

面向模块重用的描述逻辑 $\mathcal{SHJF}$ 本地性规则研究

徐德智 廖晖寰 徐连君

(中南大学信息科学与工程学院 长沙 410083)

摘要 可重用本体模块的抽取是本体重用的一个关键环节。与传统工程应用中使用的基于本体层次的结构化方法抽取本体模块相比,使用逻辑的方法能充分利用本体提供的语义信息,抽取的本体模块更具完整性和正确性。在研究保守扩展的本体模块理论基础上,根据 Grau B C 提出的 $\mathcal{SHJF}$ 本地性规则,提出并证明了描述逻辑 $\mathcal{SHJF}$ 对应的语义本地性规则和句法本地性规则,为基于该规则抽取可重用本体模块提供了理论基础。

关键词 本体重用,模块抽取,描述逻辑 $\mathcal{SHJF}$,语义本地性,句法本地性

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.1.055

Research on Locality Rules of Description Logic $\mathcal{SHJF}$ for Module Reusing

XU De-zhi LIAO Hui-huan XU Lian-jun

(School of Information Science and Engineering, Central South University, Changsha 410083, China)

Abstract The extraction of ontology module is an essential step in ontology reuse. Compared to the structural approaches based on ontology hierarchy used in traditional engineering and applications, the logic approaches can make full use of the semantic information that ontologies provide. The extracted modules are more integrated and consistent. Based on conservative extension and locality rules of $\mathcal{SHJF}$ -based ontologies proposed by Grau B C, this paper provided and proved a SEMLOC rule and a SYNLOC rule for description logic $\mathcal{SHJF}$ -based ontologies, which can provide theoretical basis for extracting reusable ontology module.

Keywords Ontology resuing, Module extraction, Description logic $\mathcal{SHJF}$, Semantic locality, Syntactic locality

1 引言

Web 的创始人 Berners-Lee 于 1998 年提出了语义网的概念^[1]。语义网是面向信息文档表示的数据,是一种能够理解人类语言的智能网络。本体(Ontology)的研究是语义网技术的关键。本体是语义网知识表示和推理的基础,是由某一领域公认的符号序列(Signature,也称为词汇)和符号序列构建的术语之间的关系组成。

大规模本体的构建需要各个领域的多位专家合作。为了提高专家工作的独立性,常常采用软件工程的思想对本体进行模块化开发。近年来,随着语义网的快速发展,本体的数目也急剧增加。定义本体已很少从零开始,几乎都是先从现有本体中搜索需要的内容。因此本体模块的抽取和重用是语义网的发展趋势,也是目前本体研究的热点^[2]。

现有的模块抽取方法大多是使用结构化方法,如:Stuckenschmidt H 和 Klein M 提出的基于类层次结构的方法^[3]、Doran P 等人在类层次结构的基础上提出的抽象图模型的本体模块化方法^[4]、Seidenberg J 和 Rector A 提出的从 GALEN 中抽取相关概念的方法^[5]、Noy N F 和 Musen M A 提出的创建本体视图的方法^[6]。这些方法虽然执行效率高,但由于没有考虑本体的语义信息,抽取的模块都比较大,且不能保证一

致性和完整性,结果不甚理想。

描述逻辑 $\mathcal{L}$ (Description Logic, DLs) 是 OWL 本体语言的数学基础,是一种知识表示的形式化方法,是一阶谓词逻辑 (First Logic Language, $\mathcal{FOL}$) 的可判定子集^[7]。OWL 语言分为 3 种:OWL Full, OWL DL, OWL Lite 语言。因为在大多情况下,OWL Lite 的表达能力足以满足需求,并且相对于 OWL DL,它更易于用户的理解和开发工具的实现,因此本文讨论的重用对象是与描述逻辑 $\mathcal{SHJF}$ 表达能力一致的不含数据类型(对应于描述逻辑的有型域($\circ$)构造子)的 OWL Lite 本体。

基于描述逻辑的本体模块抽取方法通过语义推理,保留与符号序列相关的模型或蕴含关系得到模块,故其严格考虑到本体的语义,能保证抽取模块的正确性和完整性。目前具有代表性的方法有文献[8]提出的基于保守扩展理论的方法和文献[9]提出的基于不可分离关系的方法。

本文在文献[8]的保守扩展理论基础上,根据文献[10]中提出的 $\mathcal{SHJF}$ 本地性规则,提出相应的基于描述逻辑 $\mathcal{SHJF}$ 的语义本地性规则 SEMLOC 和句法本地性规则 SYNLOC,并证明 SEMLOC 规则是抽取本体模块的充分条件,再证明 SEMLOC 规则的正确性和可判定性,最后推理证明满足已处理的 SYNLOC 规则公理必定满足 SEMLOC。本文期望通过

到稿日期:2013-02-01 返修日期:2014-05-08 本文受国家自然科学基金面上项目(61379110),中南大学教师研究基金资助。

徐德智(1963—),男,博士后,教授,主要研究方向为语义 Web, E-mail: hunan_xu@mail.csu.edu.cn;廖晖寰(1988—),男,硕士生,主要研究方向为本体映射;徐连君(1988—),女,硕士生,主要研究方向为本体映射。

提出这些规则,得到新的本体模块抽取方法。

2 相关知识

2.1 描述逻辑基础

本文讨论的本体是基于描述逻辑 $\mathcal{SHF}$ 的 OWL Lite 本体,该 $\mathcal{L}$ 在基本的描述逻辑 $\mathcal{ALC}$ 的基础上,添加了传递角色 $\text{Trans}(R)$ 、角色包含 $\mathcal{H}$ 、逆角色 $\mathcal{J}$ 和函数角色 $\mathcal{F}$ 这些构造子。同时还定义了复杂角色 $R_{(i)}$ 和复杂概念 $C_{(i)}$,以及公理 $(\alpha, \beta, \dots)$ 。 $\mathcal{L}$ 是 $\mathcal{ALC}^+$ 的缩写。本体语言 OWL Lite 并不包含描述逻辑 $\mathcal{SHF}$ 的所有构造子,如概念的并 (unionOf)、补 (complementOf) 等,但是 OWL Lite 的表达能力和 $\mathcal{SHF}$ 是一样的,OWL Lite 本体的推理问题都可归约为 $\mathcal{SHF}$ 知识库的可满足性问题^[11]。

$\mathcal{SHF}$ 角色和概念的形式化定义如表 1 所列。

表 1 $\mathcal{L}$ 为 $\mathcal{SHF}$ 的描述逻辑层次结构

描述逻辑	构造子		公理	
	角色	概念	Tbox	Abox
$\mathcal{ALC}$	r	$\top, A, C_1 \sqcap C_2, \exists R. C, \neg C$	$A \sqsubseteq C, C_1 \sqsubseteq C_2$	$C(a), r(a, b)$
$\mathcal{J}$	同上	同上	同上, $\text{Trans}(R)$	同上
$\mathcal{H}$			$R_1 \sqsubseteq R_2$	
$\mathcal{F}$	r^-		$\text{Funct}(R)$	

其中, $\mathcal{SHF}$ 的其他构造子如 $\perp, C_1 \sqcup C_2, \forall R. C$ 都可以转化成 $\neg \top, \neg(\neg C_1 \sqcap \neg C_2), \neg(\exists R. \neg C)$ 的形式。 $C_1 \sqsubseteq C_2$ 是术语公理表达式,其中 C_1, C_2 是概念表达式; $R_1 \sqsubseteq R_2$ 是角色公理表达式, R_1, R_2 是角色表达式。

$\mathcal{SHF}$ 本体的语义可以用解释 $I = (\Delta^I, \cdot^I)$ 来表示,其中 Δ^I 是解释 I 的领域,由无限可数个体组成; $\cdot^I$ 是解释函数,表示个体、概念或角色到领域元素、领域子集或领域中二元关系的映射。

$\mathcal{SHF}$ 中的 Tbox 和 Abox 中的公理 α 的解释 I 的定义如下:

- (1) 若 $C_1 \sqsubseteq C_2$, 则 $I \models C_1 \sqsubseteq C_2$;
- (2) 若 $a^I \in C^I$, 则 $I \models C(a)$;
- (3) 对于 $\forall x, y, z \in \Delta^I$, 若 $\langle x, y \rangle \in r^I \wedge \langle y, z \rangle \in r^I \Rightarrow \langle x, z \rangle \in r^I$, 则 $I \models \text{Trans}(r)$;
- (4) 若 $R_1 \sqsubseteq R_2$, 则 $I \models R_1 \sqsubseteq R_2$;
- (5) 对于 $\forall x, y \in \Delta^I$, 若 $\langle x, y \rangle \in r^I \Rightarrow \langle y, x \rangle \in r^I$, 则 $I \models r^-$;
- (6) 对于 $\forall x, y, z \in \Delta^I$, 若 $x \in \Delta^I \wedge \langle x, y \rangle \in R^I \wedge \# \{y \in \Delta^I \mid \langle x, y \rangle \in R^I\} \leq 1$ 或者 $x \in \Delta^I \wedge \langle x, y \rangle \in R^I \wedge \# \{y \in \Delta^I \mid \langle x, y \rangle \in R^I\} \geq 2$, 则 $I \models \text{Funct}(R)$ 。

描述逻辑 $\mathcal{SHF}$ 的 Tableau 算法是通过构造概念 D 的完整树来证明 D 的可满足性^[10]。

2.2 本体模块重用

本体工程 O 在构建过程中,需要用到与符号序列 S 相关的知识,已知本体 O' 中已经定义了与 S 相关的概念和关系,那么将 O' 应用于 O 中就是本体重用。但是由于本体 O' 同时也包含了许多与 S 无关的内容,在 O 中重用整个本体 O' 不仅会造成本体 O 繁琐,难以理解,而且也增加 O 中公理冲突的概率。因此,从 O' 中抽取与 S 相关的本体模块是本体重用成功的必要阶段。

从 O' 中抽取的模块 M ,必须包含与 S 相关的所有信息,规模又必须尽可能小。由于本体 O 的工程师可能不熟悉 O' 领域的知识或者 O' 的规模非常大,他们无法判断抽取的模块公理是否会互相冲突,因此必须保证模块抽取算法抽取的所有公理具有一致性。

目前,逻辑领域对本体模块的定义借用了 Horrocks I 等人提出的保守扩展^[12,13]的概念:存在本体 $O', M \sqsubseteq O'$, 符号序列 S , 如果 O' 是 M 的模型 S -保守扩展,则称 M 是本体 O' 关于 S 的模块。由该定义可知,模型保守扩展可以作为模块抽取的充分条件,但是判定一个本体是否是另一个本体的模型保守扩展是个不可判定问题,所以为了使基于逻辑的模块抽取算法在实际应用中切实可行,本文使用依赖于重用符号 S 的基于本地性模块的定义。

3 本地性规则

3.1 基于本地性的本体模块

本地性这个概念最早出现在文献[8]中。若 S 表示符号序列, α 表示公理,如果 S 的平凡扩展 $S \cup \text{Sig}(\alpha)$ 的解释是公理 α 的一个模型,那么称公理 α 关于 S 是本地的,即 α 是 S 的本地公理。若 S 所有本地公理的集合表示为 $\text{local}(S)$, 如果对于本体 O , 有 $O \sqsubseteq \text{local}(S)$, 则称 O 关于 S 是本地的。

例 1 假设有公理 $K1: A \sqsubseteq B \sqcap \exists r. C, S = \{C, D\}$ 。令 S 的一个模型为 $I, S \cup \text{Sig}(K1)$ 的模型为 I' 。除了 A, B, r 之外, I 与 I' 是相同的。这里假设原子概念 A 和 B 以及原子角色 r 都没有出现在 I 中,它们都解释为空集。公理 $K1$ 左边和右边均为空,则 $I' \models S = I \models S$ 且 I 是 $K1$ 的模型,因此 S 的平凡扩展 $S \cup \text{Sig}(\alpha)$ 的解释 I' 满足 $K1$, 公理 $K1$ 关于 S 是本地的。

假设 $S = \{A\}$, $K1$ 关于 S 就不是本地的。因为对于 I 的任意平凡扩展 I' , 公理左边的 A 解释为非空集合,但右边解释为空集, I' 不满足 $K1$ 。

命题 1 (本地性 $\Rightarrow$ 模型保守扩展) 本地性是本体之间模型保守扩展关系的充分条件。

证明: (1) 若要证明以上命题,就要证明“对于本体 O, O' 和符号序列 S , 若 O 关于 $S \cup \text{Sig}(O')$ 是本地的, 那么 $O \cup O'$ 是 O' 的模型 S -保守扩展”, 也就是要证明“若 I 是 O' 的模型, 则 $O \cup O'$ 是 O' 的模型 S -保守扩展”, 亦即“ $O \cup O'$ 存在一个模型 I' 使得 $I' \models S = I \models S$ 成立”;

(2) 令 I' 是 I 关于符号 $S \cup \text{Sig}(O')$ 到 $S \cup \text{Sig}(O') \cup \text{Sig}(O)$ 的平凡扩展, 则关于 $S \cup \text{Sig}(O')$ 有 $I' = I$ 。那么: a. 关于 $\text{Sig}(O')$, 有 $I' = I$; b. 关于 S , 有 $I' = I$;

(3) 由于 O 关于 $S \cup \text{Sig}(O')$ 是本地的, 那么根据本地性的定义就有: I' 就是 O 的一个模型, 即 $I' \models O$;

(4) 综合 (2)a、(2)b 和 (3) 可得 I' 是 $O \cup O'$ 的模型, 且 $I' \models S = I \models S$; 因此 $O \cup O'$ 是 O' 的模型 S -保守扩展。证毕。

根据 2.2 节本体模块的定义可知模型保守扩展是模块的充分条件; 再结合命题 1 的“本地性 $\Rightarrow$ 模型保守扩展”可推导出本地性是模块的充分条件。

定义 1 (基于本地性的本体模块) 若有本体 $O', M \sqsubseteq O'$ 和符号序列 S , 如果 $O' \setminus M$ 关于 $S \cup \text{Sig}(M)$ 是本地的, 那么 M 就是 O' 关于 S 的一个模块。

定义 1 是基于本地性的模块形式化定义, 这是进行本体模块抽取的基础。

3.2 语义本地性规则

测试公理 α 关于符号序列 S 的本地性的充分条件是将符号序列 S 之外的原子概念和原子角色都解释为空,然后检查 α 中剩余的符号式子是否是可满足的。

本地性概念提出之后,经国外专家们的不断研究完善,Grau B C 等人在文献[10]中提出了 *SHF* 的语义本地性规则。由于本文讨论的是 OWL Lite 本体,本节给出相应的基于描述逻辑 *SHF* 的语义本地性规则,简称 SEMLOC 规则。这里使用 $\lambda(O', S)$ 函数表示本体 O' 关于符号序列 S 是本地的; $\lambda(\alpha, S)$ 表示公理 α 关于符号序列 S 是本地的; $\lambda(C, S)$ 表示概念 C 关于符号序列 S 是本地的。

规则 1 (语义本地性—SEMLOC (SEManicLOCality rules))

- (1) $\lambda(O', S) ::= \bigcup_{\alpha \in O'} (\alpha, S)$
- (2) $\lambda(\alpha, S) ::= \lambda(C_1 \sqsubseteq C_2, S) \mid \lambda(R_1 \sqsubseteq R_2, S) \mid \lambda(C(a), S) \mid \lambda(r(a, b), S) \mid \lambda(Trans(r), S) \mid \lambda(Funct(R), S)$
- (3) $\lambda(C, S) ::= \lambda(T, S) \mid \lambda(A, S) \mid \lambda(C_1 \sqcap C_2, S) \mid \lambda(\neg C_1, S) \mid \lambda(\exists R. C_1, S)$

其中,本地性的公理 α 和概念 C 具备以下性质:

- (a) $\lambda(T, S) = \top$
- (b) 若 $A \notin S$, 则 $\lambda(A, S) = \perp$, 否则 $\lambda(A, S) = A$;
- (c) $\lambda(C_1 \sqcap C_2, S) = \lambda(C_1, S) \sqcap \lambda(C_2, S)$;
- (d) $\lambda(\neg C_1, S) = \neg \lambda(C_1, S)$;
- (e) 若 $sig(R) \not\subseteq S$, 则 $\lambda(\exists R. C_1, S) = \perp$; 否则 $\lambda(\exists R. C_1, S) = \exists \lambda(R. C_1, S)$;
- (f) $\lambda(C_1 \sqsubseteq C_2, S) = \lambda(C_1, S) \sqsubseteq \lambda(C_2, S)$;
- (g) 对于 $\lambda(R_1 \sqsubseteq R_2, S)$, 若 $sig(R_1) \not\subseteq S$, 则 $\perp \sqsubseteq \perp$; 若 $sig(R_2) \not\subseteq S$, 则 $\exists R. \top \sqsubseteq \perp$; 否则 $R_1 \sqsubseteq R_2$;
- (h) $\lambda(C(a), S) = C(a)$;
- (i) 对于 $\lambda(r(a, b), S)$, 若 $r \notin S$, 则 $\top \sqsubseteq \perp$; 否则 $\lambda(r(a, b), S) = r(a, b)$;
- (j) 对于 $\lambda(Trans(r), S)$, 若 $r \notin S$, 则 $\perp \sqsubseteq \perp$; 否则 $\lambda(Trans(r), S) = Trans(r)$;
- (k) 对于 $\lambda(Funct(R), S)$, 若 $sig(R) \not\subseteq S$, 则 $\perp \sqsubseteq \perp$; 否则 $\lambda(Funct(R), S) = Funct(R)$ 。

规则 1(k) 中的角色 R 仅限于简单角色,不具传递性,也不能是传递性角色的子角色。因为若受函数约束的角色具有传递性属性,那么它的 Tableau 算法推理会非常复杂,可能把一连串传递的本体概念合并到一个结点,还有可能导致规则 1 不可判定。

命题 2 当且仅当 $\lambda(O', S)$ 是重言式时,本体 O' 关于 S 是本地的。

证明:(1)根据规则 1(3)即可消除符号序列 S 不包含的原子概念 A 和原子角色 r 。令 $J_{A \rightarrow \emptyset}^{\alpha}(S)$ 表示基于 *SHF* 的本体解释 $\mathcal{J}$ 的集合,其中 $\mathcal{J}$ 对于每个原子概念 $A \notin S$, 有 $A^{\mathcal{J}} = \emptyset$; 对于每个原子角色 $r \notin S$, 都有 $r^{\mathcal{J}} = \emptyset$ 。由此推导出 $J_{A \rightarrow \emptyset}^{\alpha}(S)$ 关于 S 是本地的。

(2)对于解释 $\mathcal{J} \in J_{A \rightarrow \emptyset}^{\alpha}(S)$, 当且仅当 $\lambda(\alpha, S)$ 是重言式时, $Sig(\alpha)$ 的解释 $\mathcal{J}$ 有 $\mathcal{J} \models \lambda(\alpha, S)$ 。对于解释 $\mathcal{J}$, 都有 $C^{\mathcal{J}} = (\lambda(C, S))^{\mathcal{J}}$ 且 $\alpha^{\mathcal{J}} = (\lambda(\alpha, S))^{\mathcal{J}}$; 并且当且仅当 $\mathcal{J} \models \lambda(\alpha, S)$ 时, 有 $\mathcal{J} \models \alpha$; 当且仅当 $\mathcal{J} \models \alpha$ 时, 公理 α 关于 S 是本地的;

(3)由(2)可知,当且仅当 $\lambda(\alpha, S)$ 是重言式时,公理 α 具有

本地性。由此亦很容易推导出“当且仅当 $\lambda(O', S)$ 是重言式时,本体 O' 关于 S 是本地的”。证毕。

Ontology of medical research projects 0;

P1 Genetic_Disorder_Project $\equiv$ Project $\sqcap \exists$ has_Fous. Genetic_Disorder

P2 Cystic_Fibrosis_EUPProject $\equiv$ EUPProject $\sqcap \exists$ has_Focus. Cystic_Fibrosis

P3 EUPProject $\sqsubseteq$ Project

P4 $\exists$ has Focus. $\top \sqsubseteq$ Project

Ontology of medical terms O' :

M1 Cystic_Fibrosis $\equiv$ Fibrosis $\sqcap \exists$ located_In. Pancreas $\sqcap \exists$ has. Origin. Genetic_Origin

M2 Genetic_Fibrosis $\equiv$ Fibrosis $\sqcap \exists$ has Origin. Genetic_Origin

M3 Fibrosis $\sqcap \exists$ located_In. Pancreas $\sqsubseteq$ Genetic_Fibrosis

M4 Genetic_Fibrosis $\sqsubseteq$ Genetic_Disorder

M5 DEFBI_Gene $\sqsubseteq$ Immuno_Protein_Gene $\sqcap \exists$ associated_With. Cystic_Fibrosis

图 1 医学术语的本体片段

例 2 本体 $O' = \{M1\}$, $M1$ 来自于图 1 医学本体的公理, $S1 = \{Cystic_Fibrosis, Pancreas\}$, $S2 = \{Fibrosis, has_Origin\}$

M1 Cystic_Fibrosis
 $\equiv \overbrace{\text{Fibrosis} \sqcap \exists}^{\perp(\text{规则b})} \overbrace{\text{located_In. Pancreas} \sqcap \exists}^{\perp(\text{规则e})}$
 $\overbrace{\text{has_Origin. Genetic_Origin}}^{\perp(\text{规则e})}$

根据规则 1(b) 和 1(e), $\lambda(M1, S1) = (Cystic_Fibrosis \equiv \perp \sqcap \perp \sqcap \perp)$ 不属于重言式; $M1$ 关于 $S1$ 不是本地的。

M1 Cystic_Fibrosis $\equiv \overbrace{\text{Fibrosis} \sqcap \exists}^{\perp(\text{规则b})} \overbrace{\text{located_In. Pancreas} \sqcap \exists}^{\perp(\text{规则e})}$
 $\exists \overbrace{\text{has_Origin. Genetic_Origin}}^{\perp(\text{规则b})}$

根据规则 1(b) 和 1(e), $\lambda(M1, S2) = (\perp \equiv \perp \sqcap \perp \sqcap \perp)$ 属于重言式; $M1$ 关于 $S2$ 属于本地的。因此本体 O' 关于 $S2$ 是本地的。

为了迎合一些本体工程的高效率要求,文献[12]提出语义本地性规则的同时,又提出了改进的 *SHF* 句法本地性规则。

3.3 句法本地性规则

由于基于 *SHF* 的本体执行 SEMLOC 规则时开销较大,本小节在 3.2 节 SEMLOC 规则的基础上,根据文献[10],介绍基于 *SHF* 的易处理的句法本地性规则 2,简称 SYNLOC 规则。

规则 2 (句法本地性—SYNLOC (SYNtacticLOCality rules))

S 表示重用符号序列,以下语法递归定义概念集合 $Con^0(S)$ 和 $Con^A(S)$ 。

(1) $Con^0(S) ::= A^0 \mid \neg C^A \mid C \sqcap C^0 \mid C^0 \sqcap C \mid \exists R^0. \top \mid \geq 2R^0. C \mid \geq 2R. C^0$

(2) $Con^A(S) ::= \top \mid \neg C^0 \mid C^A \sqcap C^A \mid$

其中, C 表示一般概念, R 表示一般角色, A^0 是原子概念且 $A^0 \notin S$, R^0 是原子角色 $r^0 \notin S$ 或其逆角色, $C^0 \in Con^0(S)$, $C^A \in Con^A(S)$, $C^A \in Con^A(S)$ 。

若公理 α 关于 S 是句法本地,则必须满足: (a1) $R^0 \sqsubseteq R$, (b1) $Trans(r^0)$, (c1) $Funct(R^0)$, (d1) $C^0 \sqsubseteq C$, (e1) $C \sqsubseteq C^A$,

$(f1)C^{\Delta}(a)$ 。

例3 本体 $O = \{M3\}$, 符号序列 $S = \{\text{Fibrosis}, \text{Pancreas}\}$, 应用规则2判断本体 O 是否是 S 的句法本地公理, 如下所示:

$$\underbrace{M3 \text{ Fibrosisi} \sqcap \exists \text{ located_In. Pancreas}}_{\substack{\in \text{Cm}^{\emptyset}(S) \text{ 规则 } (\exists R^{\emptyset}, C) \\ \in \text{Cm}^{\emptyset}(S) \text{ 规则 } (C \sqcap C^{\emptyset})}} \\ \underbrace{\sqsubseteq \text{Genetic_Fibrosis}}_{\in \text{Cm}^{\emptyset}(S) \text{ 规则 } (A^{\emptyset})}$$

应用规则2(1)得到 $M3 = \{\perp \sqsubseteq \perp\}$, 它属于句法本地公理。

命题3 句法本地公理集包含于语义本地公理集。

证明:(1)设满足 S 的句法本地性公理集为 $Ax'_{A \rightarrow \emptyset} \leftarrow \emptyset(S)$, 若 $O \sqsubseteq Ax'_{A \rightarrow \emptyset} \leftarrow \emptyset(S)$, 则称本体 O 满足句法本地性。

(2)由规则2的(1)和(2)对 $Con^{\emptyset}(S)$ 和 $Con^{\Delta}(S)$ 的定义可推导出, 对于 $I_{A \rightarrow \emptyset}^{\leftarrow \emptyset}(S)$ 中的每一个解释 $I = (\Delta^I, \cdot^I)$, 都有 $(C^{\emptyset})^I = \emptyset$, 且对于每一个 $C^{\emptyset} \in Con^{\emptyset}(S)$ 和 $C^{\Delta} \in Con^{\Delta}(S)$, 都有 $(C^{\emptyset})^I = \emptyset^I$ 和 $(C^{\Delta})^I = \Delta^I$ 。因此, 每一条句法本地性公理 α 都满足 $I \in I_{A \rightarrow \emptyset}^{\leftarrow \emptyset}(S)$ 。

(3)因此, 若根据3.2节 SEMLOC 规则1得到的公理集合为 $Ax_{A \rightarrow \emptyset}^{\leftarrow \emptyset}(S)$, 那么对于 S 就有 $Ax'_{A \rightarrow \emptyset} \leftarrow \emptyset(S) \sqsubseteq Ax_{A \rightarrow \emptyset}^{\leftarrow \emptyset}(S)$ 。

由命题3可知, 句法本地公理一定属于语义本地公理。但是, 语义本地公理并不一定是句法本地公理。

例4 有符号序列 $S = (C_1, C_2)$, 公理 $\beta = (C_1 \sqsubseteq C_1 \perp C_2)$ 是重言式, 所以它是语义本地公理。但它只涉及 S 的相关符号, 因此它不是句法本地公理。

从例4可看出, SYNLOC 规则不能识别每一条重言式, 但是由于在实际应用中, 不能被 SYNLOC 识别的重言式公理并不经常出现, 因此“句法本地公理集并不比语义本地公理集规模大很多”这样的假设是合理的。

根据规则2匹配每一条公理, 使用二进制编码约束函数数量, 则复杂度是多项式阶。若 $|O'|$ 和 $|S|$ 分别表示本体 O' 和符号序列 S 中出现的符号数目, 那么抽取本体 O' 的 SYNLOC 模块的复杂度是 $|O'| + |S|$ 的多项式。

结束语 本体重用对于快速构建高质量的本地具有十分重要的意义, 是目前本体领域研究的热点。与传统的使用结构化方法抽取本体模块相比, 使用描述逻辑的方法保证了抽取模块的完整性与正确性。本文首先证明本地性规则是基于逻辑的模块定义的充分条件, 即“若本体 $O' \setminus M$ 关于符号序列 $S \cup \text{Sig}(M)$ 是本地的, 那么 M 就是 O' 关于 S 的一个模块”; 然后在现有的 $SHQF2$ 本体的保守扩展和本地性研究基础上, 提出了适用于 $SHQF$ 本体的本地性规则——SEMLOC 规则和

SYNLOC 规则, 证明了 SEMLOC 规则是重言式的充分必要条件, 及较高效率的 SYNLOC 规则具备 SEMLOC 规则的所有特征, 且其复杂度为多项式阶。这些都为在基于 $SHQF$ 的本体中抽取本体模块提供了理论支持。

下一步工作将在本文提出的本地性规则基础上, 提出对应的本体模块抽取方法。

参考文献

- [1] Berners-Lee T. Semantic web road map[EB/OL]. <http://www.w3.org/DesignIssues/Semantic.html>
- [2] Shvaiko P, Euzenat J. Ontology matching: state of the art and future challenges[J]. Knowledge and Data Engineering, 2012, 25(1), 158-176
- [3] Stuckenschmidt H, Klein M. Structure-based partitioning of large concept hierarchies[M]// The Semantic Web-ISWC 2004. Springer Berlin Heidelberg, 2004: 289-303
- [4] Doran P, Tamma V, Iannone L. Ontology module extraction for ontology reuse: an ontology engineering perspective[C]// Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management. ACM, 2007: 61-70
- [5] Seidenberg J, Rector A. Web ontology segmentation: analysis, classification and use[C]// Proceedings of the 15th international conference on World Wide Web. ACM, 2006: 13-22
- [6] Noy N F, Musen M A. Specifying ontology views by traversal[M]// The Semantic Web-ISWC 2004. Springer Berlin Heidelberg, 2004: 713-725
- [7] Krötzsch M, Simancik F, Horrocks I. A description logic primer[J/OL]. <http://arxiv.org/pdf/1202.4089.pdf>
- [8] Grau B C, Parsia B, Sirin E, et al. Modularity and Web Ontologies[C]// KR. 2006: 198-209
- [9] Kontchakov R, Pulina L, Sattler U, et al. Minimal Module Extraction from DL-Lite Ontologies Using QBF Solvers[C]// IJ-CAI. 2009, 9: 836-841
- [10] Horrocks I, Sattler U. A Tableau Decision Procedure for [J]. Journal of Automated Reasoning, 2007, 39(3): 249-276
- [11] Horrocks I, Sattler U. A description logic with transitive and inverse roles and role hierarchies[J]. Journal of logic and computation, 1999, 9(3): 385-410
- [12] Grau B C, Horrocks I, Kazakov Y, et al. Modular reuse of ontologies: Theory and practice[J]. Journal of Artificial Intelligence Research, 2008, 31(1): 273-318
- [13] Armas Romero A, Cuenca Grau B, Horrocks I. Modular combination of reasoners for ontology classification[C]// 11th International Semantic Web conference. 2012: 1-16
- [31] Buhl H. A novel data quality metric for timeliness considering supplemental data[C]// Proceedings of the 17th European Conference on Information Systems. 2009: 2701-2713
- [32] Heinrich B, Klier M, Kaiser M. A procedure to develop metrics for currency and its application in CRM [J]. Journal of Data and Information Quality(JDIQ), 2009, 1(1): 5

(上接第248页)

- [29] Heinrich B, Kaiser M, Klier M. DQ metrics: a novel approach to quantify timeliness and its application in CRM[C]// ICIQ. 2007: 431-445
- [30] Heinrich B, Klier M. Assessing data currency — a probabilistic approach[J]. Journal of Information Science, 2011, 37(1): 86-100