

基于简化分辨矩阵的粗糙集属性约简算法^{*})

田卫东 周创德 胡学钢 周红鹃 李培培

(合肥工业大学计算机与信息学院 合肥 230009)

摘要 针对改进的经典算法求取属性约简存在的时间和空间性能不理想问题,本文提出了一种新的属性约简算法 ARSDM。该算法先将决策表按决策属性的类别划分,后采用边生成矩阵元素边约简边排序的思想求取属性约简,有效地加快了约简速度。实验表明 ARSDM 算法与经典算法相比具有较好的时间和空间性能。

关键词 数据挖掘,粗糙集,不一致性决策表,属性约简,分辨矩阵

A Simplified-discernibility-matrix-based Algorithm for Attribute Reduction in Rough Set

TIAN Wei-Dong ZHOU Chuang-De HU Xue-Gang ZHOU Hong-Juan LI Pei-Pei

(School of Computer and Information, Hefei University of Technology, Hefei 230009)

Abstract Traditional algorithm has relatively poor efficiency in both time and space when obtaining attribute reduction. Based on SDM, a new algorithm of attribute reduction called ARSDM is proposed in this paper. ARSDM takes the idea of classifying the universe of decision table according to the value of decision attribute firstly, then reducing and sorting elements of SDM while constructing them. The experimental study shows that the algorithm of ARSDM outperforms the traditional algorithms largely on both time and space.

Keywords Data mining, Rough set, Inconsistent decision table, Attribute reduction, Discernability matrix

1 引言

粗糙集理论^[1]是1982年由波兰数学家 Z. Pawlak 提出的一种处理模糊和不确定性知识的数学工具,经过二十多年的研究和发展,该理论已经广泛地应用于机器学习、决策分析、过程控制、故障诊断、模式识别、数据挖掘、动态目标识别及跟踪等领域^[2,3],并取得了较大的成功。

决策表属性约简是粗糙集理论中的核心内容之一,也是粗糙集理论研究较多的分支之一。属性约简的核心是消除多余属性,保留与问题相关的核心部分,从而提高决策的效率和效益,在机器学习与数据挖掘领域具有重要的应用价值。目前,已有许多关于粗糙集理论的属性约简算法,如 Z. Pawlak 提出的简单数据分析方法,Skowron^[8]提出的经典的分辨矩阵方法,叶东毅^[4]提出的基于属性重要度的贪心算法,王国胤等^[5]提出的用属性信息熵作为衡量属性重要度的标准而求取属性约简算法等。基于属性重要度的方法具有较好的时间性能,但这些属性重要度的定义可能存在不完备性,因此不一定能得到决策表的最佳属性约简。Skowron 的分辨矩阵方法能求取规模较小的数据集的所有属性约简及最佳属性约简,但由于该方法存在组合爆炸问题,难以适合规模较大的数据集。文^[6]提出了基于 Skowron 经典方法的一种改进方法,该方法避免了约简过程中从合取式到析取式的转化,并在属性组合前根据吸收率约简部分矩阵元素,一定程度上节省了时间,但也存在时空性能不理想等不足,使得知识的搜索和推理难以在一个较大的状态空间下进行。

本文针对改进后的经典算法(后文中简称经典算法)存在的不足,提出了一种基于简化的分辨矩阵 SDM (Simplified

Discernibility Matrix)求取属性约简的算法 ARSDM(Attribute Reduction Based on SDM)。该算法先对数据集进行预处理,删除冗余和不一致对象,并将决策表按决策属性的类别划分。本文算法在构造 SDM 得到矩阵元素时,将其作为约简条件删除识别函数中包含该条件的元素,避免了经典算法先得到所有矩阵元素,后排序再删除组合带来的时空消耗过大的弊端,较大程度减少了后期约简的时间开销。

2 粗糙集的基本概念和定义

定义 1 信息系统 S 是一个四元组: $S = \langle U, A, V, f \rangle$, 其中: $U = \{x_1, x_2, \dots, x_n\}$ 是对象或事例的非空有限集合; $A = \{a_1, a_2, \dots, a_m\}$ 是属性的非空有限集合。对每一个属性 a_i ($i = 1, \dots, m$), 信息函数 $f_{a_i}: U \rightarrow V_{a_i}$; V_{a_i} 是 a_i 的值域。

如果属性 A 可划分为两个不相交的集合即 $A = C \cup D$, 且 $C \cap D = \emptyset$, 则称信息系统为决策信息系统或决策表, 称 C 为条件属性集, D 为决策属性集(D 一般只含有一个属性)。

定义 2 在信息系统 S 中, 对于每个信息子集 $B \subseteq A$, 定义了一个不可分辨二元关系 $IND(B)$, 即 $IND(B) = \{(x, y) \in U \times U \mid \forall b \in B \text{ 有 } b(x) = b(y)\}$ 。显然, $IND(B)$ 是一个等价关系, $[x]_B$ 表示包含元素 x 的 $IND(B)$ 的等价类。

在一个决策表 $S = \langle U, C \cup D, V, f \rangle$ 中, 设 $U/IND(C) = \{U_1, U_2, \dots, U_m\}$ 表示条件属性集 C 对论域的划分, 其中 U_i ($i = 1, 2, \dots, m$) 称为基本块, 记为 U/C ; $U/IND(D) = \{D_1, D_2, \dots, D_p\}$ 表示由决策属性集 D 对论域 U 的划分, 记为 U/D 。

定义 3 在决策表 $S = \langle U, C \cup D, V, f \rangle$ 中, 若 $\forall B \subseteq C$, $POS_B(D) = POS_C(D)$ 且 B 相对于 D 独立, 则称 B 是 C 相对于 D 的属性约简。

^{*})基金项目:安徽省自然科学基金(050420207)。田卫东 副教授,博士,主要研究方向:图形图像与并行计算;周创德 硕士,研究方向:数据挖掘与人工智能;胡学钢 博士,教授,研究方向:知识工程,数据挖掘与人工智能。

设 $RED(C)$ 表示 C 的所有属性约简集合, 则有下面关系成立:

定理 1 条件属性 C 的核等于 C 的所有属性约简的交集, 即 $CORE(C) = \bigcap RED(C)$

定义 4 在决策表 $S = (U, CUD, V, f)$ 中, $a_i(x_j)$ 表示元素 x_j 在条件属性 a_i 上的取值, $d(x_j)$ 表示元素 x_j 在决策属性 d 上的取值。如果满足 $f(x_i, a) = f(x_j, a) \rightarrow f(x_i, d) = f(x_j, d) (a \in C; x_i, x_j \in U; d \in D)$, 则 S 为一致性决策表, 否则为非一致决策表。

定义 5 在 U_i 中, 若满足 $f(x_p, a) = f(x_q, a) \rightarrow f(x_p, d) = f(x_q, d) (\forall a \in C; x_p, x_q \in U_i; d \in D)$, 则称 U_i 是一致性基本块, 否则为不一致基本块。

如果 U_i 是一致性的基本块, 且 $|U_i| > 1$ (其中 $|\cdot|$ 为基数), 则基本块存在冗余。若 $S' = (U', CUD, V', f')$, 满足 $U' = (x'_1, x'_2, \dots, x'_m), x'_i \in U_i (i=1, 2, \dots, m)$, 则 S' 为简化的决策表。

定义 6 简化的分辨矩阵是定义在简化决策表 S' 上的一个 $m \times m$ 对称矩阵 $M_{m \times m} = (m_{st})$ 。矩阵元素定义: $m_{st} = \begin{cases} \{a | a \in C: a(x_s) \neq a(x_t), D(x_s) \neq D(x_t)\} \\ \emptyset & \text{其它} \end{cases}$, 其中 $s < t, x_s, x_t \in U'$ 且 $s, t = 1, 2, \dots, m$ 。

定义 7^[6] 设 X, Y 为 U 中的两个对象, A 为对象的属性集, 定义 X 和 Y 之间的距离函数 $\delta_A = \{a | a \in A, a(X) \neq a(Y)\}$; 且属性集 A 的识别函数: $\Delta A = \delta_A(X, Y) | (X, Y) \in U \times U$, 该定义说明识别函数 ΔA 为所有非空分辨矩阵元素组成的集合。

定义 8 设 m_{st}, m_{ij} 为分辨矩阵 SDM 的两元素, 且 $m_{st} \neq \emptyset$, 如果满足 $m_{st} \subseteq m_{ij}$, 即矩阵元素 m_{ij} 是 m_{st} 的超集, 则称 m_{st} 为 m_{ij} 的一个约简条件。

定义 9 设合取逻辑表达式:

$$f = \bigwedge_{1 \leq s, t \leq k, m_{st} \in \Delta A} \bigvee m_{st} \quad (1)$$

将式(1)转化为极小析取式 $f = \bigvee_{\theta=1}^N \bigwedge \partial_\theta$ (式 2), 其中, N 是式(1)展开后析取项的项数。则每个 ∂_θ 都是决策表的属性约简, 取 ∂_θ 所含属性元素最少的为该决策表的属性约简。

定理 2 若 $\forall m_{pq}, m_{ij} \in \Delta A$, 且 m_{pq} 是 m_{ij} 的一个约简条件, 则根据 ΔA 所得的属性约简与根据 $\Delta A - \{m_{ij}\}$ 所得的属性约简结果一致。

证明: 由于 m_{pq} 是 m_{ij} 的一个约简条件, 则 $m_{pq} \subseteq m_{ij}$ 。又由定义 6 中(1)式知, $f = \bigwedge_{1 \leq s, t \leq k, m_{st} \in \Delta A} \bigvee m_{st}$ (式 1), 令 $f = \{m_{ij}\} \wedge \{m_{pq}\} \wedge \dots \wedge m_{st}$, 且其析取式为 $f = \bigvee_{\theta=1}^N \bigwedge \partial_\theta$ (式 2), 若 $P = \{m_{pq}\}, Q = \{m_{ij}\}$ 为两命题, 由 $m_{pq} \subseteq m_{ij}$ 可知 $P \subseteq Q$, 根据吸收率可得 $P \wedge Q \Leftrightarrow P$, 因此, $f \Leftrightarrow f' = \{m_{pq}\} \wedge \dots \wedge m_{st}$ 。设 f' 的析取表达式为: $f' = \bigvee_{k=1}^{N_1} \bigwedge \partial_k$ (式 3), 现证明 $\forall \partial_k \in f' \Rightarrow \exists \partial_\theta \in f$, 有 $\partial_k \subseteq \partial_\theta$ 。因为 $f \Leftrightarrow$ 式 2, 且 $f' \Leftrightarrow$ 式 3, 现只需证明 $f' \subseteq f$ 即可, 从以上分析易知, $f' \subseteq f$ 结论明显成立, 同理可证 $f' \supseteq f$ 。故命题成立。

3 属性约简的算法描述

经典算法求取属性约简方法如下: 构造分辨矩阵得到识别函数 ΔA , 将 ΔA 中的元素按属性的个数从小到大排序, 判断删除 ΔA 中元素。从 ΔA 中剩余的每个集合元素中任取一

个属性元素形成组合, 所含属性元素个数最小的组合就是决策表的属性约简, 即 ∂_θ 中所含属性元素最少的析取式。分析得知, 经典算法存在如下不足: 构造分辨矩阵需存储所有矩阵元素, 消耗大量内存空间; 时间性能较差, 得到的 ΔA 数据规模较大, 对其排序, 时间开销巨大。

目前, 部分学者在利用分辨矩阵求取属性约简时, 通常先求出决策表的核属性, 后以核属性为启发信息约去包含该核属性的矩阵元素, 再求取决策表的属性约简, 具有较好的时间性能。但在许多决策表中无核属性或核属性个数较少, 因此可能会造成以核属性为启发信息的约简算法的时间消耗与经典算法相当。

针对以上情况, 本文提出 ARSDM 算法。

3.1 ARSDM 算法

ARSDM 算法分为两个阶段:

1. 预处理阶段: 删除不一致性对象并求取简化决策表, 返回决策属性的类别个数 ClassNum, 并将不同类别的对象分别存储在链表 ClassLabel[] 中。

2. 属性约简阶段: 将相异链表 ClassLabel[] 中的对象构造 SDM, 以其为启发信息判断是否有序并入 ΔA 或删除 ΔA 的元素。

ARSDM 算法基本框架如下:

```

输入: 决策表  $S' = (U, CUD)$ ; 属性维数 dimN; 存储数据项链表 datalist 和对象个数 recordNum;
输出: 属性约简的决策表  $S = (U, RED(A))$ 
Procedure ARSDM(S, dimN, datalist, recordNum)
{
  Preprocessor(S, dimN, datalist);
   $\Delta A = NULL$ ;
  For( $i=0; i < classNum; i++$ )
  {
    Idf = get the record index from ClassLabel[i];
    FirStr = get the Idfth element from datalist;
    For( $j=i+1; j < classNum+1; j++$ )
      Ids = get the record index from ClassLabel[j];
      SecStr = get the Idsth element from datalist;
      CompareAttr(FirStr, SecStr, datalist,  $\Delta A$ );
  }
  Return RED(A);
}
    
```

根据 ARSDM 算法, 函数 CompareAttr() 操作如下: 每得到一矩阵元素 m_{ij} 时, 即对识别函数 ΔA 进行有序搜索。如果 1) 从前至 m_{st} ($m_{st} \subseteq \Delta A$, 且 $|m_{st}| \leq |m_{ij}|$) 搜索 ΔA 中元素, 判断是否存在属于 ΔA 的某元素包含于 m_{ij} , 如果是, 则跳出; 否则 2) 把 m_{ij} 插到 m_{st} 元素后, 并从后至 m_{ij} 下一元素搜索 ΔA , 如果存在某元素 $m_{pq} \supseteq m_{ij}$, 则执行 $\Delta A = \Delta A - \{m_{pq}\}$ 。而经典算法先对求得的所有分辨矩阵的元素集合 ΔA 排序后, 再约简, 最后组合 ΔA 中元素得到属性约简, 其时间代价较大, 空间性能也较差。如 1000 条记录的数据集, 经典算法在最坏的情况下得到近 50 万个矩阵元素, 然后对其排序、删除、组合, 空间和时间上的消耗都是巨大的。而在 ARSDM 算法中每得到一矩阵元素即搜索 ΔA , 判断是否有序插入 ΔA 中或删除 ΔA 中的元素, 使得 ΔA 中元素之间不存在包含关系, 避免了经典算法内存消耗过大的问题, 也在很大程度上提高了时间性能。

3.2 ARSDM 算法的正确性和复杂度分析

经典算法先得到所有分辨矩阵元素, 后排序删除使得最终 ΔA 中矩阵元素不存在包含关系, 再从 ΔA 中的每个矩阵元素中任取一个属性元素形成组合, 所含属性元素个数最小的组合就是决策表的属性约简。若有 $\Delta A = \{ab, ac\}$, 则组合后得 $\{a\}, \{ac\}, \{ba\}, \{bc\}$, 若 ΔA 是某决策表 S 的识别函数, 则 $\{a\}$ 是 S 的最佳属性约简。ARSDM 算法先对决策表进行

预处理,删除冗余及不一性对象。冗余的对象在构造分辨矩阵时,会造成冗余的矩阵元素,根据定理 2 可知,采用留一法可以正确求取决策表属性约简;文[7]指出,在不一致对象中不会提取任何有用的规则,但会在规则提取中增加计算量,因此需对数据集进行一致化操作。ARSDM 算法在生成分辨矩阵元素时,即判断是否有序并入 ΔA 或删除 ΔA 的元素,将经典算法后续工作提前,并一步完成。定理 2 中证明了以某矩阵元素为启发信息约简识别函数中矩阵元素不改变决策表的属性约简质量。因此,ARSDM 算法与经典算法的属性约简质量和数量一致。

根据文献知,经典算法求取一个属性约简时间复杂度为 $O(C \| U \|^2 + |\Delta A| (\log |\Delta A| + 1))$,其中 u 为决策表的论域, ΔA 为识别函数。ARSDM 算法在预处理过程中,已对决策表的决策属性标识归类,在求取分辨矩阵时,对于同类别决策属性的对象,不需再求取分辨矩阵元素。设决策属性的类别数为 D_n ,令 $U/D = \{D_1, D_2, \dots, D_{D_n}\}$,且 N_i 是基本块 D_i 中对象的个数,其中 $i=1, 2, \dots, D_n$, ARSDM 算法求取一个属性约简时间复杂度为 $O(|C| \sum_{i=1, \dots, D_n-1} (N_i \sum_{j=i+1, \dots, D_n} N_j))$ 。在最差的情况下,时间复杂度为 $O(|C| |U'|^2)$ 。为了进一步验证本文算法的有效性,下面以实例加以说明之。

3.3 实例分析

表 1 为预处理后的决策表 $S=(U, C \cup D)$, $U=(x_1, x_2, \dots, x_7)$; $C=(a, b, c, d, e)$, f 为决策属性。

表 1 决策表 S'

	a	b	c	d	e	F
X1	1	0	2	1	1	1
X2	1	1	2	0	1	0
X3	1	2	0	0	2	2
X4	1	2	2	1	0	1
X5	2	1	0	0	2	2
X6	2	1	0	2	2	0
X7	2	1	1	2	2	1

表 3 Results of the ARSDM

数据库	实例数	原属数 c_n	预处理后实例数	约简后属性数 C_m	经典算法耗时(s)	ARSDM 耗时(s)	约简率
Balloons	20	4	16	2	0.010	0.000	50%
Zoo	101	16	59	14	0.250	0.032	12.5%
Breast-cancer	286	9	266	7	2.591	0.125	22.2%
Monks	432	6	432	3	29.132	0.344	50%
Tic-Tac-Toe	958	9	958	8	397.142	4.141	11.1%
Solar Flare	1066	12	363	9	56.173	0.015	33.3%
Car	1728	6	1728	6	751.751	9.469	0
Mushroom	2346	22	2346	4	1016.582	15.159	81.8%
Nursery	3000	8	3000	7	3356.126	55.128	12.5%
ChessEnd Game	3196	36	3196	29	4167.765	95.766	19.4%

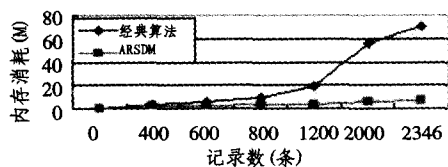


图 1 随记录数增加内存消耗情况

为了验证 ARSDM 算法空间性能,特选取 mushroom 数据库作为实例,分析该算法与经典算法的空间性能,结果如图

ARSDM 算法构造分辨矩阵如下:(\emptyset)外是根据 ARSDM 算法得出的元素值,(\emptyset)内是根据经典算法得到的矩阵元素)

表 2 分辨矩阵

U	x1	x2	x3	x4	x5	x6	x7
x1	\emptyset	\emptyset (bd)	\emptyset (bcde)	\emptyset (\emptyset)	\emptyset (abcde)	\emptyset (abcde)	\emptyset (\emptyset)
x2		\emptyset	\emptyset (bce)	\emptyset (bde)	\emptyset (ace)	\emptyset (\emptyset)	\emptyset (acde)
x3			\emptyset	\emptyset (cde)	\emptyset (\emptyset)	\emptyset (abde)	\emptyset (abcd)
x4				\emptyset	\emptyset (abcde)	\emptyset (abcde)	\emptyset (\emptyset)
x5					\emptyset	d(d)	\emptyset (cd)
x6						\emptyset	c(c)
x7							\emptyset

根据 ARSDM 算法,先对决策表 S 进行预处理,决策属性相同的对象以同一链表存储,得到 $\rightarrow x_1 \rightarrow x_4 \rightarrow x_7, \rightarrow x_2 \rightarrow x_6, \rightarrow x_3 \rightarrow x_5$ 三链表。根据定义构造 SDM,得到 $\Delta A = \{bd\}$, $m_{15} = \{abcde\}$ 。由于 $\{bd\} \subseteq \{abcde\}$,因此 m_{15} 并不并入 ΔA 中。当得到 $m_{43} = \{cde\}$ 时,因为 $\Delta A = \{bd\} \not\subseteq \{cde\}$,且 $|m_{12}| < |m_{43}|$,把 m_{23} 并入 ΔA ,则 $\Delta A = \{bd, cde\}$ 。根据 ARSDM 算法,最终得到的识别函数 $\Delta A = \{c, d\}$,所以决策表 S' 的约简为 $S' = (U', cd)$ 。经典算法首先得到所有的矩阵元素 $\Delta A = \{bd, bcde, abcde, bce, bde, ace, acde, cde, abde, abcd, d, cd, c\}$,然后排序 $\Delta A = \{d, c, bd, cd, bce, bde, ace, cde, bcde, abde, abcd, abcde\}$,再执行删除操作,得到 $\Delta A = \{d, c\}$ 。易得 ARSDM 算法与经典算法属性约简结果一致。

3.4 实验结果及性能分析

试验数据选取了 UCI 机器学习数据库[9]中 10 个具有离散属性的典型数据库,分别利用经典算法和本文提出的 ARSDM 算法对它们进行属性约简。实验以 P4, 2.66GHz, 256M 内存为硬件环境,以 WinsXP, VC.NET 为软件开发环境。为保证公平性,经典算法也是在约简后的决策表上进行试验的。实验结果如表 3 所示。

1 所示,ARSDM 算法内存消耗远低于经典算法。经典算法内存消耗速度增长快速。本文算法在生成矩阵元素的过程即做了约简,因此,随着记录数的增加,内存消耗增长速度缓慢,空间性能有较大的提高。

结束语 决策表属性约简理论的主要思想是在保持分类能力不变的前提下,删除知识库中的冗余知识,对条件属性进行约简,进而得到简化后的分类规则。本文 ARSDM 算法首先对数据集进行预处理,删除冗余和不一性对象,并把决策表按决策属性的类别标识归类,避免了相同决策类的对象之间

的操作,从而减少了时间和空间消耗。ARSDM 算法以约简条件为启发信息,采用边得到矩阵元素边删除 ΔA 中元素边排序的方法,有效地节省了约简的时间和空间消耗,提高了约简的效率。从实验结果可知,ARSDM 算法对较大规模数据库的处理比经典算法有明显优势。

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, et al. Rough Set [J]. Communication of the ACM, 1995, 38(11): 89~95
- 3 徐捷,徐从富,耿卫东. 基于粗糙集理论的动态目标识别及跟踪 [J]. 电子学报, 2002, 30(4): 605~607

- 4 赵斌,黄翠微,叶东毅等. 粗糙集中属性约简的一个贪心算法 [J]. 系统工程与电子技术, 2000, 22(9): 63~65
- 5 王国胤,于洪,杨大春. 基于条件信息熵的决策表约简 [J]. 计算机学报, 2002, 25(7): 759~766
- 6 王清毅,范焱,蔡庆生. 知识的约简研究 [J]. 小型微型计算机系统, 2000, 21(6): 623~627
- 7 谭天乐,宋执环,李平. 信息系统数据清洗. 规则提取的矩阵算法 [J]. 信息与控制, 2003, 32(4): 289~294
- 8 Skowron A, Rauszer C. The discernibility matrices and functions in information system [A]. Intelligent Decision Support Handbook of Applications and Advances of the Rough Sets Theory [C], Dordrecht: Kluwer Academic Publishers, 1992. 331~338
- 9 <http://www.ics.uci.edu/~mlearn/databases/>

(上接第 116 页)

· 数据源: 地理信息数据库, 语义信息数据库。

4.2 系统的主要功能

4.2.1 映射与命令分发

当用户发出请求后,请求首先到达系统内置的元数据模型(CIDOC CRM 与地理本体结合实现)。通过元数据模型到各个数据库的映射机制,对请求进行转换,分解成针对各个数据库的命令,发送到数据库层。对数据库进行操作后,所得到的信息再反馈回用户端,显示给用户。

4.2.2 信息浏览

通过图形化用户界面,系统实现与用户的交互。用户可以在 GIS 数据库提供的地图上直观地看到各个文物点的分布,并且选择浏览他感兴趣的文物的相关信息。与文物相关的语义信息由语义信息数据库提供。

4.2.3 信息查询

提供各类数据的查询功能,如提供文物保护相关数据的查询,包括数据的简单查询(根据文物类别,选择要查询的数据表和字段,获取用户需要的数据);数据综合查询(通过指定数据查询类型,任意组合数据过滤条件,获取用户所需要的数据)。

4.2.4 信息发布功能

可通过互联网工具进行一些公众文物保护信息和相关法律、法规信息的发布等。

结论 本文从文物领域本体模型 CIDOC CRM 对不可移动文物数据描述的局限性出发,分析了 CIDOC CRM 和地理本体两个元数据模型对不可移动文物的描述方式,并提出了将两者融合的方法,给出了实例并加以说明。同时,结合了 GIS 在表现不可移动文物空间特征的优势,提出了基于地理本体的不可移动文物数字博物馆的系统架构,为不可移动文物的保护管理提供了新的思路 and 手段。

参 考 文 献

- 1 William S, Austin T. Ontologies. IEEE Intelligent Systems, Jan./Feb. 1999. 18~19
- 2 CIDOC Conceptual Reference Model. <http://cidoc.ics.forth.gr/>
- 3 Brickley D, Hunter J, Lagoze C. ABC: A Logical Model for Metadata Interoperability? Harmony Project, Working Paper. 1999. http://www.ilt.bris.ac.uk/discovery/harmony/docs/abc/abc_draft.html
- 4 The Getty Information Institute and College Art Association. Categories for the Description of Works of Art(CDWA). <http://www.getty.edu/gri/standard/cdwa/>
- 5 Liu Hong-zhe, Bao Hong. An Information Flow-based Approach to Semantic Integration of Distributed Digital Museums. In: Proceedings Machine Learning and Cybernetics, Volume 7 of 7, 2006. 4430~4437
- 6 Bao Hong, Liu Hongzhe, Yu Jiehua. Ontology-based Semantic Integration for Digital Museums. In: Lecture Notes in Computer Science (LNCS), 3739 Chapter, Springer, 2005. 0302~9743
- 7 刘宏哲,鲍泓,余杰华. 基于 CIDOC CRM 的虚拟博物馆语义网络架构. 计算机应用研究, 2006(3): 050~054
- 8 鲍泓,刘宏哲. 基于 Web Services 的虚拟文物博物馆架构. 系统仿真学报, 2005, 17(6): 1412~1417
- 9 陈建军,周成虎,王敬贵. 地理本体的研究进展与分析. 地学前沿, 2006, 13(3): 081~090
- 10 Wang Jingui, Fen Zhen. Oceanographic ontology-based spatial knowledge query. Acta Oceanologica Sinica, 2005, 24(4): 66~71
- 11 Fonseca F. Ontology-driven Geographic Information Systems. The University of Maine, 2001
- 12 Fonseca F, Egenhofer M, Agouris P, et al. Using ontologies for integrated geographic information systems. Transactions in GIS, 2002, 6(3): 231~257