

基于演化传播模型的本体内聚程度分析

刘 磊,张 鹏,吕 帅,张 睿

(吉林大学计算机科学与技术学院,吉林长春 130012)

摘 要: 本体演化是本体工程领域的研究热点之一,当前研究大多侧重于演化策略或演化过程中使用的技术的研究,对于如何量化本体适应变化的能力至今没有明确结论,本文从一个全新的角度来解决这一问题,通过计算给定本体的内聚程度来量化本体适应变化的能力.本文提出了一种包含本体变更影响信息的演化传播模型;定义了能使该模型形式化表示的编码 ETMC;基于演化传播模型和 ETMC 分析了影响本体内聚程度的指标,并给出了使用本体内聚程度衡量本体适应变化能力的分析方法;在描述能力等价的本体上做了本体演化适应能力的判别实验,实验表明,本文提出的方法有较好的效果和实用性,可以为如何构建适应变化能力较强的本体提供理论指导.

关键词: 本体演化;演化传播模型;内聚度

中图分类号: TP311 **文献标识码:** A **文章编号:** 0372-2112 (2011) 08-1845-07

Ontology Cohesion Analysis Based on Change Transmissibility Model

LIU Lei, ZHANG Peng, LÜ Shuai, ZHANG Rui

(College of Computer Science and Technology, Jilin University, Changchun, Jilin 130012, China)

Abstract: Ontology evolution is a research hotspot in ontology engineering. Nowadays, most of the researches are related with evolution strategies and the technology used in evolution, but there are no clear conclusions for the problem how to scale the ability of ontologies for evolution. This paper solves the problem from a new point of view, that it uses ontology cohesion to determine the degree of ontology evolution. We propose a new Change Transmissibility Model that contains the information of ontology evolution; we define an encoding method called ETMC, which is used to represent the model; we analyze the factors that affect the ontology cohesion and propose an approach to compute the ontology cohesion based Change Transmissibility Mode and ETMC encoding method; we carry out our experiments on two ontologies which describe the same domain knowledge, and the experimental results show that our approach is feasible on the problem of determining the ability of ontology evolution.

Key words: ontology evolution; change transmissibility model; cohesion

1 引言

随着语义网的发展,本体已成为描述资源信息的一种重要手段,在语义 Web 中,基于本体的 Web 应用程序很大程度上依赖于支撑此应用的领域本体,然而,领域的变化、适应不同的应用背景以及概念化世界的改变都可能引起本体的演化^[1]. Stojanovic 等人将本体演化定义为“本体对出现变化的适时修改以及这些变化到附属产物的一致性传播(或称变化管理)”^[2]. 由于本体演化的普遍性和重要性,国内外的许多学者都对本体演化进行了研究.

在国外,本体演化问题最早是在与数据库模式演化的对比研究中提出的, Noy 等人分析了本体演化在使用模式、显示语义和知识模型方面与数据库模式演化的不

同,并分析了这些不同对本体演化框架的影响^[3]; Hasse 等人讨论了 OWL 本体的一致性演化问题,给出了 OWL 本体结构一致性、逻辑一致性和用户定义一致性的模型^[4]; Atluri 等人通过语义形式化和结构形式化规范说明解决了本体演化过程中数据合法性的不确定性问题^[5];

在国内,中山大学的明仲等人通过扩展继承的数学理论,系统研究了本体的继承语义、继承机制等问题^[6];吉林大学的刘磊等人提出了一种本体图模型,建立了本体邻接矩阵和可达矩阵,凭借矩阵变换与运算对本体演化中的波及效应进行了初步分析和量化定界^[7];中国科学院的韩燕波等人提出了一种可以有效减小本体变更影响范围的本体演化算法,该算法分析了本体实体和服务之间的依赖关系并提出了量化变更影响范围的数学

公式,根据量化结果将本体演化过程转化为图的启发式搜索过程^[8].

上述研究从不同角度对本体演化中的问题进行了有意义的尝试,但如何量化本体适应变化的能力,当前的研究仍没有一个明确的结论.本文的研究从本体自身出发,提出了一种包含变更影响信息的演化传播模型,基于该模型给出了一种计算本体内聚度的方法,根据本体的内聚程度衡量本体适应演化的能力,本体内聚度越低,本体越适合演化,内聚度越高,越不适合演化.

2 演化传播模型

2.1 基本定义

定义 1 本体由一个五元组 $\langle C, P, R, A, I \rangle$ 构成, C 表示领域中的概念, P 表示概念的一些基本属性, R 表示概念间的关联关系, A 表示用来约束概念间的关系 R 的公理, I 表示概念 C 的实例.

本文通过本体抽象结构分析本体的内聚程度,这种抽象结构指概念间的相互关联,所以本文的研究没有考虑实例节点.

定义 2 一个本体 O 的演化传播模型是一种有向图 $G = \langle V, E \rangle$, 其中 V 包括四种节点,分别为类节点、关系节点、类属性节点和关系属性节点, E 是一种有向依赖关系.

在本体 O 中,任何一个概念在演化传播模型中对应的节点为类节点;任何一个概念所对应的基本属性集合抽象为一个节点,该抽象化的节点为类属性节点;任何一个关系在演化传播模型中对应的节点为关系节点;任何一个关系所对应的公理集合抽象为一个节点,该抽象化的节点为关系属性节点.通过分析,可以得出结论:一个类节点对应一个类属性节点,一个关系节点对应一个关系属性节点.

定义 3 在演化传播模型中,如果一个节点的变化会影响另一个节点,则称二者存在着依赖关系,如果两个节点直接相连,则称这种关系为直接依赖关系,否则为间接依赖关系.

定义 4 演化传播模型中的两个节点 A 和 B , 如果 A 发生变化会影响 B , 则两节点间依赖关系的方向为 A 到 B .

根据本体构建演化传播模型时,节点间依赖关系的方向由一组依赖规则决定,规则的具体定义如下:

规则 1 子类节点依赖于父类节点;

规则 2 子关系节点依赖于父关系节点;

规则 3 关系节点依赖于与关系对应的定义域类节点;

规则 4 关系节点依赖于与关系对应的值域类节

点;

规则 5 关系属性节点依赖于与关系属性节点对应的关系节点;

规则 6 类节点依赖于与类节点对应的类属性节点;

规则 7 类属性节点依赖于与类属性节点对应的类节点;

规则 8 关系节点依赖于与关系节点对应的关系属性节点;

规则 9 若本体 O 中存在公理 $C = C1 \cup C2$, 则在演化传播模型中,概念 C 所对应的类节点分别依赖于概念 $C1$ 和概念 $C2$ 所对应的类节点;

规则 10 若本体 O 中存在公理 $C = C1 \cap C2$, 则在演化传播模型中,概念 C 所对应的类节点分别依赖于概念 $C1$ 和概念 $C2$ 所对应的类节点;

规则 11 若本体 O 中存在公理 $C = \neg C1$, 则在演化传播模型中,概念 C 所对应的类节点依赖于概念 $C1$ 所对应的类节点;

规则 12 若本体 O 中存在公理 $C = \exists R. C1$, 则在演化传播模型中,概念 C 所对应的类节点分别依赖于概念 $C1$ 所对应的类节点和关系 R 所对应的关系节点;

规则 13 若本体 O 中存在公理 $C = \forall R. C1$, 则在演化传播模型中,概念 C 所对应的类节点分别依赖于概念 $C1$ 所对应的类节点和关系 R 所对应的关系节点;

规则 14 若本体 O 中存在公理 $C = \geq nR$, 则在演化传播模型中,概念 C 所对应的类节点依赖于关系 R 所对应的关系节点;

规则 15 若本体 O 中存在公理 $C = \leq nR$, 则在演化传播模型中,概念 C 所对应的类节点依赖于关系 R 所对应的关系节点.

2.2 演化传播模型的构建

给定一个本体 O , 本文将按算法 1 来构建演化传播模型.

算法 1 函数 Build_Mode

```
Build_Mode( In: Ontology  $O$ , Out: Model  $G$  ) {
     $G.V = \emptyset$ ;
     $G.E = \emptyset$ ;
    For each  $c \in S$ 
         $G.V = G.V \cup \{inter(c)\} \cup \{inter(PR(c))\}$ ;
    For each  $r \in R$ 
         $G.V = G.V \cup \{inter(r)\} \cup \{inter(PR(r))\}$ ;
    For each  $x_1 \in G.V$ 
        For each  $x_2 \in G.V$ 
            If ( Rely_On( $x_1, x_2$ ) )
                 $G.E = G.E \cup \{L(x_1, x_2)\}$ ;
}
```

算法 1 中, S 表示本体 O 中概念组成的集合, R 表示本体 O 中关系组成的集合, G 表示生成演化传播模型, $G.V$ 表示演化传播模型中的点集合, $G.E$ 表示演化传播模型中的边集合; $L(x, y)$ 表示有向图中的边, x 为边的起点, y 为边的终点; $inter$ 函数的作用是将本体中的节点转换为演化传播模型中的节点, $Rely_On(x, y)$ 根据依赖规则判别演化传播模型中的两个节点 x 和 y 是否具有依赖关系, 若 y 依赖于 x 则返回真值, 否则返回假值.

$Build_Model$ 函数对给定的本体 O 完成了与其对应的演化传播模型 G 的构建, 先根据 O 中的节点生成 G 中的点集, 再根据 G 中节点间的依赖关系生成 G 的边集.

3 ETMC 编码

演化传播模型中包含本体变更所需要的信息, 为分析本体内聚程度提供指导. 传统的有向图的表示方式为邻接表或邻接矩阵, 但这两种方式不适合信息的形式化表示, 而且矩阵占的空间比较大, 不适合用来表示演化传播模型. 本文考虑用一种编码的方式来表示演化传播模型, 将该编码方式称为 ETMC (Evolution Transmit Model Code).

编号规则: 对一演化传播模型, 分别依次为其所有类节点和关系节点从 0 开始分配序号, 若类节点 c 的序号为 X , 则 c 的编号为 $C \circ X$, 若关系节点 r 的序号为 Y , 则 r 的编号为 $R \circ Y$.

根据编号规则, 同样可以为类属性节点和关系属性节点编号, 若一类节点或关系节点的编号为 X , 则其对应的类属性节点或关系属性节点的编号为 $P \circ X$.

使用编号规则为演化传播模型中的类节点和关系节点编号, 用编码规则为演化传播模型中的节点生成对应的 ETMC 编码.

编码规则 1 取演化传播模型中任一类节点 x , 若 x 根据编号规则得到的编号为 CX , x 直接依赖的类节点和关系节点的编号依次连接所构成的字符串为 S , 则节点 x 的 ETMC 编码为 $CX \circ S$.

编码规则 2 若演化传播模型中一类节点的 ETMC 编码为 $code$, 则与其对应的类属性节点的 ETMC 编码为字符 $P \circ code$.

编码规则 3 取演化传播模型中任一关系节点 x , 若 x 根据编号规则得到的编号为 RX , x 直接依赖的类节点和关系节点的编号依次连接所构成的字符串为 S , 则节点 x 的 ETMC 编码为 $RX \circ S$.

编码规则 4 若演化传播模型中一关系节点的 ETMC 编码为 $code$, 则与其对应的类属性节点的 ETMC 编码为字符 $P \circ code$.

4 量化计算

为量化分析本体适应演化的能力, 本文提出两个影响本体内聚程度的指标: 波及度和影效值. 本节将介绍在演化传播模型中节点波及度和影效值的定义及计算方法.

4.1 量化指标

定义 5 在本体 O 所对应的演化传播模型中, 如果一个节点 X 发生了变化, 则将演化传播模型中受影响的节点个数称为节点 X 的波及度, 记做 $B(X)$.

定义 6 在本体 O 所对应的演化传播模型中, 如果一个节点 X 发生了变化, 则将该变化对本体 O 产生影响的大小称为节点 X 的影效值, 记做 $Eff(X)$.

4.2 波及度的分析与计算

由于演化传播模型共包含 4 种节点, 因此节点 X 发生变化也会影响 4 种节点. 令 $B(X, C)$ 返回节点 X 发生变化时, 演化传播模型中受影响的类节点的数目; $B(X, P)$ 返回节点 X 发生变化时, 演化传播模型中受影响的类属性节点的数目; $B(X, R)$ 返回节点 X 发生变化时, 演化传播模型中受影响的关系节点的数目; $B(X, A)$ 返回节点 X 发生变化时, 演化传播模型中受影响的关系属性节点的数目. 则有公式:

$$B(X) = B(X, C) + B(X, P) + B(X, R) + B(X, A) \quad (1)$$

4.2.1 波及度特征分析

当 X 为类节点时, 由于类属性节点与类节点一一对应, 根据依赖规则 7 可知, X 发生变化受影响的类属性节点个数等于受影响的类节点个数加 1, 这个 1 是 X 对应的类属性节点, 即 $B(X, P) = B(X, C) + 1$; 同样由于关系属性节点与关系节点一一对应, 根据依赖规则 5 可知, X 发生变化受影响的关系属性节点与受影响的关系节点数目相等, 即 $B(X, A) = B(X, R)$. 综上所述, 则有如下公式:

$$B(X) = 2 \times (B(X, C) + B(X, R)) + 1 \quad (2)$$

当 X 为其他三种节点时, 通过类似的分析, 可以分别得到式(3)~(5), 其中 $PtoC(X)$ 和 $PtoR(X)$ 分别返回演化传播模型中与 X 对应的类节点和关系节点. 值得注意的是式(2)和式(4)在结论上是一样的, 但公式得到的方式是不同的.

$$B(X) = 2 \times (B(PtoC(X), C) + B(PtoC(X), R)) + 1 \quad (3)$$

$$B(X) = 2 \times (B(X, C) + B(X, R)) + 1 \quad (4)$$

$$B(X) = 2 \times (B(PtoR(X), C) + B(PtoR(X), R)) + 1 \quad (5)$$

4.2.2 $B(X)$ 的计算方法

观察式(2)、(3)、(4)和(5), 可以得到一个结论: 只

要可以计算类节点或关系节点发生变化时,受影响的类节点和关系节点的数目之和,就可以计算演化传播模型中任意节点发生变化时,受影响节点的总数,即 $B(X)$ 的值.

当演化传播模型中的类节点或关系节点发生变化时,本文将按算法 2 求得该节点所影响的类节点和关系节点的数目之和,进而求得该节点的波及度.

算法 2 函数 SumCRB

```

SumCRB(In: ETMC node, Out: Int n) {
    Q = ∅;
    S = ∅;
    num = NUM(node);
    Q.push(num)
    While(Q! = ∅) {
        numx = Q.pop();
        S = S ∪ {numx};
        For each nodex ∈ ETS
            If (ISub(numx, nodex) && NUM(nodex) ∉ S)
                Q.push(NUM(nodex));
    }
    n = Amount(S);
}

```

算法中, Q 是一个存放类节点或关系节点的编号的队列, $Q.push$ 函数用来完成向队列中添加元素的操作, $Q.pop$ 函数完成从队列中取出头元素的操作; S 是一个存放类节点或关系节点的编号的集合; $Amount(S)$ 返回集合 S 中元素的个数; $NUM(node)$ 返回 $node$ 对应节点的编号; $ISub(num, node)$ 判定 $node$ 是否为编号为 num 的节点的子类节点; ETS 表示由演化传播模型中所有类节点和关系节点的 ETMC 编码所构成的集合.

SumCRB 函数完成了类节点 $node$ 发生变化时受影响的类节点和关系节点数目之和的计算,函数中, $node$ 为类节点或关系节点的 ETMC 编码. 为保证演化传播模型中出现环时仍能正确计算出 $node$ 节点发生变化所影响的节点个数,本文采用一个集合 S 来保存依赖于 $node$ 节点的类节点和关系节点,直到 S 收敛时,才计算 S 中的元素个数.

4.3 影效值的计算

波及度反映了变化波及的广度,但并不能反映变化影响的大小,因为变化对不同节点的影响是不同的,影响会随着传播距离的增加而降低,影响会随着变化节点 X 种类的不同而不同.

设本体 O 中节点 X 发生变化会影响 n 个节点,则 $Eff(X) = k_x \cdot \sum_{i=1}^n \frac{k_i}{d_i}$, 其中 d_i 表示受 X 变化影响的第 i 个节点与节点 X 间的最短路径长度, k_x 表示节点 X 的权重, k_i 表示受 X 变化影响的第 i 个节点的权重,各节

点权重的取值可以根据需求人为指定.

在本文中,类节点和关系节点的权重都取值为 1,类属性节点和关系属性节点的权重分别取值为该节点在原本体模型中所对应的基本属性节点或约束节点的个数. 本文用 ETMC 编码来表示演化传播模型,只要能从演化传播模型的 ETMC 编码得到演化传播模型,就可以根据 Dijkstra 算法求得演化传播模型中任意两节点间的最短路径长度. 将 ETMC 编码转换为演化传播模型的过程由函数 CodeToG 完成.

算法 3 函数 CodeToG

```

CodeToG(In: ETMC * ETS, Out: Model G) {
    G.V = ∅;
    G.E = ∅;
    For each node ∈ ETS
        G.V = G.V ∪ {NUM(node)};
    For each numx ∈ G.V
        For each node ∈ ETS
            If (ISub(numx, node))
                G.E = G.E ∪ {L(numx, NUM(node))};
    For each node ∈ ETS {
        G.V = G.V ∪ {P ◦ (NUM(node))};
        G.E = G.E ∪ {L(P ◦ (NUM(node)), NUM(node))};
        G.E = G.E ∪ {L((NUM(node)), P ◦ NUM(node))};
    }
}

```

该函数先根据 ETMC 编码生成对应演化模型的类节点和关系节点以及体现这些节点间依赖关系的边,再根据类节点和关系节点生成与其对应的类属性节点和关系属性节点,最后生成体现类节点与类属性节点及关系节点与关系属性节点间依赖关系的边,完成由 ETMC 编码生成演化传播模型的过程. 根据 Dijkstra 算法可以求得 CodeToG 函数生成的图中的任意两个节点间的最短路径长度,进而求得影效值,对本体的内聚程度进行分析.

5 本体内聚度分析与量化

只要能量化本体演化过程中每个变化动作的波及效应,就可以量化整个本体演化过程的影响,进而判断判断本体演化的难易. 一个本体演化过程的影响越大,本体演化越难;本体演化过程的影响越小,本体演化越容易. 在本体演化中,最基本的变化动作包括删除、添加和变更三种,其他的复杂变化都是基本变化的复合,所以本文只对基本变化的波及效应进行分析,进而得出影响本体内聚度的相关因素.

(1) 删除一个节点: 删除该节点所波及的范围可由 $B(X)$ 得出, 如果该节点的 $B(X) = 0$, 则说明删除该节点不会影响本体中的任何节点, 可以直接删除该节点;

删除该节点对整个本体产生影响的大小可由 $Eff(X)$ 来计算,如果该节点的 $Eff(X) = 0$,则说明尽管删除该节点可能会对本体中的其他节点造成影响,但这个影响的值为 0,所以同样可以直接删除该节点.在本体需要发生变化时,可以根据 $Eff(X)$ 的值对本体中的所有节点进行排序,尽量选择 $Eff(X)$ 小的节点发生变化,这样可以降低变化所带来的影响.

(2) 添加一个节点:增加一个节点时,需要对原来 ETMC 编码做适当修改并增添新增节点的 ETMC 编码,之后根据 $B(X)$ 和 $Eff(X)$ 的值来确定增加节点的波及范围和影响大小.

(3) 变更一个节点:这个变化可以看作是删除和添加的复合,所以该操作的波及效应就是前两种基本操作的叠加.

从上面可以看出,本体中节点的波及效应与演化传播模型中对应节点的波及度和影效值有关,进而说明本体内聚度与演化传播模型中所有节点的波及度和影效值有关.节点波及度一定的情况下,节点影效值越大,该节点与所影响节点间的关联越密切;节点影效值一定的情况下,节点波及度越小,该节点与所影响节点的关联越密切.

设与本体 O 对应的演化传播模型中有 n 个节点,第 $i(1 \leq i \leq n)$ 个节点为 X_i ,则 X_i 与其所影响的节点间的关联度 $Rela(X_i) = \begin{cases} \frac{Eff(X_i)}{B(X_i)}, & \text{当 } B(X_i) \neq 0 \text{ 时} \\ 0, & \text{当 } B(X_i) = 0 \text{ 时} \end{cases}$,本体

$$O \text{ 的内聚度 } Cohesion(O) = \frac{1}{k} \cdot \frac{1}{n} \sum_{i=1}^n Rela(X_i), \text{ 其中 } k \text{ 为调和系数,取值为本体 } O \text{ 中节点个数与对应演化传播模型中节点的比值,本体 } O \text{ 中节点个数指该本体中概念个数、关系个数、基本属性个数及公理个数之和.}$$

本体内聚度从宏观的角度体现了本体中节点间联系的紧密程度,它能衡量本体变更的难易程度,进而衡量本体适合演化的能力.本体内聚性越大,本体越难修改,本体适应变化的能力越弱;本体内聚性越小,本体越容易修改,本体适应变化的能力越强.

6 实验结果与分析

本文的实验分为两部分.第一部分通过对一对相似本体 O_1 和 O_2 的内聚程度进行分析对比来判别本文提出方法的有效性;第二部分对本文提出的方法和文献[7]提出的方法做比较分析,进而说明本文提出方法的实用性.

6.1 有效性判别实验

本文通过一对相似本体 O_1 和 O_2 做有效性验证试验,本体 O_1 和 O_2 的结构分别如图 1 和图 2 所示.当前

关于本体内聚度的计算缺乏一种统一的标准,本文采取咨询本体专家的方法来解决这一问题.首先请若干位本体专家对这对相似本体的内聚度给出一个大致的评估,然后利用本文提出的方法分别计算这对本体的内聚度,看计算结果与专家意见是否一致.

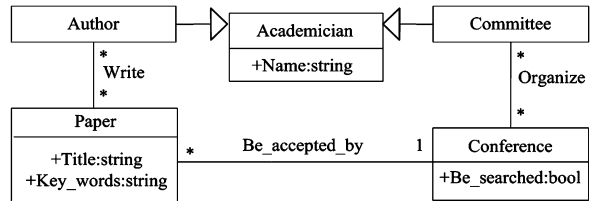


图1 本体 O_1 模型图

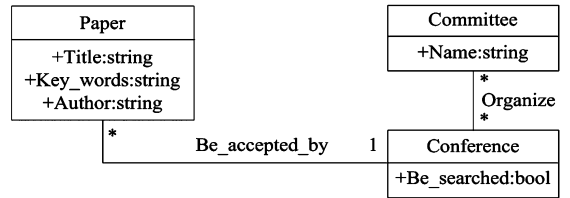


图2 本体 O_2 模型图

在做具体计算之前,本文请 5 位本体专家对本体 O_1 和 O_2 的内聚程度进行了比较,这些专家一致认为本体 O_2 的内聚程度高于本体 O_1 的内聚程度,本体 O_1 比本体 O_2 更适合本体演化.下面本文将给出计算本体 O_1 和 O_2 的内聚度的实验过程和结果.

设本体 O_1 和 O_2 所对应的演化传播模型分别为 G_1 和 G_2 ,表 1 和表 2 分别对 G_1 和 G_2 中的节点进行了 ETMC 编码和量化计算,计算了演化传播模型中节点的波及度、影效值及关联度.

表 1 G_1 中节点的 ETMC 编码与量化计算

Node	ETMC	$B(X)$	$Eff(X)$	$Rela(X)$
Academician	C0	9	3.67	0.41
Author	C1 C0	3	1.50	0.50
Committee	C2 C0	3	1.50	0.50
Paper	C3	5	5.00	1.00
Conference	C4	5	4.00	0.80
Academician attribute	PC0	9	3.17	0.35
Author attribute	PC1 C0	3	0.00	0.00
Committee attribute	PC2 C0	3	0.00	0.00
Paper attribute	PC3	5	5.33	1.06
Conference attribute	PC4	5	2.67	0.53
Write	R0 C1 C3	1	1.00	1.00
Organize	R1 C2 C4	1	1.00	1.00
Be_accepted_by	R2 C3 C4	1	1.00	1.00
Write restriction	PR0 C1 C3	1	1.00	1.00
Organize restriction	PR1 C2 C4	1	1.00	1.00
Be_accepted_by restriction	PR2 C3 C4	1	1.00	1.00

表 1 和表 2 表明,通过 $B(X)$ 的值可以分别看出本体 O_1 和 O_2 中实体的影响范围, $B(X)$ 越大,实体的影

响范围越大, $B(X)$ 越小, 实体的影响范围越小; 通过 $Eff(X)$ 的值可以分别看出本体 O_1 和 O_2 中实体的重要程度, $Eff(X)$ 越大, 实体越重要, $Eff(X)$ 越小, 实体越轻微; 通过 $Rela(X)$ 的值可以分别看出本体 O_1 和 O_2 中实体和与其相关联实体间联系的紧密程度, $Rela(X)$ 越大, 联系越紧密, $Rela(X)$ 越小, 联系越松散.

表 2 G_2 中节点的 ETMC 编码与量化计算

Node	ETMC	$B(X)$	$Eff(X)$	$Rela(X)$
Committee	C0	3	2.5	0.83
Paper	C1	3	4.50	1.50
Conference	C2	5	4.00	0.80
Committee attribute	PC0	3	2.67	0.89
Paper attribute	PC1	3	5.50	1.83
Conference attribute	PC2	5	2.67	0.53
Organize	ROCO C2	1	1.00	1.00
Be_accepted_by	R1C1 C2	1	1.00	1.00
Organize restriction	PR0CO C2	1	1.00	1.00
Be_accepted_by restriction	PR1C1 C2	1	1.00	1.00

表 3 显示了本文提出方法对这对相似本体内聚度的计算结果, 实验结果表明, 本体 O_1 的内聚度低于本体 O_2 的内聚度, 在描述能力等价的情况下, 本体 O_1 比本体 O_2 更适合本体演化. 这与之前专家评估的结果一致, 说明在本体描述能力等价的情况下, 本文提出演化传播模型在判断哪种本体更加适合演化的问题上有较好的效果.

表 3 本体 O_1 和 O_2 的内聚度比对

Ontology	k	Cohesion (O)
O_1	0.9375	0.7433
O_2	1.2000	0.8650

6.2 实用性比对分析

文献[7]对本体演化的波及效应提出了一种计算方法, 该文提出了一种本体图模型, 通过邻接矩阵和可达矩阵表示该模型, 凭借矩阵变换与运算对本体演化中的波及效应进行量化分析. 在本体规模较小时, 该方法可以有效的量化本体演化的波及效应, 但是在本体规模较大时, 该方法会带来空间灾难, 很难量化本体演化的波及效应. 相对而言, 本文提出的方法可以有效地减少在量化计算过程中使用的空间, 具有更好的实用性, 下面就文献[7]中的例子对两种方法进行比较说明.

文献[7]在讲述所提出的方法时用到的本体实例模型如图 3 所示, 该本体对应的邻接矩阵和可达矩阵分

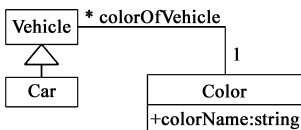


图 3 本体实例模型图

别如图 4 和图 5 所示, 本文提出方法中的量化计算依赖于 ETMC 编码, 设本体实例对应的演化传播模型为 Model, 则 Model 中节点的 ETMC 编码如表 4 所示.

$$\begin{bmatrix}
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0
 \end{bmatrix}$$

图 4 本体邻接矩阵

$$\begin{bmatrix}
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0
 \end{bmatrix}$$

图 5 本体可达矩阵

表 4 本体实例的 ETMC 编码

Model 中节点名称	ETMC 编码
Vehicle	C0
Car	C1C0
Color	C2
Vehicle attribute	PC0
Car attribute	PC1C0
Color attribute	PC2
colorOfVehicle	ROCO C2
colorOfVehicle restriction	PR0CO C2

从图 4、图 5 和表 4 可以看出: 按文献[7]提出的方法对本体实例做波及效应量化计算时需要 2 个 10×10 的矩阵, 即占用 200 个整数空间; 而本文提出的方法对本体实例做量化计算时需要对 8 个节点进行 ETMC 编码, 令编码中的每一位占用一个字符空间, 仅需要 34 个字符空间. 该结果表明, 本文提出的方法相对文献[7]中的方法具有较高的空间利用率. 如果问题规模增大, 本文提出方法的优势将更加明显. 与此同时, 在本体发生变化时, 文献[7]中的方法需要重新生成邻接矩阵和可达矩阵, 而本文提出的方法并不需要完全破坏原来的 ETMC 编码, 只需改变受变更节点影响的节点的 ETMC 编码. 通过上述分析, 可以看出本文提出的方法相对文献[7]中的方法更具实用性.

7 结束语

本文的研究将如何量化本体适应变化能力的问题

转换为本体内聚程度的计算问题,基于演化传播模型和 ETMC 编码给出了本体内聚度的计算方法,实验表明,本文提出的方法可以有效的衡量给定本体适应演化的能力.本文提出的演化传播模型不仅仅可以用来计算本体的内聚度,在未来的工作中,我们将考虑使用演化传播模型量化每种变更动作的影响,进而量化每种演化策略的影响,为本体以最优或较优的演化策略做变更提供理论指导.

参考文献

- [1] Haase P, Volker J, et al. Management of dynamic knowledge [J]. *Journal of Knowledge Management*, 2005, 9(5): 97 - 107.
- [2] Stojanovic L, Maedche A, et al. User-driven ontology evolution management[A]. EKAW2002[C]. London: Springer-Verlag, 2002. 2473:133 - 140.
- [3] Noy N F, Klein M. Ontology evolution: Not the same as schema evolution[J]. *Knowledge and Information Systems*, 2004, 6(4):428 - 440.
- [4] Haase P, Stojanovic L. Consistent evolution of OWL ontologies [A]. ESWC2005[C]. Crete: Springer-Verlag, 2005. 3532:182 - 197.
- [5] Li Qin, Atluri V. Evaluating the validity of data instances against ontology evolution over the Semantic Web[J]. *Information and Software Technology*, 2009, 51(1):83 - 97.
- [6] 明仲,蔡树彬,等.本体的继承及一致性分析[J].*电子学报*, 2005, 33(4):660 - 666.
Ming Zhong, LI Shi-xian, et al. Ontology inheritance and consistency analysis[J]. *Acta Electronica Sinica*, 2005, 33(4):660 - 666. (in Chinese)
- [7] 金龙飞,刘磊.一种本体演化波及效应分析方法[J].*电子学报*, 2006, 34(8):469 - 474.
Jin Long-fei, Liu Lei. A ripple-effect analysis method for ontology evolution[J]. *Acta Electronica Sinica*, 2006, 34(8):469 - 474. (in Chinese)
- [8] 刘晨,韩燕波,等. MINI 一种可减小变更影响范围的本体演化算法[J].*计算机学报*, 2008, 31(5):711 - 720.

Liu Chen, Han Yan-bo, et al. MIMI: an ontology evolution algorithm for reducing impact ranges [J]. *Chinese Journal of Computers*, 2008, 31(5):711 - 720. (in Chinese)

作者简介



刘磊 男,1960 年生于吉林长春.吉林大学计算机科学与技术学院教授,博士生导师.研究方向为软件形式化,语义网与本体工程.
E-mail: liulei@jlu.edu.cn



张鹏 男,1986 年生于吉林辉南.博士研究生,研究方向为语义网与本体工程.
E-mail: zhangpeng0521@mails.jlu.edu.cn



吕帅 男,1981 年生于吉林公主岭.吉林大学计算机科学与技术学院博士,讲师.研究方向为智能规划与自动推理.
E-mail: lus@jlu.edu.cn



张睿(通讯作者) 男,1979 年生于吉林德惠.博士,吉林大学计算机科学与技术学院讲师.研究方向为语义网、知识表示、访问控制.
E-Mail: rui@jlu.edu.cn