

基于本体 DL 的语义推理研究

王金环, 李宝敏

(西安工业大学 计算机科学与工程学院, 陕西 西安 710032)

摘 要:随着 Internet 的迅速发展,语义 Web 作为新一代 Web 正倍受关注,因为语义表述与推理是语义网需要解决的关键问题。为了解决这一问题,文中在概述了本体语言、描述逻辑的概念和描述逻辑的原理的基础上,提出了一种新的方法:让本体表示语言与描述逻辑相结合,从而为语义推理的合理性和有效性提供了保证。文中重点介绍了通过对基于本体 DL 的语义推理机制进行阐述分析,较详细地探讨了在本体构建过程中,运用基于描述逻辑的语义推理进行一致性判定和相容性判定的重要性和实现技术,并以果品领域的本体构建为例,对构建过程中的推理判定进行了验证。

关键词:本体;描述逻辑;语义推理;相容性;一致性

中图分类号:TP301.2

文献标识码:A

文章编号:1673-629X(2009)11-0094-03

Study of Semantic Reasoning Based on Ontology Description Logic

WANG Jin-huan, LI Bao-min

(Institute of Computer Science and Engineering, Xi'an Technology University, Xi'an 710032, China)

Abstract:With the rapid development of internet, the semantic Web is being doubly paid close attention as new generation Web, because semantic interpretation and reasoning are the keys to be solved in the semantic Web. In order to solve this problem, based on the concept and the principle of ontology language and description logic foundation, have brought forward one kind of new method: let the body express language and describe that the logic combines each other, be that semanteme speculate rationality and validity have provided guarantee thereby. The principle of semantic reasoning about DL is analyzed, and the issues are discussed such as the importance and implemental technologies of consistent test and subsuming test based on DL reasoning in the process of ontology construction. Finally, the fruits domain's main body being taken as the example of ontology construction, the practical reasoning tasks about a constructs are performed and testified.

Key words:ontology; description logic; semantic reasoning; compatibility; consistency

0 引言

本体^[1](Ontology)的概念源于哲学。在计算机科学中,本体是共享概念模型的明确的形式化规范说明^[2]。这一概念包含四层含义:

(1)概念模型:通过抽象出客观世界中一些现象的相关概念而得到的模型,概念模型表现的含义独立于具体的环境状态;

(2)明确:概念和概念的约束都有明确的、无歧异的定义;

(3)形式化:本体能通过本体语言编码,使计算机可读,且可以被计算机处理;

(4)共享:本体体现的是共同认可的知识,反映的是相关领域内公认的概念集。本体的目标是捕获相关领域的公有知识,提供对该领域知识的共同理解,确定该领域内共同认可的术语,并从不同层次的形式化模型上给出这些术语和术语间相互关系的明确定义,实现对领域知识的推理。

1 描述逻辑

描述逻辑(Description Logic, DL)是基于对象的知识表示的形式化,它称为概念表示语言或术语逻辑,是一阶谓词逻辑的一个可判定子集。描述逻辑上的推理基于知识库(Knowledge Base, KB)。在描述逻辑里知识库被分成两个部分 $KB(T, A) = \langle TBox, ABox \rangle$ 。TBox(Terminology Box)引入应用领域中的概念,是描述概念、概念间的关系,以及关系与关系之间的公理集合^[2]。例如,定义 Happy-Father 'Man $\cap \exists$ has-child. Female $\cap \exists$ has-child Male $\cap \forall$ has-child Profess'。

收稿日期:2009-03-23;修回日期:2009-06-27

基金项目:国家“星火计划”项目(2004EA850069)

作者简介:王金环(1979-),女,山东菏泽人,硕士研究生,研究方向为计算机网络与语义网;李宝敏,教授,硕士生导师,研究方向为计算机系统结构、计算机网络与语义网。

ABox(Assertion Box)为断言集合,分为两类:概念断言和关系断言。概念断言判断某一个体对象是否属于某概念,如 John:Happy-Father;关系断言判断个体对象之间是否满足某种联系,如<John,Mary>:has Child。在描述逻辑中,推理机制建立在 KB(T,A)上,有以下定义:

定义 1 知识库 $KB = TBox + ABox$,记为 $KB(T, A)^{[3]}$;

定义 2 知识库解释 $I: I = \langle \Delta^I, I \rangle$,其中, Δ^I 为论域, $KB(T, A)$ 中的概念集是 Δ^I 的子集, $KB(T, A)$ 中的关系集是 $\Delta^I \times \Delta^I$ 的子集, $KB(T, A)$ 中的任意个体是 Δ^I 中的元素。 I 为解释函数^[4]。

命题 1 解释 I 满足 TBox T , 当且仅当 I 满足概念 T 中所有的概念定义及公理, 称 I 为 T 的一个模型 (Model)^[5]。

1.1 TBox 上的推理

TBox 上的推理工作主要是检测概念的可满足性和包容性。可满足性又称为一致性,从逻辑观点来看,一致性查验概念 C 的定义是否符合逻辑,即是否存在 T 的一个模型 I ,使得概念 C 在解释 I 下不为空。包容性检测用于概念之间的归属关系的判定。在描述逻辑中,推理任务有明确的定义,设 T 为 TBox。

命题 2 可满足性 (Satisfiability) 称概念 C 关于 T 是可满足的, 当且仅当存在 T 的一个模型 I , 满足 $C^I \neq \emptyset$ ^[4]。

命题 3 包容性 (Subsumption): 若对 T 任意的模型 I 均满足 $D^I \supseteq C^I$, 则关于 T 有 $D \supseteq C$, 即概念 D 包容概念 C , C 为 D 的子类^[6]。

命题 4 相等性 (Equivalence): 若对 T 任意的模型 I 均满足 $C^I = D^I$, 则关于 T 有 $C \equiv D$, 即概念 C 与概念 D 相等^[3]。

命题 5 不交性 (Disjointness): 若对 T 任意的模型 I 均满足 $C^I \cap D^I = \emptyset$, 则 $C \neq D$, 即概念 C 与概念 D 不交^[6]。

文献[6]进一步证明了 DL 上的推理均可归结为包容检测或可满足性检测。因此,一般情况下,DL 的推理机制以相容检测和可满足性检测为核心,实现多种推理任务。

1.2 ABox 上的推理

ABox 上的推理体现在对 TBox 的一致性验证上。在 $KB(T, A)$ 系统中, ABox 中的命题往往是形如 $C(a)$ 或 $R(a, b)$ 的针对概念和关系的断言。这样的知识断言必定满足 TBox 中的规则设定, 否则推出不合逻辑的结论。如 TBox T 中定义了“ $Mother \cap Father = \emptyset$ ”, 即概念 $Mother$ 和 $Father$ 不交, 若 ABox A 中出现了断言“ $Mother(李明)$ ”和“ $Father(李明)$ ”, 则系统必得出

断言不一致的结论, 知识的描述中存在矛盾, 给出 $KB(T, A)$ 一致性的形式化定义。

定义 3 若存在一个既为 ABox A 又为 TBox T 的解释 I , 则 ABox A 关于 TBox T 是一致的^[6]。

ABox 上最重要的推理任务为实例检测。设 α 为 ABox A 中的断言, 若对 $KB = (T, A)$ 的任意模型, α 都是可满足的, 记作“ $A \vdash \alpha$ ”, 以概念断言为例, 概念断言为 $C(a)$, 实例检测问题即是判断 $KB(T, A)$ 的实例 a 与概念 C , 对知识库的任意解释 I 是否都满足 $a \in C^I$ 。“ $A \vdash C(a)$ ”问题通常归结为对 ABox A 的一致性判定, 有以下关系存在:

命题 6 $A \vdash C(a)$ 当且仅当 $A \cup \{\neg C(a)\}$ 是不可满足的。

对 $\{\neg C(a)\}$ 可满足性的判断等同于对概念的一致性判定问题。

命题 7 对任意概念 C , $\{C(x)\}$ 是可满足的, 其中 x 为任意的个体。

由于 $KB \langle T, A \rangle$ 上的所有推理问题都能够归结为可满足性验证。因此, 研究一致性验证的算法成为实现 DL 推理机制的关键。

2 语义推理

由上述分析可看出, 描述逻辑的推理主要体现在相容判定和一致性判定。在本体构建过程中, 相容判定主要用于概念的自动分类, 分类推理就自动把新概念插入层次结构的合理位置, 从而提高本体设计效率, 减少出错率。一致性检测用于查验概念定义是否存在矛盾, 检验知识描述的正确性。从而帮助设计人员尽快找出本体设计中存在的问题, 确保知识描述的正确性, 在本体的设计及维护阶段, 这一推理是非常重要的。所以, 本体离不开描述逻辑的智能推理。在设计阶段运用推理可以帮助检测概念定义中的冲突, 推导出概念间隐含的层次关联。维护阶段利用“实例检测”查验知识的可满足性, 对知识库中追加的新知识予以检测, 可确保更新知识的正确性^[7]。

3 基于本体 DL 的语义推理的实现

3.1 Apple-Onto 描述

笔者定义的苹果本体 Apple-Onto 由一个三元组^[8]表示 (ConTypeS, ConAssS, ConTypeH)。其中, ConTypeS 为本体的概念类集, ConAssS 为概念类关联集, ConTypeH 为概念类抽象层次结构, 如图 1 所示。Apple-Onto 中的概念关联表达了概念间的语义关系。

用描述逻辑的形式语言描述该本体定义的两个重要的概念:

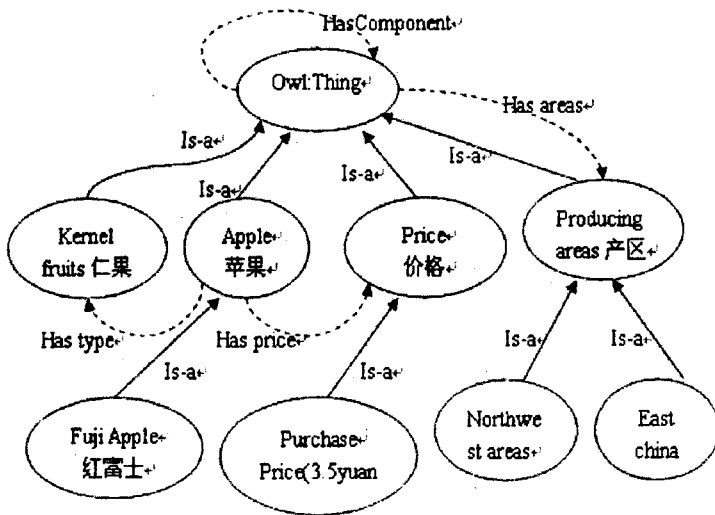


图 1 Apple-Onto 概念关联

$Apple \equiv T \cap (\leq 1 \text{ has Kernel fruits})$ “最多属于一个 Kernel fruits”, $Fuji \text{ Apple} \equiv Apple \cap (\exists \text{ has Price. Purchase Price} \cap \forall \text{ has Producing areas. East china areas})$ “至少包含一个 Purchase Price 以及一个值为 East china areas 的 Fuji Apple 的一类苹果”。本体 Apple-Onto 以 OWL 语言描述, 存放至文件 owl Apple Onto. owl 文件中, 为知识库中的 TBox。概念实例放在 owl Apple Data. xml 文件中, 即知识库中的 ABox。如在 ABox 中有以下关于实例的断言: “Kernel fruits (011)”, “Purchase Price (3.5yuan)”, “Apple (Fuji Apple)”, “has Price (3.5yuan, Fuji Apple)”等。

3.2 相容性判定

相容性判定用于概念的自动分类。假设在 TBox 中增加一个概念 New Type Apple, 描述逻辑的形式化定义为: $New \text{ Type Apple} \equiv T \cap (\forall \text{ has Kernel fruits. 011}) \cap (\forall \text{ has Producing areas}) \cap (\forall \text{ has Price 3.5yuan})$, 即该果子的属性是 Kernel fruits 为 011, Price 为 3.5yuan, 另一个组件为 East China areas。针对该概念进行相容性判定, 即检验新定义的 New-Type Apple 与其他概念的层次关联。利用 Jena 实现, 推理结论见表 1。

表 1 关于 New Type Apple 的 Jena 实现

输出结果:

```
New Type Apple * :
- (eg: New Type Apple rdf:type rdfs:Class)
- (eg: New Type Apple rdfs:subClassOf owl:Thing)
- (eg: New Type Apple rdfs:subClassOf eg: Fuji Apple)
- (eg: New Type Apple rdfs:subClassOf eg: Apple)
...
```

由结论可看出, 虽然在定义中并未声明该概念与其它概念间的相容关系, 但基于描述逻辑 TBox + ABox 上的推理结论是 New Type Fuji Apple 为 Fuji Ap-

ple 的子类, 在本体的概念层次结构中, New Type Apple 将自动插在概念 Fuji Apple 的叶子结点位置。相容性判定发掘出本体概念间隐含的层次关联。

3.3 实例认证

查询“所有的 Fuji Apple”。即在元数据库中查找所有可归属为“Fuji Apple”的苹果类。设 ABox 有一新品种的红富士 qiufu1 号 (见表 2)。qiufu1 号被声明为一类苹果, 拥有一个合适的 Price “3.5yuan”。结合 has Component 的传递性, 以及 has Kernel fruits 为 has Component 的子属性推知它同时拥有一个 “Producing areas” 的组件, 可以确认 qiufu1 号属于 Fuji Apple。因此, 除了检索出声明为 Fuji Apple 的苹果, 通过推理也发掘出隐含的所属关系 “qiufu1 号 \in Fuji Apple”。

表 2 关于 qiufu1 号的断言声明

```
Purchase Price(3.5yuan)
Apple (qiufu1 号)
has Kernel fruits (qiufu1 号, unknown MB)
has Producing areas (unknown MB, East china areas)
has Price(qiufu1 号, 3.5yuan)
```

3.4 数据一致性检测

利用推理进行一致性检测, 可以验证知识表达的正确性。设 ABox 中描述了两个同名 Name023, 不同价格的苹果 (见表 3)。在未声明情况下, 推理机制自动认定同名资源完全一致。因此, Name023 涉及的两种苹果应为同一种苹果, 即 “Name Special MIB \equiv 012”。但 TBox 中同时又声明 “Name Special MIB \cap 012 = \emptyset ”, 冲突由此产生。结果提示 (见表 4), 冲突产生的原因为不相交类的定义或函数属性的设置造成的。结合前面分析, 产生冲突是因为违背了 has Kernel fruits 的函数属性, 即一种水果只属于一种果类 (如 “仁果类”), “ $Apple \equiv T \cap (\leq 1 \text{ has - Kernel fruits})$ ”。

表 3 关于 Name023 的 OWL 描述

```
<Apple rdf:about="&eg; Name023">
<has Kernel fruits rdf:resource="&eg; Name Special MIB" />
<has Price rdf:resource="&eg; Name Special Price" />
</Apple>
<Apple rdf:about="&eg; Name023">
<has Kernel fruits rdf:resource="&eg; 012" />
<has Price rdf:resource="&eg; Name Special Price" />
</Apple>
<rdf:Description rdf:about="&eg; Name Special MIB">
<owl:differentFrom rdf:resource="&eg; 012" />
</rdf:Description>
```

在本体的设计构造阶段, 一致性检测将时时帮助查找设计中产生的各种错误。

(下转第 100 页)

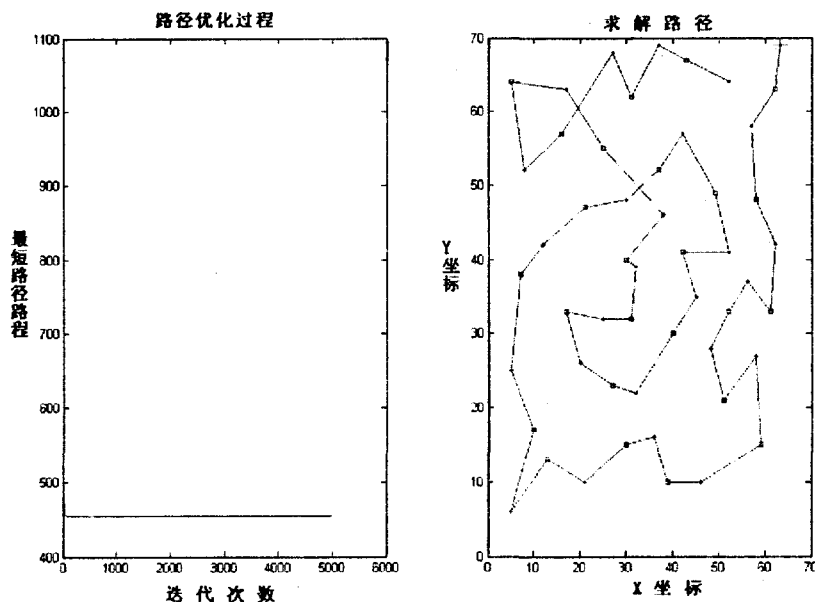


图 2 遗传算法计算的结果

4 结束语

遗传算法的主要特点是简单、通用、鲁棒性强,适用于并行分布处理,应用范围广。文中分析了遗传算法和模拟退火算法的基本原理,并将它应用于解决 TSP 问题,取得了较好的优化结果。通过实验得出遗传算法在求解 TSP 问题时,能够获得比模拟退火更好的答案,而且也比模拟退火算法的搜索速度要快,遗传算法的搜索过程是从一个母体到另一个母体的过程,使它能够有效放弃潜在的局部解,而用更加有效的算法

获得最优解。

参考文献:

- [1] Grefenstette J. Genetic algorithms for the traveling salesman problem [C]//in: proc. of 1 int. conf. on genetic algorithms and their applications. [s. l.]: Lawrence Erlbaum Association, 1985: 166-168.
- [2] 邢桂华. 用 MATLAB 实现中国旅行商问题的求解[J]. 微计算机应用, 2004(2): 218-222.
- [3] 李敏, 吴浪, 张开碧. 求解旅行商问题的几种算法的比较研究[J]. 重庆邮电大学学报, 2008(10): 624-625.
- [4] 高经纬, 张煦, 李峰, 等. 求解 TSP 问题的遗传算法实现[J]. 计算机时代, 2004(2): 19-21.
- [5] 宋丹, 傅明, 朱享荣, 等. 一种改进的遗传算法及其在 TSP 中的实现[J]. 微机发展(现更名: 计算机技术与发展), 2004, 14(6): 21-23.
- [6] 魏平, 徐成贤. 基于模拟退火算法优化分析与研究[J]. 装备制造技术, 2008(7): 1-3.
- [7] Yu Shiaw-Shian, Tsai Wen-Hsiang. A new thinning algorithm for gray-scales by the relaxation technique[J]. Pattern Recognition, 1990, 23(10): 1067-1077.
- [8] 许智宏, 宋勃, 董建波. 用蚂蚁算法和模拟退火算法解大规模 TSP 问题的研究[J]. 计算机工程与科学, 2008(10): 43-44.

(上接第 96 页)

4 结束语

本体在知识表示与知识推理方面发挥着重要作用, 本体表示语言与描述逻辑相结合, 通过对基于本体 DL 的语义推理机制进行阐述分析, 较详细地探讨了在本体构建过程中, 利用本体进行逻辑判定, 实现自动推理语义, 描述逻辑的作用十分重要。描述逻辑提供了完备高效的推理服务^[5]。其推理机制以归约判定和一致性判定为核心, 能够实现本体知识库上的多种推理, 在本体的构建阶段发挥了重要的作用。

表 4 一致性检测实例

```
//数据一致性检测
If ( ValidityTest(in Price)) System. Out .println ("ok");
检测结果:
冲突
-Error("conflict"): "Two individuals both same and different,
May be due to disjoint classes or price properties"
Culprit = eg: 012
Implicated node: eg: Name023SpeciaMIB
```

参考文献:

- [1] 刘琼, 李保敏. 一种果品领域本体库的构建方法[J]. 计算机技术与发展, 2009, 19(1): 197-203.
- [2] Berners-Lee T. The Semantic Web[J]. Scientific American, 2001(6): 1-6.
- [3] 李文斌, 刘椿年. 对 OWL 及其逻辑基础的研究[J]. 计算机应用, 2004, 24: 146-148.
- [4] Baader F. Handbook of Description Logic[M]. Cambridge: Cambridge University, 2003.
- [5] 聂卉, 龙朝晖. 描述逻辑语义推理机制的应用研究[J]. 信息检索技术, 2006(11): 61-64.
- [6] Noy N F, McGuinness D L. Ontology development: a guide to increasing your first ontology[R]. USA: Standford University, 2001.
- [7] 张娜, 李宝敏. 语义检索及其关键技术研究[J]. 计算机技术与发展, 2006, 16(11): 22-25.
- [8] 浓国海, 穆斌, 胡学钢. 语义 Web 本体及本体库系统设计技术[J]. 微机发展(现更名: 计算机技术与发展), 2004, 14(7): 78-81.