第 37 卷 第 1 期　江西师范大学学报(自然科学版)　Vol. 37 No. 1
2013 年 1 月　Journal of Jiangxi Normal University(Natural Science)　Jan. 2013

文章编号: 1000-5862(2013)01-0023-05

基于秩统计量的粗糙集精度的度量方法

吴根秀,刘佩红,罗冰辉,谢　君
(江西师范大学数学与信息科学学院,江西 南昌 330022)

摘要: 提出一种基于秩统计量的粗糙集精度的度量方法,该方法既考虑了知识颗粒块数据值的大小的信息,又考虑了论域大小,并给出这一度量的若干性质,实例表明所给出的精度度量是合理的、有效的.

关键词: 粗糙集; 秩统计量; 精度; 改进知识含量测度
中图分类号: TP 11　　**文献标志码**: A

0　引言

Z. Pawlak 等[1] 在 20 世纪 80 年代初提出的粗糙集理论是处理不完全和不精确信息的一种新的数学工具[2],粗糙集的不确定性主要由系统的不确定性和概念的不确定性 2 个原因引起的. 经典的粗糙集的近似精度,没有考虑到由等价关系导出的划分颗粒的大小,存在一定的局限性[3]. 文献[4]利用过剩熵给出了粗糙集的不确定性度量. 文献[5]给出一种基于知识含量的粗糙集不确定性度量. 文献[6]提出了论域 U 的 2 个等价类之间的距离,然后从距离角度定义划分的散度来描述划分的颗粒块的大小,再给出了近似精度的定义.

本文在分析传统精度的局限性基础上,在综合文献[4-6]工作的基础上,充分考虑到知识颗粒块数据值的大小的信息对精度的影响,提出了一种基于秩统计量的粗糙集精度的度量方法,并给出这一度量的优良性质,最后通过 2 个实例,说明了所给出的精度度量的合理性、有效性.

1　粗糙集理论基础知识

设 U 是一个非空有限集合,称为论域[7-8]. 称任何子集 $X \subseteq U$ 为 U 中的 1 个概念或范畴. 设 R 是 U 上的 1 个等价关系,U/R 是 R 的所有等价类构成的集合,符号 $[x]_R$ 表示包含 $x \in U$ 的 R 等价类. 设 A 是 U 上的一族等价关系,称二元组 $K = (U,A)$ 为一知识库(或称为 Pawlak 近似空间).

定义 1(不可分辨关系)[9]　设 $K = (U,A)$ 为一知识库,$P \subseteq R$,$P \neq \varnothing$,则 $\cap P$ 也是一种等价关系,用 $IND(P) = \{(x,y) \in U \times U \mid \forall a \in P, f(x,a) = f(y,a)\}$ 表示属性集 P 上的不可分辨关系,$IND(P) = \cap P$ 是 U 上的等价关系,$[x]_{IND(P)} = \bigcap_{R_i \in P} [x]_{R_i}$,$x \in U$ 表示不可分辨关系 $IND(P)$ 的所有等价类. 为了表述简便,将 $U/IND(P)$ 记为 U/P.

定义 2(上、下近似及边界)[10-11]　$\forall X \subseteq U$,称 $\underline{R}(X) = \{x \in U \mid [x]_R \subseteq X\}$ 为 X 关于近似空间 (U,A) 的下近似集,称 $\overline{R}(X) = \{x \in U \mid [x]_R \cap X \neq \varnothing\}$ 为 X 关于近似空间 (U,A) 的上近似集,称 $Pos_R(X) = \underline{R}(X)$ 为 X 的 R 正域,称 $Neg_R(X) = U - \overline{R}(X)$ 为 X 的 R 负域,称 $Bn_R(X) = \overline{R}(X) - \underline{R}(X)$ 为 X 的 R 边界域. 显然有 $\overline{R}(X) = \underline{R}(X) \cup Bn_R(X)$.

当 $\overline{R}(X) \neq \underline{R}(X)$ 时,称二元组 $(\underline{R}(X), \overline{R}(X))$ 为近似空间中的粗糙集. $Bn_R(X)$ 中的元素表示根据知识 R 不能分辨是属于 X 还是属于 $\sim X$(即 $U - X$)的 U 中元素; $Neg_R(X)$ 中的元素表示根据知识 R 分辨一定不属于 X 的 U 中元素; $\underline{R}(X)$ 或 $Pos_R(X)$ 中的元素表示根据知识 R 分辨一定属于 X 的 U 中元素; $\overline{R}(X)$ 中的元素表示根据知识 R 分辨可能属于 X 的 U 中元素.

在粗糙集理论中,集合的不精确性是由于边界区域的存在而引起,集合的边界区域越大,其精确性越低. 为了更准确地表达这一点,引入精度的概念.

定义 3(近似精度)　关于等价关系 R 的集合 X 的近似精度为

收稿日期: 2012-09-15
基金项目: 江西省自然科学基金(20114BAB201)和江西师范大学研究生创新基金(YJS2012081)资助项目.
作者简介: 吴根秀(1965-),女,江西南丰人,教授,主要从事不确定推理与信息融合的研究.

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

其中 $X \neq \varnothing$,$|X|$是集合 X 的基数.

精度 $\alpha_R(X)$ 描述了根据已有知识对概念 X 的了解程度. 显然,对于每个 R 和 $X \subseteq U$ 有 $0 \leqslant \alpha_R(X) \leqslant 1$. 当 $\alpha_R(X) = 1$ 时,集合 X 的 R 边界域为 $\varnothing$,集合 X 是 R 可定义的; 当 $\alpha_R(X) < 1$ 时,集合 X 有非空 R 边界域,集合 X 为 R 不可定义的. 集合 X 的 R 粗糙度 $\rho_R(X)$ 定义为 $\rho_R(X) = 1 - \alpha_R(X)$,它与精度相反,表示对概念 X 的不了解程度.

2 一种基于秩统计量的粗糙集精度的度量方法

2.1 Pawlak 精度的局限性

定义 4 设 $P = \{P_1, P_2, \cdots, P_n\}$,$Q = \{Q_1, Q_2, \cdots, Q_n\}$ 是论域 U 的 2 个划分,如果 $\forall P_i \in P(i = 1, \cdots, n)$,$\exists Q_j(j = 1, \cdots, m)$ 使得 $P_i \subseteq Q_j$,则称划分 P 比划分 Q 更细,记为 $P \leq Q$. 如果 $P \leq Q$ 且 $P \neq Q$,则称划分 P 严格细于划分 Q,记为 $P < Q$.

事实上,关于等价关系 R 的划分颗粒表现了论域 U 中的对象的知识. 划分越细反映对某个特定对象的信息就越多,然而,传统粗糙集理论的精度并不能反映完全包括在下近似集内的颗粒的大小,即在 Pawlak 粗糙集中,当 $R_1 \subseteq R_2$ 时,$\forall X \subseteq U$ 有 $\alpha_{R_1}(X) \leqslant \alpha_{R_2}(X)$,但当 $R_1 \subset R_2$ 时,却没有 $\alpha_{R_1}(X) < \alpha_{R_2}(X)$.

例 1 设论域 $U = \{1,2,3,4,5,6,7,8,9,10\}$,集合 $X = \{1,2,3,4,6,7\}$,R_1, R_2, R_3 为 U 上的 3 个等价关系,分别导出的划分为

$U/R_1 = \{\{1,2,3,4\},\{5,6,7\},\{8,9,10\}\}$,

$U/R_2 = \{\{1,2\},\{3,4\},\{5,6,7\},\{8,9\},\{10\}\}$,

$U/R_3 = \{\{1\},\{2\},\{3\},\{4\},\{5,6,7\},\{8\},\{9\},\{10\}\}$,

则 X 的 R_1, R_2, R_3 的下近似集和上近似集分别为

$\underline{R_1}X = \underline{R_2}X = \underline{R_3}X = \{1,2,3,4\}$,

$\overline{R_1}X = \overline{R_2}X = \overline{R_3}X = \{1,2,3,4,5,6,7\}$.

由 Pawlak 近似精度的定义可得 $\alpha_{R_1}(X) = \alpha_{R_2}(X) = \alpha_{R_3}(X) = 4/7$,从而认为 X 关于 R_1, R_2, R_3 的近似精度相同. 但显然,$R_3 < R_2 < R_1$,即知识 R_3 比 R_2 精细,知识 R_2 比 R_1 精细. 从上述的分析可知 R_3 获得的信息应更多,但它们的近似精度相等. 所以,Pawlak 近似精度 $\alpha_R(X)$ 具有局限性,只能部分反映 X 的不确定性,不能完全反映 X 的不确定性.

2.2 一种秩统计量的知识含量测度度量方法

当划分最细($|X_i| = 1, i = 1, \cdots, n$)时,文献[5]给出的精度突然变大,未考虑论域的大小. 此种度量也有一定的局限性. 而粗糙集不确定性的 2 个原因之一是论域上二元等价关系对论域中对象进行分类的能力,即知识的粒度. 由此,讨论知识粒度并对其进行改进.

为便于讨论,对非空论域 U,近似空间 $K = (U, A)$,$U/R = \{X_1, X_2, \cdots, X_n\}$,$1 \leqslant n \leqslant |U|$,$\sum_{i=1}^{n}|X_i| = |U|$,假设 $|X_1|, |X_2|, \cdots, |X_n|$ 由小到大排列成 $z_{(1)} < z_{(2)} < \cdots < z_{(n)}$. 若 $|X_i| = z_{(r_i)}$,则称 $|X_i|$ 的秩为 r_i,$r_i = 1, 2, \cdots, n$. 并称 $\theta = \{\{x\} \mid x \in U\}$ 为 U 的"离散划分",称 $\Omega = \{U\}$ 为 U 的"全体划分".

定义 5 称 $S(R) = \frac{1}{n}\sum_{i=1}^{n}(r_i |X_i|^2 / \sum_{j=1}^{n} r_j)$ 为基于秩统计量的知识 R 的颗粒块测度,其中 r_i 为 $|X_i|$ 的秩,$1 \leqslant n \leqslant |U|$,$\sum_{i=1}^{n}|X_i| = |U|$,$|X_i|$ 表示集合 X_i 的基数.

先给出重要的引理 1,再由定义 5 可给出 $S(R)$ 的性质.

引理 1 对于论域 U 的任意 2 个划分 P, Q,设 $U/P = \{X_1, X_2, \cdots, X_n\}$,$|X_1| \leqslant |X_2| \leqslant \cdots \leqslant |X_n|$,$U/Q = \{X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_{j-1}, X_{j+1}, \cdots, X_{j+k-1}, X_i \cup X_j, X_{j+k}, \cdots, X_n\}$,其中 $1 \leqslant k \leqslant n - j + 1$,则有 $S(P) < S(Q)$.

证 用 x_i 表示知识颗粒块 X_i 中的元素个数,$i = 1, 2, \cdots, n$,且 $|X_i|$ 的秩为 r_i,$r_i = 1, 2, \cdots, n$. 由定义 5 知,$S(P) = \frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{\sum_{i=1}^{n} r_i}|X_i|^2$,$S(Q) =$

$$\frac{2}{(n-1)^2 n} \cdot \left[\sum_{l=1}^{i-1} l x_l^2 + \sum_{l=i}^{j-2} l x_{l+1}^2 + \sum_{l=j-1}^{j+k-3} l x_{l+2}^2 + (j+k-2)(x_i^2 + x_j^2 + 2x_i x_j) + \sum_{l=j+k-1}^{n-1} l x_{l+1}^2\right],$$

$$S(Q) - S(P) = \frac{2}{n^2(n-1)^2(n+1)} \cdot \left[\sum_{l=1}^{i-1}(3n-1) l x_l^2 + \sum_{l=i+1}^{n}(3n-1) l x_l^2 - n(n+1)\sum_{l=i+1}^{n} x_l^2 - \sum_{l=j+1}^{j+k-1} n(n+1) x_l^2 + (n(n+1)(j+k-2) - (n -\right.$$

$1)^2 i)x_i^2 + n(n+1)(k-1)x_j^2 + (j+k-2)n(n+1)2x_ix_j\Big] \geqslant \frac{2}{n^2(n-1)^2(n+1)}\Big[\sum_{l=1}^{i-1}(3n-1)lx_l^2 + \sum_{l=i+1}^{n}(3n-1)lx_l^2 - n(n+1)\sum_{l=i+1}^{n}x_l^2 - \sum_{l=j+1}^{j+k-1}n(n+1)x_{j+k-1}^2 + (n(n+1)(j+k-2)-(n-1)^2 i)x_i^2 + n(n+1)(k-1)x_j^2 + (j+k-2)n(n+1)2x_ix_j\Big] = \frac{2}{n^2(n-1)^2(n+1)}\Big[\sum_{l=1}^{i-1}(3n-1)lx_l^2 + \sum_{l=i+1}^{n}(3n-1)lx_l^2 - n(n+1)\sum_{l=i+1}^{n}x_l^2 + M + N\Big]$,

由于 $x_i + x_j \geqslant x_{j+k-1}$,所以 $M = n(n+1)(k-1)\cdot[(x_i+x_j)^2 - x_{j+k-1}^2] + (j-1)n(n+1)2x_ix_j \geqslant 0$.

又由于 $j > i$,$j-1 \geqslant i \geqslant 1$,所以 $N = [(j-1)n(n+1) - (n-1)^2 i]x_i^2 \geqslant (3n-1)(j-1) > 0$,

从而 $S(Q) - S(P) > \frac{2}{n^2(n-1)^2(n+1)}\Big[\sum_{l=1}^{i-1}(3n-1)lx_l^2 + \sum_{l=i+1}^{n}(3n-1)lx_l^2 - n(n+1)\sum_{l=i+1}^{n}x_l^2\Big]$.

令 $T_n = \sum_{l=i+1}^{n}(3n-1)lx_l^2 - n(n+1)\sum_{l=i+1}^{n}x_l^2$,下面用数学归纳法证明: 若 $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n$,则 $\forall n$,$1 \leqslant i < n$,有 $T_n > 0$.

事实上(ⅰ)当 $n = 2$ 且 $i = 1$ 时,有 $T_2 = 4x_2^2 > 0$,所以当 $n = 2$ 时成立.

(ⅱ)当 $n = 3$ 时,若 $i = 1$ 则 $T_3 = 4x_2^2 + 12x_3^2 > 0$; 若 $i = 2$ 则有 $T_3 = 12x_3^2 > 0$,所以当 $n = 3$ 时成立.

(ⅲ)假设 $T_n > 0$,下面往证 $T_{n+1} > 0$.

当 $1 \leqslant i < n$ 时,有 $T_{n+1} = \sum_{l=i+1}^{n+1}(3n-1+3)\cdot lx_l^2 - (n+1)(n+2)\sum_{l=i+1}^{n+1}x_l^2 = T_n + 2n(n+1)x_{n+1}^2 + 3\sum_{l=i+1}^{n}lx_l^2 - 2(n+1)\sum_{l=i+1}^{n}x_l^2 > 2n(n+1)x_{n+1}^2 + 3\sum_{l=i+1}^{n}lx_l^2 - 2(n+1)\sum_{l=i+1}^{n}x_l^2 \geqslant 2n(n+1)x_{n+1}^2 + 3\sum_{l=i+1}^{n}lx_l^2 - 2(n+1)(n-i-1+1)x_{n+1}^2 = 2(n+1)ix_{n+1}^2 + 3\sum_{l=i+1}^{n}lx_l^2 > 0$.

当 $i = n$ 时,有 $T_{n+1} = (3n+3-1)(n+1)x_{n+1}^2 - (n+1)(n+2)x_{n+1}^2 = 2n(n+1)x_{n+1}^2 > 0$.

故 $T_{n+1} > 0$.

因此,$S(Q) - S(P) > \frac{2}{n^2(n-1)^2(n+1)}\cdot\Big[\sum_{l=1}^{i-1}(3n-1)lx_l^2 + T_n\Big] > 0$,即 $S(P) < S(Q)$.

性质 1　设 R 是 U 上的等价关系,则 $1/|U| \leqslant S(R) \leqslant |U|^2$,且当 $U/R = \theta$ 时,$S(R)$ 取最小值为 $1/|U|$; 当 $U/R = \Omega$ 时,$S(R)$ 取最大值为 $|U|^2$.

证　由定义 5 及引理 1 知,对任意的等价关系 R,若 $\theta < R < \Omega$,则 $S(\theta) < S(R) < S(\Omega)$,当 $U/R = \theta$ 时,$S(\theta) = \frac{1}{n}\sum_{i=1}^{n}(r_i \times 1^2/\sum_{j=1}^{n}r_j) = 1/n = 1/|U|$. 当 $U/R = \Omega$ 时,$S(\Omega) = |U|^2$,故 $1/|U| = S(\theta) \leqslant S(R) \leqslant S(\Omega) = |U|^2$.

性质 2　对于论域 U 的任意 2 个划分 P,Q,若 $P \prec Q$,则 $S(P) < S(Q)$.

证　重复应用引理 1 可证得 $S(P) < S(Q)$.

定义 6　称

$$I^*(R) = \begin{cases} 1 - \frac{1}{|U|^3 - 1}(|U|S(R) - 1), & |U| \neq 1 \\ 1, & |U| = 1 \end{cases}$$

为基于秩统计量的改进的知识 R 的含量测度,其中 $S(R)$ 为基于秩统计量的知识 R 颗粒块测度.

下面给出基于秩统计量的改进的知识 R 的含量测度的性质.

性质 3　设 (U,R_1),(U,R_2) 为 2 个知识库,若 $U/R_1 = U/R_2$,则 $I^*(R_1) = I^*(R_2)$.

性质 4　设 (U,R_1),(U,R_2) 为 2 个知识库,若 $R_1 \prec R_2$,则 $I^*(R_1) > I^*(R_2)$.

性质 5　设 U 为非空有限集,R 是 U 上的等价关系,则 $0 \leqslant I^*(R) \leqslant 1$.

证　当 $U/R = \Omega$ 时,$S(\Omega) = |U|^2$,$I^*(\Omega) = 0$; 当 $U/R = \theta$ 时,$S(\theta) = 1/|U|$,$I^*(\theta) = 1$; 所以对任意的等价关系 R,有 $\theta < R < \Omega$,由性质 4 知,$0 = I^*(\Omega) < I^*(R) < I^*(\theta) = 1$,故 $0 \leqslant I^*(R) \leqslant 1$.

2.3　一种新的粗糙集精度的度量方法

度量概念 X 的精度应该只与 $\overline{R}X$ 有关,而 $\Omega - \overline{R}X$ 与 X 度量无关,因此考虑精度的度量时应将论域 U 缩小到 $\overline{R}X$ 上进行研究. 先给出定义 R 在上近似集 $\overline{R}X$ 上的限制[12].

定义 7　设 U 为非空有限集,R 为 U 的 1 个划分,且 $R = \{R_1,R_2,\cdots,R_m\}(1 \leqslant m \leqslant |U|)$,$\forall X \subseteq U$,$X$ 关于 R 的上近似为 $\overline{R}X = R'_1 \cup R'_2 \cup \cdots \cup R'_k$,$k \leqslant m$,$R'_i \in R$,$1 \leqslant i \leqslant k$,则定义 R 在 $\overline{R}X$ 上的限制为 $R|_{\overline{R}X} = \{R'_1,R'_2,\cdots,R'_k\}$,显然有 $R|_{\overline{R}X}$ 是 $\overline{R}X$ 的

1 个划分.

下面给出 1 种新的粗糙集精度的度量方法定义.

定义 8 设 $K=(U,A)$ 为一知识库,R 是 U 上的 1 个等价关系,$\varnothing\subset X\subseteq U$,称 $\alpha_R^*(X)=1-\rho_R(X)(1-I^*(R\mid_{\overline{R}X}))$ 为基于秩统计量限制在上近似集考虑的改进知识含量测度的新的粗糙集精度度量,其中 $\rho_R(X)=1-|\underline{R}X|/|\overline{R}X|$.

由定义 8 可知,

$$\alpha^*(X)=\begin{cases}1-\rho(X)\left[\dfrac{1}{|\overline{R}X|^3-1}(|\overline{R}X|S(R_{\overline{R}X})-1)\right],|\overline{R}X|\neq 1,\\ 1, \quad |\overline{R}X|=1.\end{cases}$$

根据文献[5]定义的改进近似精度计算例 1 的结果为 $D_{R_1}(X)=0.854$,$D_{R_2}(X)=0.906$,$D_{R_3}(X)=0.931$,则有 $D_{R_1}(X)<D_{R_2}(X)<D_{R_3}(X)$;应用 $\alpha_R^*(X)$ 计算例 1 得 $\alpha_{R1}^*(X)=0.941\ 2$,$\alpha_{R2}^*(X)=0.982\ 2$,$\alpha_{R3}^*(X)=0.994\ 8$,则有 $\alpha_{R1}^*(X)<\alpha_{R2}^*(X)<\alpha_{R3}^*(X)$. 由例 1 说明本文定义的限制在上近似集考虑的改进知识含量测度新的粗糙集精度 $\alpha_R^*(X)$ 同样是可区分的,与文献[5]效果相同.

例 2 论域 $U=\{1,2,3,4,5,6,7,8,9\}$,集 $X'=\{1,2,3,4,5\}$,R_4,R_5,R_6 为 U 上的 3 个等价关系,分别导出的划分为

$U/R_4=\{\{1,2,3,4\},\{5,6\},\{7,8,9\}\}$,

$U/R_5=\{\{1,2,3,4\},\{5,6\},\{7,8\},\{9\}\}$,

$U/R_6=\{\{1,2,3,4\},\{5,6\},\{7\},\{8\},\{9\}\}$,

由传统的精度的定义可得

$\underline{R_4}X'=\underline{R_5}X'=\underline{R_6}X'=\{1,2,3,4\}$,$\overline{R_4}X'=\overline{R_5}X'=\overline{R_6}X'=\{1,2,3,4,5,6\}$,$\alpha_{R_4}(X')=\alpha_{R_5}(X')=\alpha_{R_6}(X')=2/3$.

而用文[5]定义的精度计算结果为 $D_{R_4}(X')=0.880\ 7$,$D_{R_5}(X')=0.897\ 1$,$D_{R_6}(X')=0.905\ 3$;应用 $\alpha_R^*(X)$ 计算例 2 可得结果为 $\alpha_{R4}^*(X')=\alpha_{R5}^*(X')=\alpha_{R6}^*(X')=0.945\ 7$,因此,对此例而言,当用知识 R 和 R' 去学习某一范畴 X 时,如果 $\overline{R}X=\overline{R'}X$,即上近似粒度完全一样,从直觉可知,此时学习的效果也应是一样的. 由粗糙集理论知,当应用已掌握的知识 R 去学习某一范畴 X 时,不确定性是由边界域的存在而产生的,而负区域对精度将不会产生影响. 所以对例 2 用本文定义的精度 $\alpha_R^*(X)$ 和传统精度计算比用文献[5]定义的精度计算更合理.

3 结论

本文给出的一种基于秩统计量限制在上近似集考虑的改进知识含量测度的新的粗糙集精度度量方法,克服了文献[5]中的精度度量未考虑论域的大小,当划分最细时精度突然变大,且当用 2 个不同的知识去学习某一同一范畴时,如果 2 个知识在这一范畴的上近似粒度完全一样,而学习效果不一样的局限性. 本文提出的方法是一种既考虑了知识颗粒块数据值的大小的信息又考虑了论域大小的精度度量方法,通过 2 个实例比较了文献[5]的精度的度量的结果,说明了本文的精度度量的合理性,且计算也不繁琐,是一种较优的精度定义.

4 参考文献

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

[2] Pawlak Z. Roughsets theoretical aspects of reasoning about data [M]. Boston: Kluwer Academic Publishers, 1991: 72-80.

[3] 张文修,吴伟志,梁吉业. 粗糙集理论与方法 [M]. 北京: 科学出版社, 2001: 1-97.

[4] Xu Baowen, Zhou Yuming, Lu Hongmin. An improved accuracy measure for rough sets [J]. Journal of Computer and Information Science, 2005, 71(2): 163-173.

[5] 刘纪芹,史开泉. 基于知识含量的粗糙集不确定性度量 [J]. 计算机科学, 2007, 34(7): 171-173.

[6] 周中华,吴根秀,连钢,等. 一种定义粗糙集精度的改进方法 [J]. 计算机工程, 2008, 34(6): 191-193.

[7] 李灿泽,吴根秀,晏伟峰,等. 覆盖粗糙集模型转化的方法研究 [J]. 江西师范大学学报: 自然科学版, 2011, 35(4): 391-395.

[8] 聂斌,王卓,杜建强,等. 基于偏最小二乘法的信息粒降维及聚类研究 [J]. 江西师范大学学报: 自然科学版, 2012, 36(5): 472-476.

[9] 吴根秀,王功,纪军,等. 多值信息系统的基于相似度的粗糙集模型 [J]. 江西师范大学学报: 自然科学版, 2011, 35(1): 88-90.

[10] 张玉琢. 一种利用属性序关系的变精度粗糙集知识约简方法 [J]. 江西师范大学学报: 自然科学版, 2008, 32(4): 449-453.

[11] 赵晓雨,雷晓蔚. 粗集模型的特征函数表示 [J]. 重庆师范大学学报: 自然科学版, 2007, 24(4): 54-57.

[12] 王静龙,梁小筠. 非参数统计分析 [M]. 北京: 高等教育出版社, 2011: 37-88.

The Accuracy Measure for Rough Sets Based on Rank Statistics

WU Gen-xiu, LIU Pei-hong, LUO Bing-hui, XIE Jun
(College of Mathematics and Informatics, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

Abstract: An accuracy measure for rough sets based on rank statistics has been proposed, which not only take into consideration the information of the size of the knowledge data value of particles but also the size of the universe. Meanwhile some properties are given. It is more reasonable and effective to measure, which are illustrated by examples.

Key words: rough sets; rank statistics; accuracy; improved knowledge capacity measure

(责任编辑:曾剑锋)

(上接第 19 页)

[9] 匡能晖. 关于两参数瑞利分布顺序统计量的分布性质[J]. 江西师范大学学报:自然科学版,2009,33(6):648-651.
[10] 张莉. 分组数据在指数分布下的近似极大似然估计[J]. 统计与决策,2009(17):153-155.
[11] 茆诗松,程依明,濮晓龙. 概率论与数理统计教程[M]. 2 版. 北京:高等教育出版社,2011.
[12] 李泽华,吴小腊,刘万荣. 变系数 EV 模型系数参数核估计的改进估计[J]. 重庆师范大学学报:自然科学版,2010,27(1):47-52.

The Estimations of Parameter from Эрланга Distribution under Missing Data Samples

LONG Bing
(Department of Mathematics and Physics, Jingchu University of Technology, Jingmen Hubei 448000, China)

Abstract: Firstly, density function of order statistics is given about Эрланга distribution, calculate related characteristic numbers of the order statistics. Secondly, in the missing data samples unbiased estimation approximate maximum likelihood estimation and confidence interval of parameter are given. In the end, several estimations of the parameter are calculated through an instance, and its superiority is proved.

Key words: Эрланга distribution; order statistics; unbiased estimation; approximate maximum likelihood estimation

(责任编辑:曾剑锋)