基于粒计算的 Rough 集模型

陈泽华 谢 刚 谢 珺 谢克明
(太原理工大学信息工程学院 太原 030024)

摘 要 上近似、下近似是 Rough 集的基本定义,它使我们能够用精确的集合讨论不精确的概念,Rough 集利用可计算的边界域实现了 G. Frege 的边界思想。然而,Rough 集本身的代数定义和其他各种扩展模型并没有提供简单直观的计算边界元素数目的算法。在二进制粒计算的基础上,通过定义粒矩阵和粒矩阵运算,建立了基于粒计算的知识表示方法和基于粒计算的 Rough 集模型,据此可以获得 Rough 集基本概念的粒矩阵表示和粒矩阵快速计算方法,为建立基于粒计算的知识发现算法提供了理论基础。举例证明了 Rough 包含与 Rough 相等的隶属度函数定义并非充要条件。同时给出了基于粒计算的 Rough 包含与 Rough 相等的充要条件。

关键词 粒计算,Rough 集理论,粒矩阵,粒关系矩阵
中图法分类号 TP182

GrC-based Rough Set Model

CHEN Ze-hua XIE Gang XIE Jun XIE Ke-ming
(College of Information Engineering of Taiyuan University of Technology, Taiyuan 030024, China)

Abstract Upper approximation and lower approximation are the basic definitions in Rough Set Theory (RST), it makes vague boundary computable, however, none of existing definition of RST brings efficient way to compute boundary. Based on Bit Granular Computing, Granular Matrix and its operation were defined and GrC-based RST model was established to complete the basic definition and computation of RST. The new model builds theoretic foundations for GrC-based knowledge discovery algorithms. Furthermore, modified sufficient and necessary conditions for rough inclusion and rough equivalent was proposed, some examples were given to illustrate the efficiency of the proposed model.

Keywords Granular computing (GrC), Rough set theory (RST), Granular matrix (GrM), Granular relation matrix

同一问题在不同知识表示下的算法难度是不同的[1]。经典 Rough 集理论的代数模型直观性差,知识约简计算繁复,其它 Rough 集模型如信息熵模型、包含度模型、信息粒度模型等[2-5],从本质上讲都是依赖于概率运算的。粒计算的提出为我们解决问题提供了新的思路[6-9]。文献[10]提出用二进制粒表示粗糙集概念,获取关联规则。本文在此基础上,通过定义粒(粒矩阵)和粒运算(矩阵运算),建立了基于粒计算的 Rough 集模型。该模型不仅意义直观且便于计算,同时为建立基于粒计算的知识约简算法提供了理论基础。

1 Rough 集合的粒定义与粒运算

1.1 等价类的二进制粒定义[10]

定义 1 设 $K=(U,R)$是一个知识库,其中 $U=\{x_1,\cdots x_k,\cdots,x_l\}$是论域,$R$ 是属性集。任意子集 $B\subseteq R$ 均可将 U 划分成互不相交的等价类 $U/IND(B)=\{Y_1,\cdots,Y_i,\cdots,Y_m\}$ $(1\leqslant i\leqslant m)$,等价类被定义成粒。$Y_i$ 可用一个长度为 l 的二进制数定义,如果 $x_k\in Y_i$,对应的二进制数的第 i 位为 1,否则为 0,其中 l 为 U 的势。取值为 1 的元素具有不可分辨关系。

1.2 二进制粒矩阵及其运算

给定知识库 $K=(U,R)$,设 P,Q 属性在论域 U 上导出的分类分别为:

$U/IND(P)=\{Y_1,Y_2,\cdots,Y_i,\cdots,Y_m\}$, $1\leqslant i\leqslant m$

$U/IND(Q)=\{X_1,X_2,\cdots,X_j,\cdots,X_n\}$, $1\leqslant j\leqslant n$

由定义 1 可知 Y_i,X_j 的二进制位串分别为

$$Y_i=\{a_{i1},a_{i2},\cdots a_{ik},\cdots,a_{il}\} \tag{1}$$

$$X_j=\{b_{j1},b_{j2},\cdots b_{jk},\cdots,b_{jl}\} \tag{2}$$

其中:

$$a_{ik}=\begin{cases}1, & \text{if } x_k\in Y_i\\ 0, & \text{if } x_k\notin Y_i\end{cases} \quad 1\leqslant k\leqslant l \tag{3}$$

$$b_{jk}=\begin{cases}1, & \text{if } x_k\in X_j\\ 0, & \text{if } x_k\notin X_j\end{cases} \quad 1\leqslant k\leqslant l \tag{4}$$

定义 2 粒矩阵(Granular Matrix, GrM)定义为$\{Y_{m\times l},X_{n\times l},C_{m\times n}\}$,其中:

$$Y_{m\times l}\triangleq Y=\begin{pmatrix}Y_1\\ \cdots\\ Y_m\end{pmatrix}=\begin{pmatrix}a_{11} & \cdots & a_{1l}\\ \cdots & \cdots & \cdots\\ a_{m1} & \cdots & a_{ml}\end{pmatrix} \tag{5}$$

$$X_{n\times l}\triangleq X=\begin{pmatrix}X_1\\ \cdots\\ X_n\end{pmatrix}=\begin{pmatrix}b_{11} & \cdots & b_{1l}\\ \cdots & \cdots & \cdots\\ b_{n1} & \cdots & b_{nl}\end{pmatrix} \tag{6}$$

到稿日期:2008-09-17 本文受山西省自然科学基金项目(20051037),高校博士点专项科研基金项目(20060112005)与山西省青年自然科学基金项目(2007021018)资助。

谢克明 博士生导师,教授,E-mail: chenzehua@tyut.edu.cn。

$$C_{m\times n}\triangleq C_{YX}\triangleq Y\times X'=[c_{ij}]_{m\times n} \tag{7}$$

$$c_{ij}=\sum_{k=1}^{l}(a_{ik}b_{kj})=card(Y_i\cap X_j)=|Y_i\cap X_j| \tag{8}$$

矩阵 Y,X 表示了知识 P 和知识 Q 导出的等价类，矩阵 $C_{m\times n}$ 反映了所有等价类 Y_i 与 X_j 之间的包含关系。定义 $C_{m\times n}$ 为知识 P,Q 的粒关系矩阵，元素 c_{ij} 反映了 Y_i 包含于 X_j 中元素的个数。$card(\cdot),|\cdot|$ 均表示集合的势。定义

$$card(Y)\triangleq|Y|=\begin{pmatrix}|Y_1|\\|Y_2|\\\cdots\\|Y_m|\end{pmatrix}=\begin{pmatrix}\sum_{k=1}^{l}a_{1k}\\\sum_{k=1}^{l}a_{2k}\\\cdots\\\sum_{k=1}^{l}a_{mk}\end{pmatrix} \tag{9}$$

$$card(U)=|U|=l \tag{10}$$

特殊地，若 X 为 U 中任一子集，粒关系矩阵 $C_{m\times n}$ 变为一个列矩阵 $C_m=[c_i]_{m\times 1}$。

2 基于粒计算的 Rough 集模型

2.1 基于粒矩阵的知识和知识库

在粒矩阵定义下，知识转变为矩阵，一个知识库 $K=(U,R)$ 对应着一族二进制粒矩阵。

2.2 基于粒矩阵的上近似和下近似

在粒矩阵定义下，可以给出 Rough 集合上、下近似和边界的数值计算公式。

定义 3 给定知识库 $K=(U,R)$，对于任一子集 $X\subseteq U$ 和一个等价关系 $P\subseteq R$，可以根据粒关系矩阵 C 来划分集合 X，于是有：

$$R_*X=\{\bigcup Y_i\,|\,i(c_i=|Y_i|)\} \tag{11}$$

$$R^*X=\{\bigcup Y_i\,|\,i(c_i\neq 0)\} \tag{12}$$

$$BN_R=R^*X-R_*X \tag{13}$$

$$card(R_*(X))=\sum_{i=1}^{m}\{c_i\,|\,c_i=|Y_i|\} \tag{14}$$

$$card(R^*(X))=\sum_{i=1}^{m}c_i \tag{15}$$

对于知识 Q 产生的分类：$U/IND(Q)=\{X_1,X_2,\cdots,X_j,\cdots,X_n\}$ 则有

$$R_*X=POS_P(Q)=\{\bigcup Y_i\,|\,NE(i)=1\} \tag{16}$$

$$R^*X=\{\bigcup Y_i\,|\,i(c_{ij}\neq 0)\} \tag{17}$$

$$BN_R=R^*X-R_*X \tag{18}$$

$$card(R_*(X))=\sum_{i=1j=1}^{m\ n}\{c_{ij}\,|\,NE(i)=1)\} \tag{19}$$

$$card(R^*(X))=\sum_{i=1j=1}^{m\ n}c_{ij} \tag{20}$$

$NE(i)$ 是粒关系矩阵 $C_{m\times n}$ 中第 i 行非零元素的个数。$POS_P(Q)$ 表示 Q 的 P 正域。

2.3 基于粒矩阵的集合的精确度

定义 4 由粒矩阵定义的集合的精确度为

$$\alpha_R(X)=\frac{card(R_*(X))}{card(R^*(X))}=\frac{\sum_{i=1}^{m}\{c_i\,|\,c_i=|Y_i|\}}{\sum_{i=1}^{m}c_i} \tag{21}$$

2.4 基于粒矩阵的分类的精确性

定义 5 由粒矩阵定义的描述分类精确性的度量为

第一个量度：分类的精度定义为

$$\alpha_R(F)=\frac{\sum card(R_*X_i)}{\sum card(R^*X_i)}=\frac{\sum_{i=1j=1}^{m\ n}\{c_{ij}\,|\,NE(i)=1)\}}{\sum_{i=1j=1}^{m\ n}c_{ij}} \tag{22}$$

第二个量度：分类的质量定义为

$$\gamma_R(F)=\frac{\sum card(R_*X_i)}{card U}=\frac{1}{l}\sum_{i=1j=1}^{m\ n}\{c_{ij}\,|\,NE(i)=1)\} \tag{23}$$

2.5 基于粒矩阵的知识的依赖性

定义 6 由粒矩阵定义的知识的依赖性为

$$k=\gamma_P(Q)=\frac{|POS_P(Q)|}{|U|}=\frac{1}{l}\sum_{i=1j=1}^{m\ n}\{c_{ij}\,|\,NE(i)=1)\} \tag{24}$$

2.6 基于粒矩阵的 Rough 隶属度函数

定义 7 通过不可分辨关系定义论域任一元素 x 对任一集合 X 的 Rough 函数隶属度函数为式(25)[11]：

$$\mu_X^R(x)=\frac{card(X\cap R(x))}{card(R(x))} \tag{25}$$

其中：$0\leqslant\mu_X^R(x)\leqslant 1$，$R(x)$ 是包含元素 x 的等价类。

定义 8 给定论域 U 和 U 上的不可分辨关系 R，$U/IND(R)=\{Y_1,\cdots Y_i,\cdots Y_m\}(1\leqslant i\leqslant m)$，通过粒矩阵定义元素 x 对任一集合 X 的 Rough 隶属度函数可写为

$$\mu_X^R(x)=\mu_X^R(Y_i)=\frac{|X\cap Y_i|}{|Y_i|}=\frac{c_i}{|Y_i|} \tag{26}$$

其中：Y_i 是包含元素 x 的等价类，$\mu_X^R(Y_i)$ 表示包含 x 在内的等价类的隶数度函数。可见定义 7,8 是一致的。

2.7 基于粒矩阵的 Rough 包含

定义 9 假定给定两个集合 $A,B\subseteq U$ 和 U 上的不可分辨关系 R，根据代数定义，定义集合 A,B 的 3 种包含：

1) 集合 A 和 B 为 R 下 Rough 包含：

$A\subseteq_{*R}B\Leftrightarrow\forall\mu_A^R(x)=1,\exists\mu_B^R(x)=1$

2) 集合 A 和 B 为 R 上 Rough 包含：

$A\subseteq_R^* B\Leftrightarrow\forall\mu_A^R(x)\neq 0,\exists\mu_B^R(x)\neq 0$

3) 集合 A 和 B 为 R Rough 包含：

$A\subseteq_R B\Leftrightarrow A\subseteq_R^* B\wedge A\subseteq_{*R}B\Leftrightarrow(\forall\mu_A^R(x)=1,\exists\mu_B^R(x)=1)\wedge(\forall\mu_A^R(x)\neq 0,\exists\mu_B^R(x)\neq 0)$

2.8 基于粒矩阵的 Rough 相等

定义 10 假定给定两个集合 $A,B\subseteq U$ 和 U 上的不可分辨关系 R，根据代数定义，定义集合 A,B 的 3 种相等：

1)集合 A 和 B 为 R 下 Rough 相等：

$A=_{*R}B\Leftrightarrow\forall(\mu_A^R(x)=1,\exists\mu_B^R(x)=1)\wedge(\forall\mu_B^R(x)=1,\exists\mu_A^R(x)=1)$

2) 集合 A 和 B 为 R 上 Rough 相等：

$A=_R^* B\Leftrightarrow(\forall\mu_A^R(x)\neq 0,\exists\mu_B^R(x)\neq 0)\wedge(\forall\mu_B^R(x)\neq 0,\exists\mu_A^R(x)\neq 0)$

3) 集合 A 和 B 为 R Rough 相等：

$$\begin{aligned}A=_R B\Leftrightarrow&(\forall\mu_A^R(x)=1,\exists\mu_B^R(x)=1)\\&\wedge(\forall\mu_B^R(x)=1,\exists\mu_A^R(x)=1)\\&\wedge(\forall\mu_A^R(x)\neq 0,\exists\mu_B^R(x)\neq 0)\\&\wedge(\forall\mu_B^R(x)\neq 0,\exists\mu_A^R(x)\neq 0)\end{aligned}$$

简言之，基于粒矩阵的 Rough 相等的充要条件是 $\mu_AR(x)$ 与 $\mu_BR(x)$ 要么同时为 1，要么同时不为 0。

2.9 基于隶属度函数的 Rough 包含与 Rough 相等

定义 11 对于 $\forall x\in U$ 利用 Rough 隶属度函数定义的 Rough 包含的充要条件为[11]

$X\subseteq_R Y\Leftrightarrow\mu_X^R(x)\leqslant\mu_Y^R(x)$

定义 12 对于 $\forall x \in U$，利用 Rough 隶属度函数定义 Rough 相等的充要条件为[11]：

$X =_R Y \Leftrightarrow \mu_X^R(x) = \mu_Y^R(x)$

粗糙集中用代数定义和隶属度函数定义的 Rough 包含与 Rough 相等是不等价的，其中基于隶属度函数的定义仅给出表达式而没有相关的证明[11]。从 2.7－2.9 可以看出，由两种不同定义方式义定义的 Rough 包含和 Rough 相等可以统一在基于粒矩阵的 Rough 隶属度函数之上。可以证明由 Rough 隶属度函数定义的 Rough 包含和 Rough 等价的充要条件仅仅是必要条件。由于篇幅关系，这里仅给出反例和基于粒计算的充要条件，数学证明将在其它文章中给出。

3 计算实例

3.1 Rough 集基本概念的计算

一个知识系统 $U=\{x_1,x_2,\cdots,x_8\}$，假设一个等价关系族 $P=\{R_1,R_2,R_3\}$，有下列等价类：

$U/R_1=\{\{x_1,x_3,x_4,x_5,x_6,x_7\},\{x_2,x_8\}\}$

$U/R_2\{\{x_1,x_3,x_4,x_5\},\{x_2,x_6,x_7,x_8\}\}$

$U/R_3=\{\{x_1,x_5,x_6\},\{x_2,x_7,x_8\}\{x_3,x_4\}\}$

等价关系 P 导出以下等价类：

$U/IND(P)=\{Y_1,Y_2,Y_3,Y_4,Y_5\}=\{\{x_1,x_5\},\{x_2,x_8\}\{x_3,x_4\},\{x_6\},\{x_7\}\}$

并设等价关系 Q 有下列等价类：

$U/Q=\{X_1,X_2,X_3,X_4\}=\{\{x_1,x_5,x_6\},\{x_2,x_7\}\{x_3,x_4\},\{x_8\}\}$

(1)求 R_*X 并计算 $cardR_*X$；

(2)求 R^*X 并计算 $cardR^*X$；

(3)计算根据 P,Q 的近似分类精度；

(4)计算根据 P,Q 的近似分类质量；

(5)根据我们定义的知识，计算知识 Q 对知识 P 的依赖性。

解：依据题意，求取 BGrM，可以得到：

$$Y=\begin{pmatrix}Y_{1\times 8}\\Y_{2\times 8}\\Y_{3\times 8}\\Y_{4\times 8}\\Y_{5\times 8}\end{pmatrix}=\begin{pmatrix}10001000\\01000001\\00110000\\00000100\\00000010\end{pmatrix}\quad X=\begin{pmatrix}X_{1\times 8}\\X_{2\times 8}\\X_{3\times 8}\\X_{4\times 8}\end{pmatrix}=\begin{pmatrix}10001100\\01000010\\00110000\\00000001\end{pmatrix}$$

$$C_{5\times 4}=Y\times X'=\begin{pmatrix}10001000\\01000001\\00110000\\00000100\\00000010\end{pmatrix}\times\begin{pmatrix}1000\\0100\\0010\\0010\\1000\\1000\\0100\\0001\end{pmatrix}=\begin{pmatrix}2&0&0&0\\0&1&0&1\\0&0&2&0\\1&0&0&0\\0&1&0&0\end{pmatrix}$$

$R_*X=\{\bigcup Y_i \mid NE(i)=1\}=\{Y_1\cup Y_3\cup Y_4\cup Y_5\}=\{x_1,x_3,x_4,x_5,x_6,x_7\}$

$card(R_*X)=\sum_{i=1}^{m}\sum_{j=1}^{n}\{c_{ij}\mid NE(i)=1)\}=c_{11}+c_{33}+c_{41}+c_{52}=6$

$R^*X=\{\bigcup Y_i \mid c_{ij}\neq 0\}=Y_1\cup Y_2\cup Y_3\cup Y_4\cup Y_5$

$=\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8\}$

$card(R^*X)=\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}=8$

根据知识 P,Q 的分类精度为

$$\alpha_R(F)=\frac{\sum card(R_*X_i)}{\sum card(R^*X_i)}=0.75$$

根据知识 P,Q 的近似分类质量为

$$\gamma_R(F)=\frac{\sum card(R_*X_i)}{card(U)}=0.75$$

知识 Q 对知识 P 的依赖性为

$k=\gamma_P(Q)=|POS_P(Q)|/U=0.75$

3.2 Rough 包含与 Rough 相等的例子

设论域 $U=\{x_1,\cdots,x_8\}$，R 是 U 上的等价关系，其等价类为 $E_1=\{x_2,x_3\}$，$E_2=\{x_1,x_4,x_5\}$，$E_3=\{x_6\}$，$E_4=\{x_7,x_8\}$。给定集合 $Z_1=\{x_2,x_3,x_4,x_5,x_7,x_8\}$，$Z_2=\{x_2,x_3,x_5,x_6,x_7,x_8\}$ 以及集合 $V_1=\{x_2,x_3,x_4,x_5,x_7\}$，$V_2=\{x_2,x_3,x_4,x_7\}$，由 BGrM 和式(25)可求得

$\mu_{Z_1}^R(x_2)=\mu_{Z_1}^R(x_3)=\mu_{Z_1}^R(E_1)=1$

$\mu_{Z_1}^R(x_1)=\mu_{Z_1}^R(x_4)=\mu_{Z_1}^R(x_5)=\mu_{Z_1}^R(E_2)=2/3$

$\mu_{Z_1}^R(x_6)=\mu_{Z_1}^R(E_3)=0$

$\mu_{Z_1}^R(x_7)=\mu_{Z_1}^R(x_8)=\mu_{Z_1}^R(E_4)=1$

$\mu_{Z_2}^R(x_2)=\mu_{Z_2}^R(x_3)=\mu_{Z_2}^R(E_1)=1$

$\mu_{Z_2}^R(x_1)=\mu_{Z_2}^R(x_4)=\mu_{Z_2}^R(x_5)=\mu_{Z_2}^R(E_2)=1/3$

$\mu_{Z_2}^R(x_6)=\mu_{Z_2}^R(E_3)=1$

$\mu_{Z_2}^R(x_7)=\mu_{Z_2}^R(x_8)=\mu_{Z_2}^R(E_4)=1$

显然，$Z_1\subseteq_R Z_2$，然而 $\mu_{Z_1}^R(x)-\mu_{Z_2}^R(x)$。

同理可验证 $V_1=_R V_2$，但是

$\mu_{V_1}^R(x_1)\neq\mu_{V_2}^R(x_1)\quad\mu_{V_1}^R(x_4)\neq\mu_{V_2}^R(x_4)\quad\mu_{V_1}^R(x_5)\neq\mu_{V_2}^R(x_5)$

可见 Rough 包含与 Rough 等价的隶属度函数定义的充要条件并不成立。

结束语 本文在二进制粒计算的基础上，定义了粒矩阵和粒矩阵的运算，建立了基于粒计算的 Rough 集模型，使得 Rough 集中的代数定义直观且便于理解；提供了求取边界的快速算法，为基于粒计算的知识发现算法和进一步的工程应用提供了数值计算方法。举例证明了 Rough 包含与 Rough 相等的隶属度函数定义的充要条件并不成立，并给出了修正的基于粒计算的 Rough 包含与 Rough 相等的充要条件。粒矩阵的定义为进一步完成粒的合成与分解，进行基于粒的知识推理与逻辑运算提供了数学工具。

参 考 文 献

[1] 王珏，袁小红，石纯一. 关于知识表示的讨论[J]. 计算机学报，1995，18(3)：212-224

[2] 苗夺谦，王珏. 粗糙集理论中知识粗糙性与信息熵关系的讨论[J]. 人工智能与模式识别，1998，11(01)：34-40

[3] 王国胤. Rough 集理论代数与信息论观点的关系研究[J]. 世界科技研究与发展，2002，24(5)：20-26

[4] 张文修，梁怡，吴伟志. 信息系统与知识发现[M]. 北京：科学出版社，2003

[5] 梁吉业，钱宇华. 粗糙集理论中的不确定性与知识粒度. 粗糙集与概念格[M]//张文修，姚一豫，梁怡，编. 西安：西安交通大学出版社，2006：113-135

(下转第 233 页)

义一致性检测后再抽取，R-Extra 表示在 C-Extra 基础上进行抽取修正。可以看出，融入推理后的抽取算法，性能明显得到提高，且其在查准率上的收益大于在查全率上的损失。在各抽取算法中，Sufferage 查准率收益最好，但查全率稍微差些，而 Hungarian 查全率损失最小。通过改变优化目标函数和算法的执行顺序，我们可以得出下面两个结论：

①当遇到 sufferage 值相等的情况较多时，修改优化的目标函数为连乘积后，查准率会略有提高；

②如果把不一致检测和结构冲突修正过程合并放到抽取之前进行，对于 Hungarian 算法，将直接影响抽取过程中的优化选取和指派，抽取质量提高的程度不如其他算法明显。而分离开来，则具有更大的适应性。

在时间复杂度上，Naïve-Desc 和 Sufferage 抽取算法均为 $O(n*m)$，其运行时间为 2～3s，而 Hungarian 为 5s 左右。可以预见，当本体规模变大时(Conference 测试集本体规模均较小)，Sufferage 的时间性能将比 Hungarian 更加好。

实验 2 把本文的映射算法应用在项目 SNAX 系统的升级版本中，使用该系统的映射子模块 SNAX_Mapping[10] 得出候选映射，再用本文的抽取算法进行抽取和修正。

表 2 SNAX_Mapping 映射抽取修正前后结果比较

System	SNAX_Mapping(抽取修正前)		SNAX_Mapping(抽取修正后)	
test	Pre.	Rec.	Pre.	Rec.
1××	1.0000	1.0000	1.0000	1.0000
2××	0.9367	0.8014	0.9445	0.7928
3××	0.8562	0.7301	0.8793	0.7274
total	0.9279	0.8326	0.9396	0.8298

表 2 中 #1××～3×× 表示标准测试数据集 benchmarks 中本体编号，Pre. 表示查准率，Rec. 表示查全率。抽取修正前的映射集的得出详见文献[10]。该实验设置参数 percentage 为 50%，σ 为 10%，从表 2 可以看出，在测试数据集[1××]上，抽取修正之前，查准率和查全率均为 1.0，抽取修正后仍然不变；对于测试数据集[2××]和[3××]，应用本文的算法后查准率都有了一定程度的提高，但查全率略有损失。其中，在[2××]上除本体 210，247，252，261，266 改进比较明显外，大部分没有太大改进，因此总体改进不大。但在[3××]上，除本体 301 外其他改进均较为明显，总体表现最佳。

通过对比各组数据的特点及实验观察，我们发现查准率的提高归结于两个方面：抽取之前的语义消歧及之后的冲突修正；全局最优的抽取算法及移除函数的制定。尤其是融入推理后，精度有了很大提高，分析如下：融入推理前，查准率公式为 $p=\frac{|R\cap A|}{|A|}$，融入推理后 A 可以分解为 A^+ 和 A^-，其中 A^- 是修正过程中去掉的“伪正确”映射对，此时查准率公式应改写为 $p'=\frac{|R\cap A^+|}{|A^+|}$。注意到 $|R\cap A^+|$ 和 $|R\cap A|$ 其值是相等的，因此最终的查准率公式可以写为 $p=\frac{|R\cap A|}{|A^+|}$。比较推理前后的两个公式便可以看出，使用推理技术能够更好地提高映射质量。

结束语 本文提出了 Sufferage 算法对候选映射对进行抽取。为了提高抽取质量，抽取之前先消除映射对间的语义冲突不一致，并对抽取之后的映射结果中可能存在的结构上的不一致进行修正。由于抽取之前的语义消歧和抽取之后的结构修正对抽取的质量影响程度不同，我们分两个阶段进行处理。尽管本文中推理是不完备的，然而实验发现只对映射对(及其涉及的某一局部范围)进行推理是很有效的，而且时间性能有了较大提高。

参 考 文 献

[1] Euzenat J, Shvaiko P. Ontology Matching [M]. Springer Verlag, 2007: 157-187

[2] 唐杰，梁邦勇，李涓子，等. 语义 Web 中的本体自动映射[J]. 计算机学报，2006，29(11)：1956-1976

[3] Meilicke C, Stuckenschmidt H. Applying logical constraints to ontology matching[C]// Proceedings of the 30th Annual German Conference on Artificial Intelligence. Germany: KI 2007, 2007: 99-113

[4] Meilicke C , Stuckenschmidt H . Analyzing mapping extraction approaches[C]// Proceedings of the ISWC'2007 Workshop on Ontology Matching OM-200. Korea, 2007: 25-36

[5] Meilicke C, Stuckenschmidt H, Tamilin A. Repairing ontology mappings[C]// Proceedings of the Twenty-Second Conference on Artificial Intelligence. Canada, 2007: 22-26

[6] Schlobach S. Debugging and semantic clarification by pinpointing [C]// Proceedings of ESWC 2005. Greece, 2005: 226-240

[7] 丁丁，罗四维，高瞻. 网格环境下一种可调目标的启发式调度策略[J]. 计算机研究与发展，2007，44(9)：1572-1578

[8] Kuhn H W. The hungarian method for the assignment problem [J]. Naval Research Logistics, 1955, 2: 83-97

[9] Svab O, Svatek V, Berka P, et al. Ontofarm: Towards an experimental collection of parallel ontologies[C]// Poster Proceedings of the International Semantic Web Conference 2005. Galway, 2005

[10] Zhang Zhiwei, Xu Dezhi, Zhang Tian. Ontology Mapping Based on Conditional Information Quantity[C]// Proceedings of IEEE International Conference on Networking, Sensing and Control . Sanya, 2008: 587-591

(上接第 202 页)

[6] Zadeh L A. Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems [J]. Soft Computing, 1998, 2 (1): 23-25

[7] Lin T Y. Granular Computing on Binary Relations: I: Data Mining and Neighborhood Systems. II: Rough Set Representations and Belief Functions[C]// Skowron A, Polkowski L, eds. Rough Sets in Knowledge Discovery. Physica-Verlag, 1998: 107-140

[8] Yao Y Y. Granular Computing: basic issues and possible solutions [A]// Proceedings of the 5th Joint Conference on Information Sciences[C]. Atlantic, USA: Association for Intelligent Machinery, 2000: 186-189

[9] 苗夺谦，王国胤，刘清，等. 粒计算：过去、现在与展望[M]. 北京：科学出版社，2007

[10] 刘清. Rough 集及 Rough 推理[M]. 北京：科学出版社，2001

[11] Pawlak Z. Rough Sets, Rough Relations and Rough Functions [J]. Fundam. Inform, 1996, 27(2/3): 103-108