

# 信息系统中的知识距离与知识粗糙熵<sup>\*</sup>)

王宝丽 梁吉业

(山西大学计算智能与中文信息处理省部共建重点实验室 太原 030006)

(山西大学计算机与信息技术学院 太原 030006)

**摘要** 本文基于知识即划分的观点,证明了论域中所有知识构成一个距离空间,研究了知识距离空间中知识距离的一些性质,并应用知识距离度量信息系统中知识的粗糙熵。这些结果有利于更加深刻地分析信息系统中知识间的关系和知识的粗糙性。

**关键词** 信息系统,知识距离,粗糙熵

## Knowledge Distance and Rough Entropy in Information Systems

WANG Bao-Li LIANG Ji-Ye

(Key Laboratory of Ministry of Education for Computation Intelligence and Chinese Information Processing, Taiyuan 030006)

(School of Computer & Information Technology Shanxi University, Taiyuan 030006)

**Abstract** In the present paper, based on the view of knowledge induced by partition, all knowledge in the universe constructs a metric space, and the properties of knowledge distance are studied. Moreover, the knowledge distance is used to measure rough entropy of knowledge. These results will be very helpful for analyzing the relationships among knowledge and the roughness of knowledge thoroughly in information systems.

**Keywords** Information system, Knowledge distance, Rough entropy

## 1 引言

信息系统作为智能信息处理领域的一个具有形式化定义和明确语义的专业术语,出现于 Z. Pawlak 提出的粗糙集理论中<sup>[1]</sup>,该理论认为知识就是一种分类能力,给定论域上的一个等价关系,可以得到论域的一个划分,也就形成了论域上的一个知识基,目标概念即论域上的任意子集可用知识基中知识颗粒来近似刻画<sup>[2,3]</sup>。目前,粗糙集理论已被广泛应用于各个领域,特别是在数据挖掘领域获得了成功。张铃、张钹教授提出的商空间理论<sup>[4,5]</sup>,将人类从不同粒度世界之间跳转分析和解决问题的能力建立模型并将它用于动态规划、机器人路径规划等问题的解决中,获得了成功的应用。商空间理论也用子集来表示概念,一簇概念构成论域的划分形成的商空间也称知识基,并证明了论域上的所有划分(即知识库)构成一个完备格。在信息系统的知识度量方面,许多学者做了深入的研究,给出了知识的不确定性及粒度度量<sup>[6~8]</sup>。作者在文[8]中从信息论的角度对完备信息系统中知识粒度、信息熵及粗糙熵进行了深入研究,同时建立了它们之间的关系,并应用于信息系统中的知识约简、规则度量、属性重要性度量等方面。文[9]和文[10]也从信息论的角度借鉴熵的思想给出了不完备信息系统中一般二元关系下知识粗糙熵和粗糙集粗糙熵的定义。

本文从新的角度研究了信息系统中知识间关系和知识的粗糙性,给出知识距离定义,分析知识度量空间及性质。在此基础上,运用知识距离来定义知识的粗糙熵,得到新的定义下

的粗糙熵同样具有不变性和随知识的分辨能力的增强而单调下降的结论。

## 2 信息系统中的知识

形式上,四元组  $S=(U, A, V, f)$  是一个信息系统,其中  $U$ : 对象的非空有限集合,称为论域;  $A$ : 属性的非空有限集合,称为属性集;  $V=\cup_{a \in A} V_a, V_a$  表示属性  $a$  的值域;  $f: U \times A \rightarrow V$  是一个信息函数,即对  $\forall x \in U, a \in A$ , 有  $f(x, a) \in V_a$ 。通常  $S=(U, A, V, f)$  也简记为  $S=(U, A)$ ,  $f(x, a)$  简记为  $a(x)$ 。

如果  $\forall u \in U, \forall a \in A, a(u)$  是一个确定的值,则信息系统  $S$  称为完备信息系统,否则称为不完备信息系统。在完备信息系统中,任意  $P \subseteq A$  决定等价关系:  $ind(P) = \{u, v\} \in U \times U \mid \forall a \in P, a(u) = a(v)\}$ 。  $U/ind(P) = \{P_1, P_2, \dots, P_m\}$  构成了论域  $U$  的一个划分,并称其为论域  $U$  上的一个知识<sup>[2]</sup>。在不产生混淆的情况下用  $P$  代替  $U/ind(P)$ , 称为知识  $P$ 。  $P_i (i=1, 2, \dots, m)$  是知识粒。设  $\mathcal{K}$  是信息系统中所有知识构成的集合,则称  $ind(S) = (U, \mathcal{K})$  为信息系统的知识库。如果对两个知识  $P, Q$ , 有  $U/ind(Q)$ , 则称知识  $P, Q$  无区别(或知识  $P, Q$  相同)。

为方便后面各节讨论,这里采用一种新的知识  $P$  的表达  $U/ind(P) = \{I_P(u_1), I_P(u_2), \dots, I_P(u_n)\}$  (元素允许有重复)。其中  $I_P(u_i)$  为所有与  $u_i$  不可区分的对象构成的等价类,也是包含  $u_i$  的一个知识颗粒。显然若  $u_j \in I_P(u_i)$ , 则  $I_P(u_j) = I_P(u_i)$ 。

<sup>\*</sup> 本文受到国家自然科学基金(70471003, 60275019)、高等学校博士学科点专项科研基金(20050108004)、山西省高等学校拔尖创新人才基金、教育部科学技术研究重点项目资助。王宝丽 硕士研究生,主要研究方向为粗糙集理论、粒度计算;梁吉业 教授,博士生导师,主要研究领域为粗糙集理论、数据挖掘、人工智能。

表1 一个信息系统

S	颜色	形状	体积
$u_1$	红	方	大
$u_2$	蓝	方	大
$u_3$	蓝	方	大
$u_4$	红	圆	大
$u_5$	蓝	圆	小
$u_6$	蓝	圆	小

例1 考虑表1给出的完备信息系统,这里  $S=(U,A)$ ,  $U=\{u_1, u_2, \dots, u_6\}$ ,  $A=\{\text{颜色, 形状, 体积}\}$ 。

$I_A(u_1)=\{u_1\}$ 表示颜色为红、形状为方、体积为大的信息粒。

$I_A(u_2)=I_A(u_3)=\{u_2, u_3\}$ 表示颜色为蓝、形状为方、体积为大的信息粒。

$I_A(u_4)=\{u_4\}$ 表示颜色为红、形状为圆、体积为大的信息粒。

$I_A(u_5)=I_A(u_6)=\{u_5, u_6\}$ 表示颜色为蓝、形状为圆、体积为小的信息粒。

信息系统中的知识A为

$U/ind(A)$

$$= \{I_A(u_1), I_A(u_2), I(u_3), I_A(u_4), I_A(u_5), I_A(u_6)\}$$

$$= \{\{u_1\}, \{u_2, u_3\}, \{u_2, u_3\} \cup \{u_4\}, \{u_5, u_6\}, \{u_5, u_6\}\}$$

### 3 知识距离及性质

根据知识即划分的思想,论域的不同划分构成了论域中不同的知识,且不同知识对同一概念的表达一般是有区别的,这一点在粗糙集理论中有深入的研究。在文[8]中,作者对具体的某一知识的在粗糙集理论中度量性质做了深入的研究并已应用于信息系统的知识约简、规则度量、属性重要性度量等方面。对于不同知识之间的关系及性质,目前仅在商空间理论中证明了知识库中所有知识构成有一定层次关系的完备格<sup>[4]</sup>。本小节从集合对称差的角度给出知识距离的定义,并分析了知识距离的一些性质。

信息系统中不同的属性集一般会诱导出论域的不同划分。考虑知识之间的距离也就是考察不同属性集之间的差别<sup>[3]</sup>。不失一般性,这里讨论的知识库是  $U=\{u_1, u_2, \dots, u_n\}$  论域上所有可能的知识(即所有可能的划分)构成的全总知识库,记为  $K(U)$ ,给定信息系统  $S=(U,A)$ ,其知识库  $ind(S)$  只是  $K(U)$  的子集。

为方便讨论,引入知识序关系的定义。

定义3.1<sup>[8]</sup> 知识库  $K(U)$  中的两个知识  $U/ind(P)=\{I_P(u_1), I_P(u_2), \dots, I_P(u_n)\}$ ,  $U/ind(Q)=\{I_Q(u_1), I_Q(u_2), \dots, I_Q(u_n)\}$ ,

若满足  $I_P(u_i) \subseteq I_Q(u_i), \forall 1 \leq i \leq n$ , 则称知识P比知识Q精细,记为  $P < Q$ 。

定义3.2  $K(U)$  中的两个知识  $P, Q, U/ind(P)=\{I_P(u_1), I_P(u_2), \dots, I_P(u_n)\}$ ,  $U/ind(Q)=\{I_Q(u_1), I_Q(u_2), \dots, I_Q(u_n)\}$ , 定义P, Q之间的距离为

$$d(P, Q) = \frac{1}{|U| \times (|U| - 1)} \sum_{i=1}^{|U|} |I_P(u_i) \oplus I_Q(u_i)|$$

其中  $\oplus$  为集合的对称差运算,  $|\cdot|$  表示集合的基数。

定理3.1  $(K(U), d)$  构成距离空间,即对于  $\forall P, Q, R \in K(U)$  满足:

(1)  $d(P, Q) \geq 0$ , 当且仅当  $P=Q$  时不等式取“=”;

(2)  $d(P, Q) = d(Q, P)$ ;

(3)  $d(P, R) \leq d(P, Q) + d(Q, R)$  成立。

证明:(1)、(2)显然成立,下证(3)成立。

设  $\forall A, B, C \subseteq U$ , 由集合论的知识有  $A \oplus B \subseteq (A \oplus C) \cup (C \oplus B)$  成立,所以对任意的  $u_i \in U (i=1, 2, \dots, n)$  有

$$|I_P(u_i) \oplus I_R(u_i)|$$

$$\leq |(I_P(u_i) \oplus I_Q(u_i)) \cup (I_Q(u_i) \oplus I_R(u_i))|$$

$$\leq |I_P(u_i) \oplus I_Q(u_i)| + |I_Q(u_i) \oplus I_R(u_i)|$$

所以

$$\sum_{i=1}^n |I_P(u_i) \oplus I_R(u_i)| \leq$$

$$(\sum_{i=1}^n |I_P(u_i) \oplus I_Q(u_i)| + \sum_{i=1}^n |I_Q(u_i) \oplus I_R(u_i)|)$$

不等式两边同乘一个常数  $\frac{1}{|U| \times (|U| - 1)}$  就可以得到  $d(P, R) \leq d(P, Q) + d(Q, R)$  成立。证毕。

给定信息系统  $S=(U, A)$ , 其知识库  $ind(S)$  只是  $K(U)$  的子集,由度量空间的知识可知  $(ind(S), d)$  是  $(K(U), d)$  的距离子空间。

性质3.1 给定论域  $U(|U| > 1)$ , 则知识库  $K(U)$  中两个不同知识的距离最大值可达到1, 最小可达到

$$\frac{2}{|U| \times (|U| - 1)}$$

证明:显然,对于知识库中任意两个不同的知识  $P, Q$ , 至少存在两个对象  $u_i, u_j \in U$ , 使得  $I_P(u_i) \neq I_Q(u_i), I_P(u_j) \neq I_Q(u_j)$  成立。

设  $A, B \subseteq U$ , 且  $A \neq \emptyset, B \neq \emptyset, A \neq B$ , 则根据集合论知识有  $1 \leq |A \oplus B| \leq |U| - 1$  成立。

给定论域  $U$ , 存在两个特殊的知识:知识完备格中最底层的知识即相等关系  $\omega$  对应的知识  $U/\omega$  和最高层的知识即全域关系  $\delta$  对应的知识  $U/\delta$ :

$$U/\omega = \{\{u_1\}, \{u_2\}, \dots, \{u_n\}\}$$

$$U/\delta = \{\{u_1, u_2, \dots, u_n\}\}$$

也记为知识  $\omega$  和知识  $\delta$ 。运用距离公式求二者距离  $d(\omega, \delta)$ , 公式的第二个因式中的每一个加项都能取到最大值  $|U| - 1$ , 这样二者的距离一定取到最大值,即  $d(\omega, \delta) = 1$ 。

设知识库中两个不同的知识  $P, Q$  为

$$U/ind(P) = \{\dots, \{u_i, u_j\}, \dots, \{u_i, u_j\}, \dots\}$$

$$U/ind(Q) = \{\dots, \{u_i\}, \dots, \{u_j\}, \dots\}$$

其中  $\dots$  表示其余的知识颗粒相同。运用距离公式求  $d(P, Q)$ , 公式的第二个因式中仅有2个加项为1, 其余加项均为0, 而又由于  $P, Q$  是不同的两个知识,由前面分析可知这种情况下距离可以达到最小值,即  $d(P, Q) = \frac{2}{|U| \times (|U| - 1)}$ 。证毕。

定理3.2  $P, Q, R$  为  $K(U)$  中任意的3个知识,若满足  $P < Q < R$ , 则必有  $d(P, R) = d(P, Q) + d(Q, R)$  成立。

证明:由于  $P < Q < R$  成立,设具有序关系的知识其中一个知识粒改变如下,其余颗粒不变。

$$R: \{L(k)\},$$

$$Q: \{L(k-m)\}, \{L(m)\},$$

$$P: \{L(k-m-l)\}, \{L(l)\}, \{L(n)\}, \{L(m-n)\}$$

( $L(k)$  表示该粒中有  $k$  个元素,其余类似)其中,

$$1 \leq k \leq |U|, 0 \leq m \leq k$$

$$0 \leq l \leq k-m, 0 \leq n \leq m$$

$$\begin{aligned} & \text{为表达简洁,记 } \alpha = \frac{1}{|U| \times (|U|-1)} \\ d(P,R) &= \alpha \times ((k-m-l)(m+l) + l(k-l) + \\ & \quad n(k-n) + (m-n)(k-m+n)) \\ &= \alpha \times (2km + 2kl + 2mn - 2ml \\ & \quad - 2m^2 - 2n^2 - 2l^2) \\ d(P,Q) &= \alpha \times (2l(k-m-l) + 2n(m-n)) \\ d(Q,R) &= \alpha \times (2m(k-m)) \\ d(P,Q) + d(Q,R) &= \alpha \times (2km + 2kl + 2mn - \\ & \quad 2ml - 2m^2 - 2n^2 - 2l^2) \end{aligned}$$

所以,  $d(P,R) = d(P,Q) + d(Q,R)$ 。

这里仅讨论  $R$  中的一个粒的二分裂分解过程,并且每次的分裂不超过二分裂为例,对于多粒多分裂的情况仅是单粒分解的重复,证明过程仅仅是计算量的增大,分析类似。证毕。

但定理 3.2 的条件不是结论成立的必要条件。

**例 2** 知识库中 3 个知识  $P_1, P_2, P_3$  表达如下:

$$\begin{aligned} U/\text{ind}(P_1) &= \{\{u_1, u_2\}, \{u_1, u_2\}, \{u_3, u_4\}, \\ & \quad \{u_3, u_4\}, \{u_5, u_6\}, \{u_5, u_6\}\}, \\ U/\text{ind}(P_2) &= \{\{u_1, u_2\}, \{u_1, u_2\}, \{u_3\}, \{u_4\}, \{u_5, u_6\}, \\ & \quad \{u_5, u_6\}\}, \\ U/\text{ind}(P_3) &= \{\{u_1, u_2, u_3\}, \{u_1, u_2, u_3\}, \{u_1, u_2, u_3\}, \{u_4, \\ & \quad u_5, u_6\}, \{u_4, u_5, u_6\}, \{u_4, u_5, u_6\}\}. \\ d(P_1, P_2) &= \frac{1}{6 \times (6-1)} \times (0+0+1+1+0+0) = \frac{1}{15}, \\ d(P_2, P_3) &= \frac{1}{6 \times (6-1)} \times (1+1+2+2+1+1) = \frac{4}{15}, \\ d(P_1, P_3) &= \frac{1}{6 \times (6-1)} \times (1+1+3+3+1+1) = \frac{5}{15} = \end{aligned}$$

$\frac{1}{3}$ , 有  $d(P_1, P_3) = d(P_1, P_2) + d(P_2, P_3)$  成立。

显然,  $P_1, P_2, P_3$  并不满足定理 3.2 的条件,却有定理 3.2 的结论成立。

$K(U)$  的知识按 3.1 所定义的序关系对应着一个哈斯图。满足定理 3.2 条件的三个知识必在同一条线上。这一结论与欧式空间的距离性质有类似之处。但在欧式距离空间中,三点不在同一条直线上必取不等号,但此性质在以上定义的知识空间中是不成立的,如例 2 所分析。

#### 4 知识的粗糙熵

知识库是具有一定层次的。显然,处于知识库最底层的知识  $\omega$  最不粗糙,而处于最高层知识  $\delta$  最粗糙。在粗糙集理论中,用知识  $R$  刻画目标概念可能会产生不精确表达,这是由于知识的分辨能力不强引起的,因此给定论域上的知识  $R$ ,其粗糙熵的度量是有意义的<sup>[8~10]</sup>。本小节根据第 3 小节的知识距离来定义信息系统中知识的粗糙熵。

**定义 4.1** 信息系统  $S=(U,A)$  中,知识  $P \in \text{ind}(S)$  的粗糙熵  $E_r(P)$  定义为

$$\begin{aligned} E_r(P) &= d(P, \omega) \\ &= \frac{1}{|U| \times (|U|-1)} \sum_{i=1}^{|U|} (|I_P(u_i)| \oplus \{u_i\}) \end{aligned}$$

其中  $U/\text{ind}(P) = \{I_P(u_1), I_P(u_2), \dots, I_P(u_n)\}$ ,  $\omega$  为知识库中最底层的知识  $U/\omega = \{\{u_1\}, \{u_2\}, \dots, \{u_n\}\}$ 。

**例 3** 计算例 2 中知识  $P_1, P_2, P_3$  的粗糙熵。

$$E_r(P_1) = \frac{3}{15}, E_r(P_2) = \frac{2}{15}, E_r(P_3) = \frac{6}{15}$$

可见 3 个知识中,  $P_2$  最不粗糙,  $P_3$  最粗糙。

**性质 4.1** 信息系统  $S=(U,A)$  中两个知识  $P, Q \in \text{ind}(S)$ , 知识距离  $d(P,Q)$  与粗糙熵  $E_r(P), E_r(Q)$  有以下关系:

$$d(P,Q) \geq |E_r(P) - E_r(Q)|$$

**证明:** 要证  $d(P,Q) \geq |E_r(P) - E_r(Q)|$  成立,由  $d(P,Q) \geq 0$ , 即证下式

$$d(P,Q) \geq E_r(P) - E_r(Q) \geq -d(P,Q) \text{ 成立, 对于知识 } P,$$

$Q, \omega$ , 由定理 3.1 知

$$d(P,Q) + d(Q, \omega) \geq d(P, \omega)$$

$$d(Q,P) + d(P, \omega) \geq d(Q, \omega)$$

即

$$d(P,Q) \geq d(P, \omega) - d(Q, \omega)$$

$$d(Q,P) \geq d(Q, \omega) - d(P, \omega)$$

而

$$d(P,Q) = d(Q,P)$$

所以,

$$d(P,Q) \geq d(P, \omega) - d(Q, \omega) \geq -d(P,Q)$$

即

$$d(P,Q) \geq E_r(P) - E_r(Q) \geq -d(P,Q)$$

所以

$$d(P,Q) \geq |E_r(P) - E_r(Q)|$$

**性质 4.2** 信息系统  $S=(U,A)$  中两个知识  $P, Q \in \text{ind}(S)$ , 如果  $P < Q$  成立, 则  $d(P,Q) = E_r(Q) - E_r(P)$ 。

**证明:** 显然有  $\omega < P < Q$  成立, 由定理 3.2 知:

$$d(\omega, P) + d(P, Q) = d(\omega, Q)$$

即

$$d(P, Q) = d(\omega, Q) - d(\omega, P)$$

亦即

$$d(P, Q) = E_r(Q) - E_r(P) \text{ 成立.}$$

性质 4.2 表明: 如果两个知识之间具有序关系, 可以根据知识的粗糙熵的差来计算两个知识之间的距离。显然, 如果  $d(P,Q) > |E_r(P) - E_r(Q)|$ , 一定没有序关系成立。

**定理 4.1** 信息系统  $S=(U,A)$  中两个知识  $P, Q \in \text{ind}(S)$  分别表示为

$$U/\text{ind}(P) = \{I_P(u_1), I_P(u_2), \dots, I_P(u_n)\}$$

$$U/\text{ind}(Q) = \{I_Q(u_1), I_Q(u_2), \dots, I_Q(u_n)\}$$

若存在一个一一映射  $f: U \rightarrow U$ , 使  $|I_P(u_i)| = |I_Q(f(u_i))|$ , ( $i=1, 2, \dots, n$ ) 成立,

则  $E_r(P) = E_r(Q)$ 。

**证明:** 显然。

定理 4.1 称为知识的粗糙不变性。

**定理 4.2** 信息系统  $S=(U,A)$  中两个知识  $P, Q \in \text{ind}(S)$ , 若序关系  $P < Q$  存在, 则  $E_r(P) \leq E_r(Q)$ 。

**证明:** 由已知条件  $P < Q$  和性质 4.2 知  $d(P,Q) = E_r(Q) - E_r(P)$  成立。

而

$$d(P,Q) \geq 0,$$

所以

$$E_r(Q) - E_r(P) \geq 0,$$

即

$$E_r(P) \leq E_r(Q).$$

定理 4.2 表明, 知识的粗糙熵随知识的分辨能力增强而单调减少。

这里根据知识距离定义的粗糙熵给信息系统中粗糙熵一

# Huber-SVR 中参数 $\mu$ 与输入噪声间的近似线性关系<sup>\*</sup>)

周晓剑 朱嘉钢 王士同

(江南大学信息工程学院 无锡 214122)

**摘要** 为使 Huber-SVR 更具鲁棒性,深入研究了 Huber-SVR 中参数与输入噪声之间的关系。运用 SVR 的贝叶斯框架,分别推导出了鲁棒的 Huber-SVR 中参数  $\mu$  与拉斯噪声和均匀噪声之间呈近似线性关系,并结合仿真结果和已有的相关结论,得到了更为一般的结论,即鲁棒的 Huber-SVR 中参数  $\mu$  与输入噪声之间呈近似线性关系。这一结论为输入样本含有分布未知噪声的情况下 Huber-SVR 参数的选择提供了理论依据。

**关键词** 支持向量机,支持向量回归机,Huber 损失函数

## Approximately Linear Dependency between $\mu$ and the Input Noise in Huber-Support Vector Regression

ZHOU Xiao-Jian ZHU Jia-Gang WANG Shi-Tong

(School of Information Engineering, Southern Yangtze University, Wuxi 214122)

**Abstract** The dependency relationship between  $\mu$  and the input noise in Huber-SVR is studied using SVR Bayesian evidence framework. First, focus is paid on the cases of Laplacian noise and Uniform noise, and the approximately linear dependencies between  $\mu$  and the variances of the two noises are then respectively derived. Second, with the relevant conclusion on Huber-SVR and experimental study, the more general claim is then proposed that the approximately linear dependency is almost kept between  $\mu$  and the input noise in Huber-SVR. Such a dependency relationship is useful to determine the optimal choice for  $\mu$  in Huber-loss function in the existence of unknown input noise.

**Keywords** Support vector machines(SVM), Support vector regression(SVR), Huber loss function

### 1 引言

近年来,SVM 广泛地应用于各种分类和回归问题中。当 SVM 应用于回归分析和预测时,通常称之为 SVR。由于真实的数据经常含有噪声,因此在用 SVR 进行回归时,如何使 SVR 具有更好的鲁棒性,即如何选择损失函数的参数使得 SVR 具有尽可能强的抑制噪声能力,就成了一个重要的研究课题。为此一些学者进行了研究,重要成果有:文[1,2]提出贝叶斯框架理论,该理论指出 SVR 优化问题等价于求对应的最大后验概率问题(MAP 问题)。文[3]在贝叶斯框架下推导出重要结论,即  $\epsilon$ -SVR 中参数  $\epsilon$  和输入噪声(高斯噪声、拉斯噪声、均匀噪声)间呈线性关系。文[4,5]在贝叶斯框架下推

导出了 Huber-SVR 中的参数  $\mu$  与高斯输入噪声间呈近似线性关系。实际中,除了高斯噪声外,拉斯噪声和均匀噪声也是典型的噪声。在文[4,5]的基础上,研究 Huber-SVR 中的参数  $\mu$  与拉斯噪声和均匀噪声间的关系,可以进一步得出 Huber-SVR 中的参数  $\mu$  与输入噪声间的更一般的关系,从而为输入噪声分布未知情形下 Huber-SVR 的参数选择提供理论依据,本文的目的就在于此。本文运用 SVR 的贝叶斯框架,推导出 Huber-SVR 中的参数  $\mu$  与拉斯噪声和均匀噪声之间均呈近似线性关系,这一理论推导结果也与实验结果相吻合。进而,在文[4,5]的基础上,综合理论和实验两方面结果,得出 Huber-SVR 的参数与输入噪声间总是呈近似线性关系这一更一般的结论。

<sup>\*</sup>)江苏省自然科学基金(BK2003017);教育部跨世纪优秀人才计划资助项目;江苏计算机技术重点实验室开放课题资助项目。周晓剑 硕士研究生,从事软件理论、人工神经网络等研究;朱嘉钢 博士,副教授,从事人工神经网络、模式识别等研究;王士同 教授,博士生导师,CCF 高级会员,从事人工神经网络、模式识别、生物信息、模糊系统等研究。

种新的解释,而且以上定理表明通过知识距离定义的粗糙熵与其它定义下的粗糙熵也同样具有粗糙不变性和随知识分辨能力增强而单调减少的性质。

**结束语** 基于知识即划分的观点,本文证明了论域中所有知识构成一个距离空间,研究了知识空间中知识距离的一些性质,并从一个全新的角度运用知识距离定义了信息系统中的知识粗糙熵,为信息系统中知识度量研究提供了新工具,有利于更加深刻地分析信息系统中知识间的关系和知识的粗糙性。

### 参考文献

- 1 Pawlak Z. Rough Sets; Theoretical Aspects of Reasoning about Data. Dordrecht; Kluwer Academic Publisher, 1991
- 2 梁吉业,李德玉. 信息系统中的不确定性与知识获取. 北京:科

- 学出版社,2005
- 3 管延勇,王洪凯,史开泉. 知识的粗识别及其评判. 计算机科学, 2004, 31 (12): 113~116
- 4 张钊,张铃. 问题求解理论及应用. 北京:清华大学出版社,1990
- 5 张钊,张铃. 模糊商空间理论(模糊粒度计算方法). 软件学报, 2003, 14 (4): 770~776
- 6 Beaubouef T, Petry F E, Arora G. Information theoretic measures of uncertainty for rough sets and rough relational databases. Information Science, 1998, 109 (1-4): 185~195
- 7 Wierman M J. Measuring uncertainty in rough set theory. International Journal of General Systems, 1999, 28 (4): 283~397
- 8 Liang J Y, Shi Z Z. The information entropy, rough entropy and knowledge granulation in rough set theory. International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, 2004, 12 (1): 37~46
- 9 黄兵,何新,周献中. 基于广义粗集覆盖约简的粗糙熵. 软件学报, 2004, 15 (2): 215~220
- 10 黄兵,周献中,史迎春. 基于一般二元关系的知识粗糙熵与粗集粗糙熵. 系统工程理论与实践, 2004 (1): 93~96