B_2(X)(x)(1-B_2(X)(x))=\frac{4}{|U|}\frac{(hk_1-kh_1)^2}{h_1k_1(h_1+k_1)}\geqslant 0$，当且仅当 $\frac{h}{h_1}=\frac{k}{k_1}$ 时等号成立。

综合①，②，③，有 $EB_2(X)\leqslant EB_1(X)$。此时，

$$\frac{|Q_i|}{|U|}\left(1-\frac{|Q_i|}{|U|}\right)+\frac{|Q_j|}{|U|}\left(1-\frac{|Q_j|}{|U|}\right)-\frac{|P_1|}{|U|}\left(1-\frac{|P_1|}{|U|}\right)$$
$$=\frac{|Q_i|}{|U|}\left(1-\frac{|Q_i|}{|U|}\right)+\frac{|Q_j|}{|U|}\left(1-\frac{|Q_j|}{|U|}\right)-\frac{|Q_i|+|Q_j|}{|U|}\left(1-\frac{|Q_i|+|Q_j|}{|U|}\right)$$
$$=\frac{2|Q_i||Q_j|}{|U|}>0$$

故 $E_{B_1}^{BN}(X)<E_{B_2}^{BN}(X)$。因此，$1-E_{B_2}^{BN}(X)<1-E_{B_1}^{BN}(X)$。

综合 1)，2)，3)，有 $EB_2(X)(1-E_{B_2}^{BN}(X))\leqslant EB_1(X)(1-E_{B_1}^{BN}(X))$，即 $F_{B_2}(X)\leqslant F_{B_1}(X)$。而且，当边界域被严格细分时“<”成立。

(4)由式(3)和式(4)易见 $0\leqslant ER(X)\leqslant 1$，$0\leqslant E_R^{BN}(X)<1$，则$F_R(X)=ER(X)(1-E_R^{BN}(X))=1$ 当且仅当 $ER(X)=1$，且 $E_R^{BN}(X)=0$。由 $ER(X)=1$，知 $\forall x\in U$，$R(X)(x)=0.5$，故$\underline{R}X=\emptyset$，$\bar{R}X=U$，所以 X 关于 R 是全不可定义的。而 $E_R^{BN}(X)=0$ 当且仅当存在唯一的 $R_i\in U/R$ 满足 $R_i=BN_R(X)$，又由$\underline{R}X=\emptyset$，$\bar{R}X=U$ 知 $BN_R(X)=U$，因此 $U/R=\{U\}$。

(5)用反证法。因 $B_1\subseteq B_2\subseteq A$，故 $U/B_2\preceq U/B_1$，由定理 1 的证明知 $\rho_{B_2}(X)\leqslant\rho_{B_1}(X)$，假设 $\rho_{B_1}(X)\neq\rho_{B_2}(X)$，则 $\rho_{B_2}(X)<\rho_{B_1}(X)$，说明边界域 $BN_{B_2}(X))$比 $BN_{B_1}(X))$要“严格小”，即存在 $P_i(P_i\in U/B_1\wedge P_i\subseteq BN_{B_1}(X))$在 U/B_2 中被“严格细分”，因此 $U/B_2\prec U/B_1$。由本命题性质(3)得 $F_{B_2}(X)<F_{B_1}(X)$，这与已知 $F_{B_2}(X)=F_{B_1}(X)$矛盾。因此，$\rho_{B_2}(X)=\rho_{B_1}(X)$。证毕。

结束语　粗糙集是用于研究不确定问题的有力工具，所以其本身就带有不确定性。粗糙度和模糊度从不同角度给出了粗糙集不确定性的度量，前者比较直观，而后者较为抽象。粗糙集不确定性来自集合本身及近似空间两个方面，将这两个方面合理地结合起来，才能更好地反映粗糙集的不确定性。本文定义了近似空间中集合的相对知识粒度，将其与 Pawlak 粗糙度结合，给出了一种新的粗糙集粗糙性的度量。该度量既刻画了近似空间对粗糙集不确定性的影响，又去除了负域的干扰，并且满足随着近似空间知识粒的细分粗糙集的不确定性单调递减。粗糙集的模糊度与近似空间知识粒大小也是密切相关的，边界域对于模糊度有决定性的影响。本文结合重新定义的边界熵，给出了新的粗糙集的模糊度量，它可以更合理地反映不同知识粒度对集合的模糊性的影响，同时具备符合人们认知规律的一些性质。

参 考 文 献

[1] 李德毅，刘常昱，杜鹢，等. 不确定性人工智能[J]. 软件学报，2004，15(11)：1583-1594

[2] 王国胤. Rough 集理论与知识获取[M]. 西安：西安交通大学出版社，2001

[3] 苗夺谦，王玨. 粗糙集理论中知识粗糙性与信息熵关系的讨论[J]. 模式识别与人工智能，1998，11(3)：34-40

[4] 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

[5] 梁吉业，李德玉. 信息系统中的不确定性与知识获取[M]. 北京：科学出版社，2005

[6] Pawlak Z. Rough sets[J]. International Journal of Computer and Information Science，1982，11：341-356

[7] 李玉榕，乔斌，蒋静坪. 粗糙集理论中不确定性粗糙信息熵表示[J]. 计算机科学，2002，29：101-103

[8] Liang Ji-ye，Shi Zhong-zhi. The information entropy，rough entropy and knowledge granulation in rough set theory[J]. International Journal of Uncertainty，2004，12(1)：37-46

[9] Zadeh L A. Fuzzy sets[J]. Information Control，1965，8：338-353

[10] Chakrabarty K，Biswas R，Nanda S. Fuzziness in rough sets[J]. Fuzzy Sets and Systems，2000，110：247-251

[11] Liang Ji-ye，Chin K S，Dang C Y. A new method for measuring uncertainty and fuzziness in rough set theory[J]. International Journal of General Systems，2002，31(4)：331-342

[12] Mi Ju-sheng，Yee Leung，Wu Wei-Zhi. An uncertainty measure in partition-based fuzzy rough sets[J]. International Journal of General Systems，2005，34：77-90

[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

[14] 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

(上接第 224 页)

[6] 孙玉娣，裴勇. 基于可视化文本挖掘的本体构建[J]. 情报杂志，2007，26(12)

[7] Roy R，Mili H，Blettner M. Development and application of a metric on semantic nets[J]. IEEE Transaction on System，1989，19(1)

[8] Song Shaoxu，Li Chunping. TCUA P：a novel app roach of text clustering using asymmetricproximity[C]//Proceedings of the 2nd Indian International Conference on Artificial Intelligence. India：IICA I，2005：604-613

[9] 孙爽，章勇. 一种基于语义相似度的文本聚类算法[J]. 南京航空航天大学学报，2006(6)

[10] 范明，孟小峰. 数据挖掘概念与技术[M]. 北京：机械工业出版社，2002

[11] 王元明，熊伟. 异常数据的检测方法[J]. 重庆工学院学报：自然科学版，2009，23(2)：86-89