

一种基于解释的知识库综合

端义锋, 胡谷雨, 潘志松

(解放军理工大学指挥自动化学院 南京 210007)

【摘要】从知识库的解释出发,对概念和概念间关系的解释进行了分析,定义了知识库系统的最小概念集合,设计了生成最小概念集合的方法,提出了基于解释的知识库综合算法,讨论了该算法在有限域上的可计算性。

关键词 综合; 知识库; 解释; 描述逻辑

中图分类号 TP182 文献标识码 A

Interpretation-Based Integration of Knowledge Bases

DUAN Yi-feng, HU Gu-yu, PAN Zhi-song

(PLA University of Science and Technology Nanjing 210007)

Abstract Information integration is a valuable research area, which is widely used in distributed databases, cooperative information systems and data warehouses. Based on the interpretation of knowledge bases, interpretations of concepts and the relation of these concepts are analyzed in detail, and the minimal concept set of knowledge bases is defined. Then a method of building minimal concept set is given, and an algorithm of interpretation-based integration of knowledge bases is devised. Finally, complexity of the algorithm in finite domain is discussed.

Key words integration; knowledge bases; interpretation; description logic

随着信息技术的发展和信息系统的广泛使用,人们越来越关注如何有效利用现有的信息系统以减少不必要的重复开发。这种关注不仅注重于对信息系统的模块化构造和综合利用,还注重于对信息本身的重用或综合。信息的重用或综合是一个重要的技术研究领域,广泛运用于分布式数据库、信息协作系统和数据仓库等实用产业^[1]。另外,与信息综合技术相关的研究还包括本体重用技术和知识共享技术^[2-5]。

在信息综合的过程中,必须考虑的两个关键问题是如何进行领域概念的建模以及综合信息如何提供对推理的支持^[1]。为了能够有效地支持信息推理,通常采用形式化语言来描述领域中的概念信息。在最近的研究中,人们对形式化语言研究的注意力集中到了描述逻辑(Description Logic, DL)上。以描述逻辑为语言建立起来的知识库由两个部分组成:Tbox和Abox。其中Tbox是由关于概念和概念间关系的断言构成的;Abox是由关于特定对象的断言(实例断言)构成的^[7]。在基于描述逻辑的信息综合方面,目前的研究主要集中在Tbox的综合上,即概念的综合和概念间关系的综合,本文的讨论也仅限于对Tbox的综合。

在对Tbox的综合上,文献[6]提出了可将多个独立知识库合并的合成操作方法。首先根据知识库之间的概念映射建立了换名操作方法,并给出了知识库的完备化操作,在此基础上定义了知识库合成操作方法,但并未说明如何获得知识库之间的概念映射。文献[1]从系统的角度分析了用于信息综合的描述逻辑框架,并将信息综合系统分成了三层:概念层,逻辑层和物理层,但未给出具体的算法。

收稿日期:2004-01-12

作者简介:端义锋(1974-),男,在职博士生,讲师,主要从事网络管理,人工智能,知识发现等方面的研究;胡谷雨(1963-),男,博士生导师,主要从事计算机网络,网络管理,人工智能等方面的研究;潘志松(1973-),男,博士后,主要从事神经网络,模式识别,入侵检测等方面的研究。

在基于本体的信息综合方面,文献[4]大多根据概念名称和描述之间的语言学分析,向用户提交信息综合的建议,并通过与用户之间的不断交互来逐步完成信息综合的目标,整个过程需要用户的不断参与。文献[3]通过引入形式化概念分析(Formal Concept Analysis, FCA)建立了多本体的概念格,从而确立多本体之间的概念关系,但仅限于满足FCA定义的概念,而不是知识库中的所有概念。

1 描述逻辑语言AL

AL语言是描述逻辑语言中最简单、最基础的一种语言。可以通过引入新的符号不断地对AL进行扩充以增强其描述能力。典型的扩充有ALC, ALCQ, ALCQI等^[7]。但是,由于本文关注的是概念和概念间关系的解释,与具体的语言元素及其描述能力无关,因此仅考虑AL这种最简单的描述逻辑语言。

AL语言由(Con, Rep)构成,其中 Con 是AL语言中的基本概念集合, Rel 是 Con 上的二元关系集合。在AL语言中,概念 C 的语法为^[7]: $C ::= \top | \perp | A | \neg A | C_1 \cap C_2 | C_1 \cup C_2 | \forall R.C | \exists R$ 。利用该语法,可以定义各种概念,它们的集合称为知识库。

给定AL的基本语言元素(基本概念及其间关系),可以给出AL的解释 I 为 (\mathcal{A}^I, \cdot^I) 。其中 \mathcal{A}^I 是一个非空集合,称为 I 的解释域, \cdot^I 为 I 的解释函数。对每个 $A \in Con$,有 $A^I \subseteq \mathcal{A}^I$;同样地,对每个 $R \in Rep$,有 $R^I \subseteq \mathcal{A}^I \times \mathcal{A}^I$ 。

根据 \cdot^I 的解释,得到任意的概念 $C(C \in Con)$ 在 I 下的解释 C^I ^[7]。概念 C 在 I 中的定义是由若干个AL语言元素依据AL的概念定义语法组成的AL语言中的一个语句。根据 \cdot^I 的解释,该语句决定了 \mathcal{A}^I 中的每个元素是否具备概念 C 的性质(即是否属于 C^I),因此将概念 C 的定义语句看成是概念 C 的内涵。概念 C 在 I 下的解释 C^I 包含了所有 \mathcal{A}^I 中符合概念 C 的实例,可以将 C^I 看成概念 C 的外延。

2 基于解释的概念分析

根据知识库及其描述语言AL的一个解释 I ,定义 Con 中两个概念(或概念间关系)在解释 I 下的相等关系:

$$\begin{aligned} C_1 = C_2, & \text{ 此时 } C_1^I \subseteq C_2^I \\ C_1 = C_2, & \text{ 此时 } C_1^I = C_2^I, C_2 = C_1 \\ R_1 = R_2, & \text{ 此时 } R_1^I \subseteq R_2^I \\ R_1 = R_2, & \text{ 此时 } R_1^I = R_2^I, R_2 = R_1 \end{aligned}$$

根据上述“=”的定义,在 Con 中,由两个完全不同的语句所定义的概念(或概念间关系) P_1 和 P_2 ,在解释 I 下可能会得到 $P_1^I = P_2^I$ 的结论。事实上可将 P_1 和 P_2 看成同类事物或关系所体现出的不同性质(即内涵)。

为了能够对 Con 中的每个概念建立一个统一的描述,希望能够得到 Con 中所有最小概念的集合 P^{min} 。该集合中不同概念的解释 P^I 之间两两不相交,并且 Con 中的任何一个概念 P 的解释 P^I 都能够写成: $P^I = P_1^I \cap P_2^I \cap P_3^I \cap P_4^I \dots$,其中, $P_i \in P^{min}, i=1, 2, 3, 4, \dots$ 。

为了得到 Con 的最小概念集合,给出生成最小概念集合的方法。该方法对于 Con 中的概念和概念间关系做如下操作:1) $P^{min} = \emptyset$; 2) 消除 Con 中的重复概念($P_1 = P_2$); 3) 如果 $P^{min} = \emptyset$,则算法结束;否则,任选 Con 中的概念 P ,将 P 从 Con 中删除,并转4); 4) 如果概念 P 满足对 Con 中的任意一个概念 P_j ,都有 $P^I \cap P_j^I = \emptyset$,则令 $P^{min} = P^{min} \cup P$,并转3); 否则转5); 5) 任取 Con 中一个概念 $P_j (P^I \cap P_j^I \neq \emptyset)$,将 P_j 从 Con 中删除,同时将 $P \cap P_j, P \cup P_j, P \rightarrow P_j, P_j \rightarrow P$ 加入到 Con 中,并转2)。

易证, P^{min} 为 Con 在 I 下的最小概念集合。

3 基于解释的知识库综合

知识库综合的主要目标是使我们能够以统一的方式对多个知识库中的概念进行访问和关联。知识库综合的方法主要可以分为两种:一种是通过协调器来访问多个知识库,该协调器中建有各知识库概念间的映射;另一种是将多个知识库中的概念综合后形成新的知识库。为了能够体现综合后的知识库与源知识库之间的关系,本文采用第一种方法。假设有知识库 K_1 和 K_2 ,建立它们之间的映射 M :

$$M \subseteq \{(C_1, C_2) | C_1 \text{ 并且 } C_2\} \cup \{(R_1, R_2) | R_1 \text{ 并且 } R_2\}$$

对知识库₁和₂做如下操作:1) 确定共同的解释域 Δ^I ;2) 建立₁和₂在 Δ^I 上的解释函数 \cdot^I_1 和 \cdot^I_2 , 构成相应的解释: $I_1 = (\Delta^I, \cdot^I_1)$, $I_2 = (\Delta^I, \cdot^I_2)$;3) 根据生成最小概念集合的方法得到₁和₂在 Δ^I 上的最小概念集合 P_1^{\min} 和 P_2^{\min} ;4) 根据概念间相等关系的定义得到 P_1^{\min} 和 P_2^{\min} 之间的映射 M^{\min} :

$$M^{\min} = \{(C_1, C_2) | C_1 \text{ 并且 } C_2 \text{ 并且 } C_1 = C_2\} \\ \cup \{(R_1, R_2) | R_1 \text{ 并且 } R_2 \text{ 并且 } R_1 = R_2\}$$

5) 根据 M^{\min} 很容易得到₁和₂之间的映射 M ,即只需对₁中的每个概念(或概念间关系) P 做如下处理:(1) 将 P 表达成 $P_1^I \vee P_2^I \vee P_3^I \vee P_4^I \dots$, 其中 $P_i \in P_1^{\min}$; (2) 对表达式中的每个 P_i , 如果存在 (P_i, P_i') M^{\min} , 则用 P_i' 代替概念表达式中的 P_i ; (3) 如果₂中某个概念(或概念间关系) P' 的表达式与变换后的 P 的表达式相同, 则把 (P, P') 添加到 M 中。

4 讨论

在提出的基于解释的知识库综合算法中, 知识库的综合过程被分割成三个阶段: 首先针对特定的解释域建立各知识库的解释, 然后根据该解释获取特定解释域上的各知识库的最小概念集合及最小概念集合间的映射, 最后在最小概念集合间映射的基础上建立各知识库间的映射。算法的第一阶段, 需要根据具体的解释域, 由知识库的解释器完成或手工完成。算法的后两个阶段的执行建立在第一阶段的基础之上, 并且是完全自动化的。在有限域模型下(即 Δ^I 中的元素为有限多个), 算法中₁和₂在 Δ^I 上的最小概念集合 P_1^{\min} 和 P_2^{\min} 能够在有限时间内找到, ₁和₂间的映射也能够有限时间内完成。

但是当 Δ^I 中的元素为无限多个时, 简单的依靠比较概念解释将导致算法无法在有限时间内完成。这种情况下, 一种可行的解决方法是简化知识库的解释域 Δ^I 。首先寻找 Δ^I 的一个有限子域, 在该子域上建立知识库之间的映射关系 M_0 。但是, 由于解释域的简化, 可能会导致 M_0 中出现错误的相等关系。此时需要人工参与, 找到这些错误的对应关系, 并针对性的扩大解释子域的范围, 再次计算映射关系得到 M_1 。这个过程反复执行, 就可以得到₁和₂间映射的序列 $\{M_0, M_1, M_2, \dots\}$ 。根据开放逻辑的研究成果^[8], 可以判定该序列收敛于₁与₂间的映射 M 。当知识库的解释域为无限域或数量巨大时, 可以采用一定的分割技术以加快映射序列的收敛速度。在以后的工作中, 将研究如何根据知识库中的概念定义来分析和制定分割的策略。

参 考 文 献

- [1] Calvanese D, Giacomo G D, Lenzerini M, et al. Description logic framework for information integration[C]. Proceedings of the 6th International Conference on Principles of Knowledge Representation and Reasoning (KR'98), Trento Italy, 1998
- [2] Pinto H S, Martins J P. Reusing ontologies[C]. Proceedings of AAAI 2000 Spring Symposium Series, Workshop on Bringing Knowledge to Business Processes, AAAI Press, 2000: 77-84
- [3] Stumme G, Maedche A. FCA-merge: bottom-up merging of ontologies[C]. Seattle, WA, U. S. : ProcIJCAI, 2001. 225-234
- [4] Natalya F N, Mark A. Musen. prompt: algorithm and tool for automated ontology merging and alignment[C]. ProcAAAI, Austin, Texas U.S. , 2000
- [5] Madhavan J, Bernstein P, Domingos P, et al. Representing and reasoning about mappings between domain models[C]. The 18th National Conference on Artificial Intelligence (AAAI'2002), Edmonton, Canada, 2002
- [6] Vitoria A, Mamede M. Integrating concept-based knowledge base[EB/OL]. <http://citeseer.nj.nec.com/282145.html>, 2003-11-02
- [7] Calvanese D, Giuseppe D G. Reasoning in expressive description logics(Chapter 12), handbook of automated reasoning[M]. Amsterdam, The Netherlands: Elsevier Science Publishers B.V.,2000
- [8] 李 未. 世纪之交的知识工程与知识科学[M]. 北京: 清华大学出版社, 2000. 197-225