维普资讯 http://www.cqvip.com



# 基于知识本体的属性分析以及概念联通*)

雷玉霞[1] 眭跃飞[2] 曹宝香[1]

(曲阜师范大学计算机学院 山东曲阜273165)[1] (中国科学院计算技术研究所 北京100080)[2]

**摘 要** 知识库是计算机智能的重要来源。要提高机器的智能程度,不仅需要大型的知识库,更需要知识之间的广泛联通。类似于MindNet的语义联通,我们以文本知识为背景提出了基于本体的概念联通。文章首先介绍了知识本体的设计思想,然后进行属性的本体分析,最后讨论了基于属性的概念联通以及A-属性联通测度。

**关键词** 知识表示,概念本体模型,概念联通,A-属性联通,联通测度

## Based-Ontology Attribute Analysis and Concept Interconnection

LEI Yu-Xia[1] SUI Yue-Fei[2] CAO Bao-Xiang[1]

(Department of Computer Science, Qufu Teachers University, Qufu 273165)[1]

(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080)[2]

**Abstract** Knowledge base is an important source of intelligence of computer. It needs both a large knowledge base and more interconnection among knowledge for computer in order to improve the degree of intelligence. In this paper, we present concept-interconnection being similar to semantic connection of MindNet. We firstly introduce knowledge-ontology, and then discuss the attribute analysis, and at last discuss based-attribute concept-interconnection and the degree of A-attribute interconnection.

**Keywords** Knowledge representation, Concept-ontology model, Concept-interconnection, The degree of interconnection, A-interconnection

## 1 引言

知识库是计算机实现智能的重要基础,是计算机智能的重要来源[1]。即使建立了大型知识库,而不把有关联的知识联系起来,也只能说为计算机堆积了大量的"知识",形成一些"概念孤岛",所以有必要把有关的知识联系起来。从知识工程角度看,要提高机器的智能程度更需要知识间广泛的联系[2~5]。

在国外,Richardson等创建的MindNet涉及到词汇之间的24类语义联系,可以说是词汇联通。MindNet是从两种机器可读的词典(MRD)中自动构建起来的词汇知识库,代表了从自然语言文本中获取、组织、访问、开采语义信息的一般方法等,充分约束的、扩展的语义关系路径(Semantic Relation Paths)已经证明,词跟词之间的关系不被连接起来就没有意义[6,7]。

为了解决专业知识库中"概念孤岛"的问题,中科院计算所的曹存根研究员首先提出了知识联通的概念。类似于MindNet的词汇联通,我们以文本知识为背景提出了概念联通。所谓概念联通就是指通过关系链或属性或属性值将两个概念联系起来的过程。关系链是指形如"$C_1,R_1,C_2,R_2,\cdots,R_{n-1},C_n$"的序列,其中$C_1,\cdots,C_n$表示概念,$R_1,\cdots,R_{n-1}$表示关系。本文重点讨论基于属性或属性值的概念联通。为了讨论概念联通,我们首先对领域知识进行了本体论分析(Ontological Analysis)。本体论分析已经成了国际上知识处理研究的一个基本手段。它从一个很高的层次刻画各种不同领域的知识,因而它所生成的一个本体具有很高的跨领域的可重用性[8]。为了提高知识的可重用性,在知识获取过程中,我们建立了不同类型知识本体。然后,在建立本体的基础上进行知识分析。最后,在知识分析的基础上讨论概念联通。

本文第2部分介绍知识本体模型(SOM和COM)的设计思想;第3部分讨论基于取值域类型的属性划分,并给出一些重要的定理;第4部分讨论基于属性或属性值的概念联通以及A-联通测度;最后总结全文并提出一些值得讨论的问题。

## 2 知识本体

我们的知识本体分为两个层次:一是领域本体;二是概念本体。领域本体描述某个领域中的概念结构以及概念系统。概念本体主要包括两个部分:属性部分和关系部分。属性集部分主要描述概念的内涵。另外,还包括一些辅助属性,关系集合表示与其他概念的关联。在本体设计中,我们充分了考虑完备性、准确性、可重用性、简洁性和可扩展性等重要原则[9]。

### 2.1 领域本体模型(Special Ontology Model)

领域本体是描述特定领域中的概念结构和概念之间关系的。下面首先给出3个定义,然后给出一些基本性质。

**定义2.1**(领域本体模型) 一个领域本体模型$SOM=(C,R,A,\top,H,O,F)$。其中,$C$表示领域中的所有概念的集合;$R$表示所有关系的集合;$A$表示所有属性的集合;$\top$不在$C$中,但在分类$H$中。$\top$表示所有实体的概念,是$C$中所有概念的上位概念;$H=(C,\leq)$表示$C$中概念的分类结构,其中$\leq$是一个偏序;$O=(O,\lesssim)$表示所有领域概念本体集合$O$上的结构,其中$\lesssim$是个偏序;$F$是个映射,且是连续的,即$F: H\to O$。另外,满足下面条件:

(2.1)$\forall (O\in O)(O'\in O)(O\lesssim O'\to O\subseteq O')$;

(2.2)对每个元素$O\in O, O=\bigcup_{O'\lesssim O}O'$

*)本文得到国家自然科学基金(项目号:60072014)资助。**雷玉霞** 硕士,感兴趣的方向:人工智能(尤其是知识获取与知识分析和KDD),算法设计与分析;**眭跃飞** 博士导师,研究员;**曹宝香** 教授,硕士生导师,感兴趣的方向:图形图像,人工智能。

维普资讯 http://www.cqvip.com

(2.3)∀ (C∈$H$)(C'∈$H$)(C≤C'→$F$(C)≤$F$(C'))；

(2.4)∀ (O∈$O$)∃ (C∈$H$)($F$(C)=O))；

(2.5)对每个 C∈$H$,$F$(C)=$\bigcup_{C'\leq C}F(C')$

**定义2.2**(极小上界) 称概念 C 是概念 $C_1$和 $C_2$的极小上界，如果 $C_1$和 $C_2$同时满足：$C_1\leq C$,$C_2\leq C$;∀ D($C_1\leq D\wedge C_2\leq D\rightarrow C\leq D$)。记作 C=lub($C_1$,$C_2$)。

**定义2.3**(极大下界) 称概念 C 是 $C_1$和 $C_2$的极大下界，如果 $C_1$和 $C_2$同时满足：$C\leq C_1$,$C\leq C_2$;∀ D($D\leq C1\wedge D\leq C_2\rightarrow D\leq C$)，记作 C=glb($C_1$,$C_2$)。

**命题2.1** 若 C=lub($C_1$,$C_2$),D=glb($C_1$,$C_2$),则有 $F$(D)⊆$F$($C_1$)∪$F$($C_2$)⊆$F$(C)。

证明：(1)由 C=lub($C_1$,$C_2$)可得 $C_1\leq C$,$C_2\leq C$,又由(2.3)可知 $F(C_1)\leq F(C)$,$F(C_2)\leq F(C)$,再由(2.1)知 F($C_1$)⊆F(C),F($C_2$)⊆F(C),所以有 F($C_1$)∪F($C_2$)⊆F(C)；

(2)显然有 F($D$)⊆F($C_1$)∪F($C_2$)成立。

**命题2.2** 若 $O\leq O'$,$O'\leq O$,则 $O=O'$。

证明：由题意及(2.1)可知：$O\subseteq O'$,$O'\subseteq O$,所以有 $O=O'$。

### 2.2 概念本体模型(Concept-Ontology Model)

概念本体是描述一个概念的数学模型。主要描述该概念的个体与属性之间的关系，该概念和其他概念之间的关系。首先，我们先给出概念本体模型的定义，然后对构件进行解释。

**定义2.4**(概念本体模型) 一个概念本体模型是个五元组，即 COM=($A$(C),$I$(C),$R$(C),$P$(C),$H$)。概念 C 的本体模型记作 $O$(C)。另外，满足条件：

(2.6)如果 C≤C'，则 $O$(C)⊆O(C')；

(2.7)如果 C∈$C$，则 $O(C)\supseteq\bigcup_{C',P(C',C)}O(C')$，其中 $P$(C',C)表示 C' 是 C 的物理部分；

(2.8)如果 C 是 C' 的一个物质，则 $O$(C)⊆$O$(C')

其中，$A$(C)⊆$A$,$R$(C)⊆$R$,$I$(C)表示概念 C 的所有个体的集合，$P$(C)表示侧面集合，$H\subseteq I(C)\times A(C)$是一个个体和属性之间的二元关系，且每个序偶(i,a)∈$H$ 表示个体 i 包含属性 a；$A_x$(C)表示与 $O$(C)有关的公理集。

**命题2.3** 若 C=lub($C_1$,$C_2$),D=glb($C_1$,$C_2$),则有 $O$(D)⊆$O$($C_1$)∪$O$($C_2$)⊆$O$(C)。

**定义2.5** 设 COM=($A$(C),$I$(C),$R$(C),$P$(C),$H$),对 X⊆$I$(C),Y⊆$A$(C),定义映射如下：

φ(X)={a∈$A$(C)|∀ i∈X,(i,a)∈$H$}

σ(Y)={i∈$I$(C)|∀ a∈Y,(i,a)∈$H$}

给定一个概念本体模型 COM。在 COM 中，集合 φ(X)表示集合 X 中所有个体的共有属性，而集合 σ(Y)则表示含有属性 Y 的所有个体的集合。

**命题2.4** 给定 COM=($A$(C),$I$(C),$R$(C),$P$(C),$H$)。映射 φ 和 σ 具有以下性质，其中 X,$X_1$,$X_2$∈$I$(C),Y,$Y_1$,$Y_2$∈$A$(C)：

(2.9)$X_1\subseteq X_2\Rightarrow\varphi(X_1)\subseteq\varphi(X_2)$；

(2.10)$Y_1\subseteq Y_2\Rightarrow\sigma(Y_1)\subseteq\sigma(Y_2)$；

(2.11)$X\subseteq\sigma(\varphi(X))$

(2.12)$Y\subseteq\varphi(\sigma(X))$

概念本体模型 COM=($A$(C),$I$(C),$R$(C),$P$(C),$H$)的结构描述一般为：

COM〈概念本体名〉[:〈上位概念本体〉]

{ Attibute Set〈对概念的属性描述〉

Relation Set〈与其他概念的关联描述〉

Restaint〈限制条件〉

$H\subseteq I(C)\times A(C)$〈个体和属性之间的二元关系〉

}

{

$A_x$〈对应于 COM 的公理集合〉

}

## 3 基于取值域类型的属性划分

对于属性有多种不同的划分方法。例如根据可能世界可划分成严格属性、非严格属性和反严格属性等[9,10]。为了讨论 A-属性联通以及联通测度，我们给出一种新的划分方法。

每个属性都有相应的取值域。下面，我们根据取值域的情况将属性分为三大类：结构属性、分类型属性和非结构属性。其中，结构属性又分为线性属性和序属性。先看几个属性。例如速度是舰艇和飞机的一个属性，其属性值是线性的。再如职业是人的一个属性，其取值域是一个分类结构。例如一个人可能是一个领域专家，是一个军事专家，是一个核武器专家。很显然，速度和职业是有区别的。设 A 表示属性，其取值域为 $D_A$。

**定义3.1**(结构属性) 称 A 是结构属性，若 $D_A$ 上有个偏序≤。

**定义3.2** 如果 A 是结构属性，则

(3.1)如果 $D_A$ 上的偏序≤是线性的，即对于 $D_A$ 中的任何两个元素 X 和 Y，有 X≤Y，或者 Y≤X，则称 A 是线性属性。例如出生时间。

(3.2)如果 $D_A$ 上的≤是个偏序，则称 A 是序属性。

$D_A$ 可能是一个集合的幂集，并且≤是个集合之间的包含关系。给定一个结构属性 A，有一个关系 $R_A$ 与之对应，$R_A$ 满足对于论域中的任何两个个体 X 和 Y，$XR_AY$ 当且仅当 X(A)≤Y(A)，并且这个关系有一个相应的名称。例如设 A=速度，X=军用飞机，Y=火车，则我们知道 Y(速度)≤X(速度)，并且军用飞机的速度不慢于火车的速度，火车的速度不快于军用飞机的速度。

**定义3.3**(线性属性 A 的边缘值) 如果 A 是线性属性，其取值域为一个区间[a,b]，称使得 max{|X−a|,|X−b|}<σ 成立的所有 X 为 A 的边缘值，其中0<σ<1/3是一个给定的常数。

**命题3.1**(结构属性的性质)

(3.3)如果 A 是个线性属性，则 $R_A$ 是线性的、传递的、反对称的和自反的；

(3.4)如果 A 是一个序属性，则 $R_A$ 是传递的、反对称的和自反的。

证明：仅需证明(3.3)成立，(3.4)的证明类似。已知 A 是个线性属性，即对于 $D_A$ 中的任何两个元素 X 和 Y，有 X≤Y，或者 Y≤X(不兼容的“或”)。从而由“$XR_AY$ 当且仅当 X(A)≤Y(A)”可知：$XR_AY$，或者 $YR_AX$，所以 $R_A$ 是线性的。

$R_A$ 是传递的：对于 $D_A$ 中的任何元素 X,Y 和 Z。若 X(A)≤Y(A)和 Y(A)≤Z(A)成立，则有 X(A)≤Z(A)。再由 $XR_AY\Leftrightarrow$X(A)≤Y(A)，$YR_AZ\Leftrightarrow$Y(A)≤Z(A)和 $XR_AZ\Leftrightarrow$X(A)≤Z(A)，所以有 $XR_AY\wedge YR_AZ\rightarrow XR_AZ$。

$R_A$ 是反对称的：因为对于 $D_A$ 中的任何两个元素 X 和 Y，有 X≤Y，或者 Y≤X。如果有 $XR_AY$ 和 $YR_AX$，即 X≤Y，并且 Y≤X，所以 X=Y。

$R_A$ 显然是自反的。证毕。

给定一个偏序≤，有一个严格偏序<，满足对任何的 X，Y∈$D_A$ 有：

维普资讯 http://www.cqvip.com

$X<Y \Leftrightarrow X \leqslant Y \wedge X \neq Y$

设 $R_A'$ 是与 $<$ 相应的关系，则有类似于 $R_A$ 的命题。

**命题3.2**($R_A'$ 的性质)

(3.5)如果 A 是线性的，则 $R_A'$ 是线性的(即对于任意的 X,Y，有 $X<Y$ 或 $Y<X$，$X=Y$)，传递的、非对称的和非自反的；

(3.6)如果 A 是序属性，则 $R_A'$ 是传递的、非对称的和非自反的。

证明：类似于命题3.1的证明。

**定义3.4**(分类型属性) 如果 $D_A$ 是 $H$ 中的子集，则称 A 是分类型属性。例如职业。

给定一个分类型属性 A。设个体 X 具有属性 A，则在 $D_A$ 中存在一个分支，其上面的任何节点，都是 X 在 A 上的值。对于分类型属性 A，有下面的公理：

**公理3.1** $\forall X \exists B \in H \forall C, C' \in B (C \subseteq C' \wedge X(A)=C \in K \rightarrow X(A)=C' \in K)$。其中，$K$ 是构造的领域知识库。

**定义3.5**(非结构属性) 若 A 的取值域 $D_A$ 是无结构的，则称 A 是非结构型属性。例如颜色。

上述不同类型的属性采用不同的联通度量方式。策略如下：首先将属性值进行量化，以进行数值处理。例如对于分类型属性，我们可以考虑属性值在 $H$ 中的最短路径的长度。

## 4 A-属性联通以及联通测度

A-属性联通是一类重要的概念联通。我们着重探讨 A-属性联通。符号说明，设 l(C)表示概念 C 所在概念分类 $H$ 中的层数；Lub(C,C')和 Glb(C,C')分别表示 C 和 C' 的极小上界集合和极大下界集合。

### 4.1 A-属性联通以及基本性质

为了讨论 A-属性联通，我们引入几个概念，这些概念引自中科院计算技术研究所的曹存根研究员的学术报告[11]。定义如下：

**定义4.1**(A-属性联通) 称 $C_1$ 和 $C_2$ 是 A-属性联通的，若 $C_1$ 和 $C_2$ 都有属性 A，并且两者在 A 有值。记为 $ita(C_1,C_2,A)$。

**定义4.2** 若 $C_1$ 和 $C_2$ 是 A-属性联通的。设 $C_1(A)=V_1$，$C_2(A)=V_2$：

(4.1)若 $V_1 \cap V_2 \neq \Phi$，则称 $C_1$ 和 $C_2$ 是 A-属性值联通；

(4.2)若 $V_1=V_2$，则称 $C_1$ 和 $C_2$ 是 A-完全属性值联通；

(4.3)若 $V_1$ 和 $V_2$ 中的元素在 $H$ 中有极小上界，则称 $C_1$ 和 $C_2$ 是属性值上位联通；

(4.4)若 $V_1$ 和 $V_2$ 中的元素是在 $H$ 中有极大下界，则称 $C_1$ 和 $C_2$ 是属性值下位联通；

(4.5)若 $V_1$ 和 $V_2$ 中的元素同时是上位和下位联通，则称 $C_1$ 和 $C_2$ 是属性值混合联通；

(4.6)若 A 是个序属性，则称 $C_1$ 和 $C_2$ 是 A-属性值序联通。

**定义4.3** 假设 A 和 A' 是线性属性。如果 $C_1$ 和 $C_2$ 同时是 A-属性值序联通和 A'-属性值序联通。设 $C_1(A)=V_1$，$C_1(A')=V_2$ 和，$C_2(A)=V_1'$，$C_2(A')=V_2'$。若 V1 和 $V_1'$ 是 $D_A$ 中的边缘值，则称 $C_1$ 和 $C_2$ 是 A-边缘联通。

**命题4.1** 若概念 $C_1$ 和 $C_2$ 是 A-完全属性值联通，则 $C_1$ 和 $C_2$ 也是 A-属性值联通。

**性质4.1** (A-属性联通的性质)

(4.7)$ita(C_1,C_2,A) \rightarrow ita(C_2,C_1,A)$(对称性)

(4.8)$ita(C_1,C_2,A) \wedge ita(C_2,C_3,A) \rightarrow ita(C_1,C_3,A)$(传递性)

(4.9)$\exists A(ita(C_1,C_2,A) \vee \neg ita(C_1,C_2,A))$(严格凸性)

### 4.2 A-属性联通测度

直观上，A-属性联通的测度应该反映出下面的主要性质：

·公共属性越多，两个概念的联通强度就应该越大；

·不同类型的属性应该采用不同的度量测度；

·与概念在属性上的取值有关。例如取值之间的联通强度越大，则概念在该属性上的联通强度就应该越大。

下面，我们先具体分析每种 A-属性联通的联通测度，然后给出求联通强度的算法。设 $ita(C_1,C_2,A)$，$C_1(A)=V_1$，$C_2(A)=V_2$。

若 $C_1$ 和 $C_2$ 是 A-属性值联通，则 $C_1$ 和 $C_2$ 在属性 A 上的联通测度定义为：

$$ind(C_1,C_2,A)=|V_1 \cap V_2|/|V_1 \cup V_2| \quad (4.1)$$

其中 $|*|$ 的含义需视 $V_1$ 和 $V_2$ 的具体类型而定。

若 $C_1$ 和 $C_2$ 是 A-属性值上位联通，则 $C_1$ 和 $C_2$ 在属性 A 上的联通测度定义为：

$$D^+(V_1,V_2)=\sum_{C' \in Lub(V_1,V_2)}\{1/|l(V_1)-L(C')|+1/|l(V_2)-l(C')|+1/(l(V_1)-l(V_2)|+1)\} \quad (4.2)$$

当两个概念有公共的子节点时，即使上联通度很小，其直观上的联通度也同样是较大的。为了反映这种情况，我们属性值下位联通度如下：

$$D^-(V_1,V_2)=\sum_{C' \in Glb(V_1,V_2)}\{1/|l(V_1)-l(C')|+1/|l(V_2)-l(C')|\} \quad (4.3)$$

注意(4.3)不包含 $C,C' \in Glb(C,C')$ 的情况，这种情况用公式(4.2)。根据(4.2)和(4.3)定义属性值混合联通测度如下：

$$D(C,C')=D^+(C,C')+D^-(C,C') \quad (4.4)$$

若 $C_1$ 和 $C_2$ 是 A-边缘联通，并且 A 中的属性值服从正态分布 $N(m,\sigma)$，则我们使用"小概率事件大联通"原则估算联通强度。若把属性看成随机变量，则落入 $3\sigma$ 区域内的 A-属性值的联通是很有意义的。对于一般的边缘联通，有两种情况：一是 $V_1$ 和 $V_2$ 是同一端的边缘值；二是 $V_1$ 和 $V_2$ 分别是两端的边缘值。

对一般的边缘联通，定义联通强度为如下的分段函数(假设 $V_1 <_D V_2$，$\sigma$ 是给定的常数)：

$$ind(C_1,C_2,A)=\begin{cases} 2-\min\{|\min(A)-V1|,|\max(A)-V2|\}/|\max(A)-\min(A)|; & \text{当 } \max\{|\min(A)-V1|,|\max(A)-V2|\}<\sigma \\ 2-|\min(A)-V1|/|\max(A)-\min(A)|; & \text{当 V1和V2是左端边缘值} \\ 2-|\max(A)-V2|/|\max(A)-\min(A)|; & \text{当 V1和V2是右端边缘值} \\ |V1-V2|/|\max(A)-\min(A)|; & \text{当 } \min\{|\min(A)-V1|,|\max(A)-v2|\}>\sigma \end{cases} \quad (4.5)$$

有了上面的分析以后，我们给出一个总的求属性联通的算法：

**联通测度算法**

已知：设 $A(C_1) \cap A(C_2)=\{A_1,A_2,\cdots,A_n\}$。个体 $C_1$ 和 $C_2$ 在 $\{A_1,A_2,\cdots,A_n\}$ 上的属性值向量分别为 $\alpha=(V_1,V_2,\cdots,V_n)$ 和 $\beta=(V_1',V_2',\cdots,V_n')$。$ind(C_1,C_2,A_i)$ 表示 $C_1$ 和 $C_2$ 在 $A_i$ 上的联通强度。$ind(C_1,C_2)$ 表示 $C_1$ 和 $C_2$ 的总联通强度。

求解：$C_1$ 和 $C_2$ 之间的联通强度 $ind(C_1,C_2)$

```
算法：FOR i=1 to n
        IF C1和C2是Ai-属性值联通，
        Then ind(C1,C2 Ai)=|V1∩V2|/|V1∪V2|;
```

维普资讯 http://www.cqvip.com

```
Else IF C1和C2是Ai-属性值混合联通,
    Then ind(C1,C2,Ai)=D(V1,V2);
  Else IF C1和C2是Ai-属性值上位联通,
      Then ind(C1,C2,Ai)=D+(V1,V2);
      Else IF C1和C2是Ai-属性值下位联通,
          Then ind(C1,C2,Ai)=D-(V1,V2);
          Else IF C1和C2是Ai-属性值序联通,
              Then IF Ai服从正态分布N(m,σ)
                Then ind(C1,C2,Ai)=1-min(P(|Vi- m|
              >3σ,P(|Vi'-m|>3σ)+常数(>1)
                  Else ind(C1,C2,Ai)=(4.4)
              Else ind(C1,C2,Ai)=1
ind(C1,C2)=ind(C1,C2)+ind(C1,C2,Ai)
```

现在,分析该算法的复杂性。设$|C|=m$,$Max\{|D_{Ai}|\,|n\geq i\geq 1\}=k$。在最坏情况下,该算法需执行指令数当$n\geq 3$时为$3n[(k^2)(m-2)+4]/4$,当$2\geq n$时为$2k^2mn$;在最好情况下,该算法需执行指令数为5n。

**总结** 本文利用本体这一手段重点分析了A-属性联通极其测度。主要内容是:第2部分讨论了领域本体模型(SOM)和概念本体模型(COM),并给出了命题2.1-命题2.4;第3部分讨论了基于属性取值域的属性划分,将属性划分为结构属性、分类型属性和非结构属性,并证明了命题3.1-命题3.2;第4部分探讨了A-属性联通以及联通测度。

基于属性的联通还有很多值得进一步讨论的问题。例如两个概念在不同属性上的联通问题;两个概念在两个相同属性上的取值以及取值之间的各种关系等。总之,我们认为基于属性的联通是一类很有意义的联通,可以运用于自然语言查询系统中,提高查询质量。

**致谢**:本论文的思想形成于在中科院计算所大规模知识处理课题组学习期间。

## 参考文献


1 史忠植著. 高级人工智能. 北京:科学出版社,1998.1~14
2 雷玉霞,眭跃飞. 基于Concept-Relation模型的知识联通. 计算机科学,2002,29(11):53~58
3 雷玉霞,眭跃飞,曹宝香. 基于固定概念分类H上的概念联通以及联通测度. 计算机科学(录用)
4 Lei Yuxia, Cao Cungen, Sui Yuefei. Based-CR Knowledge Interconnection in National Knowledge Infrastructure. ICII'2001, 2001,C:134~139
5 Gallant S I. Connectionist expert systems. Communications of the ACM,1988,31:152~169
6 Richardson S D, Dolan W B, Vanderwende L. MindNet: acquiring and structuring semantic information from text. Microsoft Research Technical Publications :[MSR-TR-98-23]. 1998
7 Richardson S. Determining similarity and inferring relations in a lexical knowledge base:[PhD. dissertation]. City University of New York,1997
8 Guarino N,Giaretta P. Ontologies and knowledge bases towards a terminological clarification. In:Proc. of KB&KS'95, 1995.25~32
9 Guarino N,Carrara M, Giaretta P. An Ontology of Meta-Level Categories. In:D. J. ,E. Sandewall,P. Torasso,eds. Principles of Knowledge Representation and Reasoning: Proc. of the Fourth Intl. Conf. (KR94). Morgan Kaufmann,San Mateo,1994. CA: 270~280
10 Guarino N, Welty C. A Formal Ontology of Properties. In: Proc. of 12th Int. Conf. on Knowledge Engineering and Knowledge Management,2000b
11 曹存根. 知识分析中的若干概念与方法:[中科院计算所学术报告]. 2001


(上接第91页)

整体更新指整个批次的数据进行更新。对于成果影像库,导入新的批次数据,迁移旧的批次数据到备份介质中,同时更新总表中批次数据变动情况。对于应用影像库,预处理后装载新数据,一般不删除旧数据,使用户可以在线浏览所有影像数据。

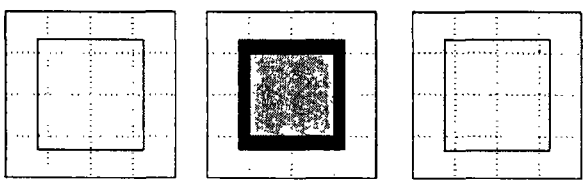


图6 遥感影像数据局部更新

局部更新指少量的标准分幅影像进行更新。对于成果影像库,一条记录对应一个分幅的影像,因此只需替换相应的数据即可。对于应用影像库则比较复杂。一个批次的影像构成一个完整的金字塔结构,金字塔的每一层对应不同分辨率的影像。更新时需要分别更新每一层的影像。以第i层影像为例,首先得到单幅影像相对该层影像的位置,图A。然后删除完全落入该范围的影像块,把其他的影像块重叠部分的像元赋值为零,见图B。之后对单幅影像按照该层影像的分辨率进行重采样并分块,把不重叠的影像块入库,重叠部分的影像块进行异或运算,见图C。其他层的更新以此类似。

4.2.2 *矢量数据更新* 矢量数据的更新比较复杂,矢量数据整体更新需要迁移旧的数据,然后把新的数据按照相同的表结构导入库中。

矢量数据局部更新分为两种情况,一种是通过客户端维护工具直接修改数据库中的矢量数据,特点是保持了数据的现势性,但丢失了历史数据。另一种是复制所要更新的图层表,对新表进行编辑修改,最后把表的变化情况记录到总表中,该方案的特点是数据的现状和历史都得到保存,但是增加了数据冗余。

### 4.3 图例的管理

在传统的GIS应用中,图例的制作管理是很费时的工作。在Oracle Spatial中,图例用数据库管理,用户可以任意添加自定义图例,方便了开发和应用。

**小结** 国内外空间数据库产品也实现了空间数据的有效管理,但是数据共享层次不够高,一般不支持第三方厂商的中间件产品访问空间数据,不能有效实现互操作。本文试图从这一角度回答这一问题。同时还存在很多问题需要研究。例如,JPEG2000算法除了高压缩比,还有渐进传输、局部解压以及兴趣点解压等优势没有很好地体现出来,还需要进一步研究和开发。

## 参考文献


1 李琦. 数字城市服务系统与示范总体框架设计和关键技术研究. 863技术报告. 基于SIG的数字城市体系结构和关键技术研究. 2003
2 郝力. 城市空间信息共享应用政策与协议研究. 863技术报告. 基于SIG框架的数字城市服务与示范研究. 2003
3 Barclay T,Eberl R,Gray J. The Microsoft TerraServer™. 1998
4 李军. 影像数据库的研究、设计和实现:[中国地质大学博士学位论文]. 2000
5 Sun Microsystems, Inc. 2000. Java™ 2 Platform Enterprise Edition Specification,1.4. 2002.12
6 Murray C. Oracle(r)Spatial User's Guide and Reference. Release 9.2. 2002.3
7 Adams M D. The JPEG-2000 Still Image Compression Standard, ISO/IEC JTC 1/SC 29/WG 1. 2002.12