# 动态粒度下的粗糙集近似*)

钱宇华 梁吉业 王 江
(山西大学计算机与信息技术学院 太原030006)

**摘 要** 粒度计算是粗糙集理论研究的一种强有力的工具。本文讨论了粒度意义下的粗糙集近似,并定义了动态粒度下的正向近似。另外,本文还从粒度的角度讨论了聚类结果和先验知识的协调度问题,并提出了一种基于动态粒度下的正向近似的聚类算法。这些结果将有助于粒度计算和粗糙集理论的研究。

**关键词** 动态粒度,粗糙集近似,聚类,协调度

## Rough Set Approximation under Dynamic Granulation

QIAN Yu-Hua LIANG Ji-Ye WANG Jiang
(School of Computer & Information Technology, Shanxi University, Taiyuan 030006)

**Abstract** Granular computing is emerging as a powerful tool for rough set theory. In this paper, we discuss rough set approximation under granulation, and establish positive approximation under dynamic granulation. In addition, measure of harmony between clustering and transcendent knowledge is defined, and a clustering arithmetic based on positive approximation under dynamic granulation is proposed. These results will be helpful for studying for granular computing and rough set theory.

**Keywords** Dynamic granulation, Rough set approximation, Clustering, Measure of harmony

## 1 引言

粒度计算覆盖了所有有关粒度的理论、方法论、技术和工具的研究[1~3]。粒度计算的基本思想在许多领域都有体现,例如区间分析、粗糙集理论、聚类分析、机器学习、数据库以及信息检索等领域[2,4]。Zadeh 依据模糊集理论提出了粒度计算的通用模型[2],用一般性约束来定义信息颗粒。约束的类型有相等、可能性、相似性、模糊性和功能性等约束,同时也提出了许多特定的粒度计算模型。Pawlak[5], Polkowski[6]和 Skowron[7]等都用粗糙集理论来解释和研究粒度计算。Lin[8]和 Y. Y. Yao[9,10]用近邻系统来研究粒度计算。Klir[11]则运用颗粒的相似性探讨了一些粒度计算的基本问题。

从粒度计算的角度来看,经典的粗糙集采用的是统一的粒度,即确定的等价关系。在经典的粗糙集理论中,Pawlak 认为用一对上、下近似集合可以粗糙地定义和描述论域中的任意一个子集[12]。在这种情况下,对对象的粗糙描述便产生了边界域,它可以定量地刻画其粗糙度。在以往的研究中,人们在描述和刻画问题时一般都是采用统一的粒度。但是,统一粒度有个最大的缺点:边界域不能够变化,不便于对它作进一步的分析研究。在实际应用中,我们往往需要从多个角度或者多个层次来分析问题、解决问题。也就是说,对同一个研究对象,需要用一个等价关系族来进行研究,而不是经典粗糙集理论里的单一等价关系。基于这种动态粒度原理,本文提出了动态粒度下的粗糙集近似,并给出了它的一些基本性质。另外,本文还从粒度的角度分析了聚类结果和先验知识的协调性问题,并给出了一种基于动态粒度原理的聚类算法。

## 2 粒度意义下的粗糙集近似

从知识粒度的角度看,在 Pawlak 的粗糙集理论中,知识库中的等价关系 $R\in\boldsymbol{R}$ 或信息系统中的属性集 $P\subseteq A$ 都对应着一个相应的粒度,或者说一个等价关系或一个属性集决定了一个相应的知识粒度。如果有偏序序列 $R_n\leq R_{n-1}\leq\cdots\leq R_1$ 存在,也就相对应着一个粒度从细到粗的序列。也就是说,一个具体的等价关系对应着一个具体的粒度。从这个角度来分析,便可以给出粒度意义下的粗糙集近似。

### 2.1 统一粒度

粒度意义下,经典的粗糙集是一种统一粒度意义下的粗糙描述。

**定义1**[12] 给定知识库 $K=(U,R)$,对于每个子集 $X\subseteq U$ 和一个等价关系 $R\in\boldsymbol{R}$,定义两个子集:

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

$$\overline{R}X=\bigcup\{Y\in U/R\mid Y\cap X\neq\varnothing\}。$$

分别称它们为下近似集和上近似集。

下近似、上近似也可用下面的等式表达:

$$\underline{R}X=\{x\in U\mid[x]_R\subseteq X\},$$

$$\overline{R}X=\{x\in U\mid[x]_R\cap X\neq\varnothing\},$$

集合 $bn_R(X)=\overline{R}X-\underline{R}X$ 称为 $X$ 的 $R$ 边界域;$pos_R(X)=\underline{R}X$ 称为 $X$ 的 $R$ 正域;$neg_R(X)=U-\overline{R}X$ 称为 $X$ 的 $R$ 负域。显然:$\overline{R}X=pos_R(X)\cup bn_R(X)$。

### 2.2 动态粒度

动态粒度原理有两种情况:(1)对同一个问题采用不同的角度(粒度)来研究;(2)对同一个问题采用多个层次(粒度)来

*)本文受到国家自然科学基金(60275019)、山西省自然科学基金(20031036)和山西省留学基金资助。钱宇华 硕士研究生,主要研究方向为数据挖掘,粒度计算。梁吉业 教授,博士,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。王 江 讲师,主要研究方向为数据挖掘。

研究。在第二种情况里，所采用的多个粒度之间要具有一定的偏序关系，而前者则不作要求。本文运用第二种思想来进行研究。

在粒度变化过程中，有逐渐细化和逐渐粗糙两种情况。前者主要处理对研究对象刻画和描述过于粗糙，仍需作进一步更精细的刻画的情况；而后者则相反，是处理目前所进行的刻画和描述过于精细，丢失了一些对象的抱团性质，需要使之粗糙一些的情形。

用一组具有偏序关系的等价关系族 $R_n \preceq R_{n-1} \preceq \cdots \preceq R_1$ 来刻画对象 $X$ 的思想称为动态粒度下的粗糙集近似。如果采用的是逐渐细化的思想，称为正向近似；如果采用的是逐渐粗糙的思想，称为逆向近似。在实际应用中，常常关注的是粗糙集的边界域，所以本文所研究的动态粒度下的粗糙集近似每次所刻画的对象都是其新的边界域。在这里我们只讨论正向近似。

2.2.1 正向近似

**定义2** 给定知识库 $K=(U,R)$，对于每个子集 $X\subseteq U$ 和一个具有偏序关系的等价关系族

$P=\{R_n,R_{n-1},\cdots,R_1\}$

且满足 $R_n\preceq R_{n-1}\preceq\cdots\preceq R_1(R_i\in R)$，初始关系为 $R_1$，定义两个子集：

$$\overline{P}X=\overline{R_n}X,$$

$$\underline{P}X=\bigcup_{i=1}^{n}\underline{R_i}X_i$$

其中 $X_1=X, X_i=X-\bigcup_{k=1}^{i-1}\underline{R_k}X_k$，分别称它们为 $X$ 的 $P$ 上近似集和 $P$ 下近似集。这里的 $X_i$ 采用的是递归的定义。

$P$ 上近似集、$P$ 下近似集也可用下面的等式表达：

$$\overline{P}X=\{x\in U\mid [x]_{R_n}\cap X\neq\varnothing\}$$

$$\underline{P}X=\{x\in U\mid [x]_{R_i}\subseteq X_i,$$

其中 $X_1=X, X_i=X-\bigcup_{k=1}^{i-1}\underline{R_k}X_k$。

集合 $bn_P(X)=\overline{P}X-\underline{P}X$ 称为 $X$ 的 $P$ 边界域；$pos_P(X)=\underline{P}X$ 称为 $X$ 的 $P$ 正域；$neg_P(X)=U-\overline{P}X$ 称为 $X$ 的 $P$ 负域。显然：$\overline{P}X=pos_P(X)\cup bn_P(X)$。

正向近似过程中，把 $X$ 中仍需要进一步研究的领域作为下一个研究对象，这样便形成了一系列不同层次的表达式。正向近似使得子集 $X\subseteq U$ 在等价关系族 $P$ 下使用知识颗粒最少且表达最为精确。

当然，有的时候人们所关注的对象也许不是边界域，但是根据上文的动态粒度思想，也可以得出相应的动态粒度表示。

## 3 粒度意义下的聚类协调性

聚类操作实质上是在样本点之间定义一种等价关系。属于同一类的任意两个样本点被看作是等价的，可以认为它们具有相近的性质，在当前的阈值尺度下是没有区别的。一个等价关系就定义了样本点集合的一个划分，它把样本点划分成一些子集，一个子集就对应着聚类形成的一个类。由一族粗细不等的等价关系便形成了不同的聚类结果，这些等价关系形成一个偏序格结构[13]。

从聚类的角度来看，先验知识中规定的某一类中的样本点，依照选定的特征空间和相似性测度，也应当聚成一类。然而不幸的是，在决大多数情况下，都不会达到这种理想境界，常常遇到的情况是领域专家认为应该归为一类的点，往往在特征空间中距离特别远；而那些被认为分属不同类的点，却距离非常近。也就是说，聚类结果和先验知识之间往往存在某种不协调性。

在这里我们借用粗糙集理论和粒度计算的体系和术语来描述聚类结果和先验知识的不协调性。

3.1 统一粒度

聚类谱系图实际上是定义了一个粒度逐渐变细的等价关系序列，选择一个阈值实际上就是选定一个等价关系 $R$，也就是选定了一个粒度 $G(R)$，进而可以得到商集，即知识体系 $U/R$。如果由先验知识规定的类 $X$ 能够使用现有的信息颗粒精确表达，就表示聚类结果和先验知识是协调的；而如果上近似和下近似不相同的话，就说明聚类结果和先验知识是不协调的。

**定义3** 给定知识库 $K=(U,R)$，对$\forall$ $X\subseteq U$，给定 $R\in R$，定义由等价关系 $R$ 所形成的聚类结果和 $X$ 的协调度为：

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

其中 $|\cdot|$ 表示集合的基数。

显然，协调度 $H(R,X)\in[0,1]$。当 $H(R,X)=0$ 时，表示聚类结果和先验知识最不协调；当 $H(R,X)=1$ 时，表示聚类结果和先验知识最协调，即现有的知识完全可以精确地描述先验知识。

3.2 动态粒度

然而在实际情况下，我们对粗糙的边界中的对象常常是感兴趣的，或者为了更加精确地描述先验知识，总之在很多时候是需要进一步研究边界对象的。在这种情况下，需要使粒度进一步细化，即采取更加细化的等价关系，以此类推，直到满足实际需求或者表达尽可能的精细为止。

聚类谱系图定义了一个粒度逐渐变细的等价关系序列，选择一组阈值实际上就是选定一个等价关系族 $P=\{R_n,R_{n-1},\cdots,R_1\}$，其满足偏序关系 $R_n\preceq R_{n-1}\preceq\cdots\preceq R_1$，进而可以得到相应的商集序列，即知识体系族 $U/R_i(i=1,2,\cdots,k)$。先验知识规定的类 $X$ 可先由知识体系 $U/R_1$ 来表达，令其边界中信息颗粒集作为新的研究对象，再用知识体系 $U/R_2$ 来表达，依次类推，直至满足实际需求或者达到目前知识体系所能表达的最大精细程度时为止。这种思想就是定义2中定义的正向近似。

**定义4** 给定知识库 $K=(U,R)$，对$\forall$ $X\subseteq U$，给定一个具有偏序关系的等价关系族

$P=\{R_n,R_{n-1},\cdots,R_1\}$

且满足 $R_n\preceq R_{n-1}\preceq\cdots\preceq R_1(R_i\in R)$，定义由 $P$ 的正向近似所形成的聚类结果和 $X$ 的协调度为：

$$H(P,X)=\frac{|\underline{P}X|}{|X|}$$

其中 $|\cdot|$ 表示集合的基数。

显然，协调度 $H(P,X)\in[0,1]$。当 $H(P,X)=0$ 时，表示聚类结果和先验知识最不协调；当 $H(P,X)=1$ 时，表示聚类结果和先验知识最协调，即现有的知识完全可以精确地描述先验知识。

在实际应用中，使用给定的等价关系族 $P=\{R_n,R_{n-1},\cdots,R_1\}$ 进行正向近似也许不需要进行完毕，可能在 $R_c(c=\{1,2,\cdots,n\})$ 处便已能够精确地表示和描述先验知识 $X$。但是这并不影响协调度的计算，它的值在进一步细化中一直保持为 $H(P,X)=1$ 不变。

采用正向近似方案的结果，就是把先验知识 $X$ 分解成一

些处在不同粒度层次子类的并集 $X=X_1\cup X_2\cup\cdots\cup X_k$，表示第 $k$ 次已经能够最为精确地表达先验类 $X$，每一个子集都采用不同的粒度，是在相应粒度之下能够精细表达的最大子集。在信息系统中，经过属性约简后，每一个属性的增减都会影响粒度的粗细。从这个角度来考虑，聚类结果和先验知识的协调度也可以用∪的个数来定量地反映。

当然我们也可以用逆向近似方案来近似表达先验类 $X$，但是这种思想在挖掘对象抱团性质的同时会相应降低聚类结果和先验知识之间的协调度。

## 4 一种基于正向近似的聚类算法

在实际应用中，先验知识 $X$ 的刻画和描述有无指导和有指导之分。在信息系统中，有指导的聚类指的是事先给定一个必须的属性集(这些属性是领域专家指定的或者描述问题所必须的)，然后再在此基础上加入新的属性来继续聚类。而无指导的聚类则是不需要任何先验指导的聚类。

给定信息系统 $S=(U,A)$，对$\forall X\subseteq U$，在这里给出 $P$ 的构造算法，即构造一个具有偏序关系的等价关系族 $P=\{P_n,P_{n-1},\cdots,P_1\}$，其满足 $P_n\preceq P_{n-1}\preceq\cdots\preceq P_1(P_i\subseteq A)$，同时也给出了基于正向近似的聚类算法。

算法1：$P$ 的构造算法

在信息系统中，经过属性约简后，每一个属性的增减都会影响粒度的粗细及其聚类结果。

1) 信息系统 $S=(U,A)$，经属性约简和计算核属性集[12]后属性集 $A_0=\{a_1,a_2,\cdots,a_m\}$和 $core(A)=\{a_1',a_2',\cdots,a_n'\}$；

2) 如果是有指导的聚类：那么指定初始属性集 $P_1=\{a_1^*,a_2^*,\cdots,a_l^*\}$；

如果是无指导的聚类：那么令初始属性集 $P_1=core(A)=\{a_1',a_2',\cdots,a_l'\}$，此时令 $s=l$；

3)$i=2$；

4)计算剩下的属性集 $A_0-P_{i-1}$中各个属性对 $P_{i-1}$的重要度[14]，找出其中重要度最大的属性，令其和 $P_{i-1}$一块构成 $P_i$，$i=i+1$；

5)如果 $i=m-s+2$则转到6)；否则转至4)；

6)这样形成的序列 $P_1,P_2,\cdots,P_{m-s+1}$便构成了一个由粗到细且具有偏序关系的等价关系族 $P$，算法结束。

算法2：正向近似聚类算法

1)用算法1求出具有偏序关系的等价关系族 $P=\{P_n,P_{n-1},\cdots,P_1\}(P_i\subseteq A)$；

2)构造一族等价关系族 $\boldsymbol{P}=\{P^1,P^2,\cdots,P^n\}$，其中 $P^i=\{P_i,P_{i-1},\cdots,P_1\}$，且满足偏序关系 $P_i\preceq P_{i-1}\preceq\cdots\preceq P_1(i=\{1,2,\cdots,n\})$；

3)$j=1$；

4)计算 $P^j$ 的正向近似的下近似$\underline{P^j}X$；

5)计算此时的聚类结果和 $X$ 的协调度为 $H(P^j,X)=\dfrac{|\underline{P^j}X|}{|X|}$；

6)如果 $H(P^j,X)=1$或者 $j>n$ 成立，那么聚类结果 $I(X)=\underline{P^j}X$；否则 $j=j+1$，转至4)；

7)最后用来近似描述先验知识 $X$ 的聚类结果为 $I(X)$，算法结束。

**例1** 一个关于某些病人的知识表达系统如表1。

表1

| 病人 | 头痛 | 肌肉痛 | 体温 |
|---|---|---|---|
| $e_1$ | 是 | 是 | 正常 |
| $e_2$ | 是 | 是 | 高 |
| $e_3$ | 是 | 是 | 很高 |
| $e_4$ | 否 | 是 | 正常 |
| $e_5$ | 否 | 否 | 高 |
| $e_6$ | 否 | 是 | 很高 |

$U=\{e_1,e_2,e_3,e_4,e_5,e_6\}$，

$A=\{a_1,a_2,a_3\}=\{$体温，头痛，肌肉痛$\}$，

$X=\{e_2,e_4,e_6\}$。

在这里进行有指导的聚类，令初始属性集为 $P_1=\{a_1\}$，剩下的属性集 $A-P_1=\{a_2,a_3\}$，现在分别计算它们对 $P_1$的重要度。其计算公式[14]为 $Sig_P(a)=1-|P\cup\{a\}|/|P|$，设 $P$ 的划分 $U/P=\{X_1,X_2,\cdots,X_n\}$，$|P|=\sum_{i=1}^{n}|X_i|^2$。

所以，$Sig_{P_1}(a_2)=1-|P_1\cup\{a_2\}|/|P_1|=1-6/12=1/2$

$Sig_{P_1}(a_3)=1-|P_1\cup\{a_3\}|/|P_1|=1-5/6=1/6$

这里有 $Sig_{P_1}(a_2)>Sig_{P_1}(a_3)$。把 $a_2$和 $P_1$合并为 $P_2=\{a_1,a_2\}$。

同理，$P_3=\{a_1,a_2,a_3\}$。

这样便构造了等价关系族 $P=\{P_1,P_2,P_3\}$，且满足偏序关系 $P_1\preceq P_2\preceq P_3$。

有了偏序关系序列，便可以根据算法2动态地进行聚类。

由算法2可以得出：

$P^1=\{P_1\}$；$P^2=\{P_1,P_2\}$；

$P^3=\{P_1,P_2,P_3\}$。

$\underline{P^1}X=\underline{P_1}X=\{\varnothing\}$，

$H(P^1,X)=0/3=0$；

$\underline{P^2}X=\{e_2,e_4,e_6\}\cup\{\varnothing\}=\{e_2,e_4,e_6\}$；

$H(P^2,X)=3/3=1$。

此时，聚类协调度为1，即用来近似描述先验知识 $X$ 的聚类结果为$\underline{P^2}X=\{e_2,e_4,e_6\}$。由于已经达到精确描述，故不需要进一步考虑$\underline{P^3}X$。

**结论** 本文讨论了粒度意义下的粗糙集近似，并定义了动态粒度下的正向近似，为粒度计算和粗糙集理论提供了新的研究角度。另外，本文还从粒度的角度讨论了聚类结果和先验知识的协调度问题，并提出了一种基于动态粒度下的正向近似的聚类算法。但是，本文提出的动态粒度下的粗糙集近似还应有一些相应的性质需要进一步研究和探讨，文中给出的聚类算法的算法效率也可以作进一步的改进。

## 参 考 文 献


1 Zadeh LA. Fuzzy logic-computing with words. IEEE Transactions on Fuzzy Systems, 1996, 4(1): 103～111

2 Zadeh LA. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy sets and Systems, 1997, 19(1): 111～127

3 Zadeh LA. Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems. Soft Computing, 1998, 2(1): 23～25

4 Zadeh LA. Fuzzy sets and information granularity. In: Gupta N, Ragade R, Yager R, eds. Advances in fuzzy set theory and applica-

tion. Amsterdam: North-Holland, 1979. 3～18
5 Pawlak Z. Granularity of knowledge, indiscernibility and rough sets. Proceedings of 1998 IEEE Intl. Conf. on Fuzzy Systems, 1998. 106～110
6 Polkowski L, Skowron A. Towards adaptive calculus of granules. In: Proc. of 1998 IEEE Intl. Conf. on Fuzzy Systems, 1998. 111～116
7 Skowron A, Stepaniuk J J. Information granules and approximation spaces. Manuscript, 1998
8 Lin T Y. Granular computing on binary relations I: data mining and neighborhood systems, II: Rough set representations and belief functions. In: Polkowski L, Skowron A, editors. Rough sets in knowledge discovery 1. Heidelberg: Physica-Verlag, 1998. 107～140
9 Yao Y Y. Granular computing using neighborhood systems. In: Roy R, Furuhashi T, Chawdhry, PK, eds. Advances in soft computing: engineering design and manufacturing. London: Springer-Verlag, 1999. 539～553
10 Yao Y Y. Rough sets, neighborhood systems, and granular computing. In: Proc. of the 18th Intl. Conf. of the North American Fuzzy Information Processing Society. IEEE Press, 1999. 800～804
11 Klir GJ. Basic issues of computing with granular computing. In: Proc. of 1998 IEEE Intl. Conf. on Fuzzy Systems; 1998. 101～105
12 张文修，吴伟志，梁吉业，李德玉. 粗糙集理论与方法. 北京：科学出版社，2001
13 卜东波，白硕，李国杰. 聚类/分类中的粒度原理. 计算机学报，2002，25(8)：800～806
14 苗夺谦，范世栋. 知识的粒度计算及其应用. 系统工程理论与实践，2002，1：48～56

(上接第174页)

6 Feng Y, et al. A Handbook of TCM Prescriptions. Guizhou Science and Technology Publishing House, 1999
7 Gu F, Cao C G. Biological Knowledge Acquisition from the Electronic Encyclopedia of China. In: Proc. of the 16th Intl. Conf. for Young Computer Scientists, 2001. 1199～1203
8 Gomez F. Acquiring Knowledge about the Habitats of Animals from Encyclopedic Texts. In: Proc. of the Workshop for Knowledge Acquisition, 1995. 1～22
9 Gomez F, Hull R, Segami C. Acquiring Knowledge from Encyclopedic Texts. In: Proc. of the 4th Conf. on Applied Natural Language Processing, 1994. 84～90
10 Hahn U, Romacker M, Schulz S. Discourse Structures in Medical Reports - Watch Out! The Generation of Referentially Coherent and Valid Text Knowledge Bases in the MEDSYNDIKATE System. International Journal of Medical Informatics, 1999, 53: 1～28
11 Hahn U, Schnattinger K. Deep Knowledge Discovery from Natural Language Texts. In: D. Heckerman, H. Mannila, D. Pregibon and R. Uthurusamy, eds. Proc. of the 3rd Intl. Conf. on Knowledge Discovery and Data Mining, Menlo Park, CA: AAAI Press, 1999. 175～178
12 Hahn U, Klenner M, Schnattinger K. Learning from Texts: A Terminological Metareasoning Perspective. In: S. Wermter, E. Riloff and G. Scheler, eds. Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing. Lecture Notes In Artificial Intelligence 1040, Springer-Verlag, Berlin, 1996. 453～468
13 Hahn U, Romacker M. Content Management in the SYNDIKATE System - How Technical Documents Are Automatically Transformed to Text Knowledge Bases. IEEE Transactions on Data & Knowledge Engineering, 2000, 35: 137～159
14 He X D. The Chinese Materia Medica (China Academy Press, 1998)
15 Hull R, Gomez F. Automatic Acquisition of Biographic Knowledge from Encyclopedic Texts. International Journal of Expert Systems with Applications, 1999, 16: 261～270
16 Kazawa K, Fujimoto K, Matsuzawa K. Attribute Dependency Acquisition from Formatted Text. In: Proc. of the 3rd Intl. Conf. on Knowledge-Based Intelligent Information Engineering Systems, 1999. 464～468
17 Lapalut S. Text Clustering to Help Knowledge Acquisition from Documents. In: N. Shadbolt, K. Ohara and G. Schreiber, eds. Advances In Knowledge Acquisition. Lecture Notes in Artificial Intelligence 1076, Springer-Verlag, Berlin, 1996. 115～130
18 Lei Y X, Cao C G. Acquiring Military Knowledge from the Encyclopedia of China. In: Proc. of the Sixth Intl. Conf. for Young Computer Scientists, Hangzhou, 2001. 368～372
19 Li Q Y. Formulae of Traditional Chinese Medicine. China Academy Press, 1998
20 Plant R T. Techniques for Knowledge Acquisition from Text. International Journal of Computer Information Systems, 1994, 35: 64～70
21 Lo H R, et al. Colored Icones of Chinese Medicine. Guangdong Science and Technology Press, 1991, 1-5
22 Lu R Q, Cao C G. Towards Knowledge Acquisition from Domain Books. In: B. Wielinga, B. Gaines, G. Schreiber and M. Vansomeren, eds. Current Trends in Knowledge Acquisition, Amsterdam, IOS Press, 1990. 289～301
23 Lu R Q, Cao C G, Chen Y H, et al. A PNLU Approach to ICAI System Generation. Science In China (Series A), 1996, 38: 1～10
24 National Committee of Codex. Codex of China (Volume 1). Chemical Engineering Press, 2000
25 Richardson S. Determining Similarity and Inferring Relations in a Lexical Knowledge Base: [PhD. Dissertation]. City University of New York, 1997
26 Richardson S, Dolan W B, Vanderwende L. MindNet: Acquiring and Structuring Semantic Information from Text. In: Proc. of the 36th Annual Meeting of the Association for Computational Linguistics and 17th Intl. Conf. on Computational Linguistics, 1998. 1098～1102
27 Sato H, Fujimoto K. A New Approach to Semantic Word-Matching for Knowledge Acquisition from Text Containing Daily-used Words. In: M. Mohammadian, ed. Advances In Intelligent Systems: Theory And Applications, 2000, 59: 135～140
28 Schmidt G. Knowledge Acquisition from Text in a Complex Domain. In: F. Belliand F. J. Radermacher, eds. The 5th Intl. Conf. on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems. Lecture Notes in Artificial Intelligence, Springer-Verlag, Berlin, 1992, 604: 529～538
29 Shi D B, et al. The Medical Volume of the Encyclopedia of China. Publishing House of Encyclopedia of China, 1991
30 Tian W, Cao C G. A Framework for Extracting Knowledge of the Human Blood System from Medical Texts. In: Proc. of the 16th Intl. Conf. for Young Computer Scientists, Hangzhou, 2001. 501～505
31 Tschaitschian B, Abecker A, Schmalhofer F. Information Tuning With KARAT: Capitalizing on Existing Documents. In Lecture Notes in Artificial Intelligence, Springer-Verlag, Berlin, 1997, 1319: 269～284
32 Lu R Q, Cao C G. Towards Knowledge Acquisition from Domain Books. In: B. Wielinga, B. Gaines, G. Schreiber and M. Vansomeren, eds. Current Trends in Knowledge Acquisition (Amsterdam, IOS Press, 1990) 289～301. N. Guarino, Formal Ontology and Information Systems, Proc. of the First Intl. Conf. (FOIS'98), June, Trento, Italy