

一种使能虚拟组织的本体动态构建机制

刘 晨^{1,2}

(1. 中国科学院计算技术研究所网络与服务计算研究中心, 北京 100080; 2. 中国科学院研究生院, 北京 100039)

摘要: 如何应对虚拟组织成员本体的变更问题是构建虚拟组织本体面临的主要挑战。该文提出一个半自动的本体构建机制 DODO, 通过动态捕获并吸收成员本体更新的领域知识来实现虚拟组织本体的不断演化完善。该机制还可辅助领域专家消除虚拟组织本体变更所产生的影响。实际应用表明, 使用 DODO 机制能够保证虚拟组织本体对变更成员本体领域知识的平均覆盖程度始终维持在一个较高的水平。

关键词: 虚拟组织; 本体构建; 本体演化

Dynamic Ontologies Construction Mechanism for Enabling Virtual Organizations

LIU Chen^{1,2}

(1. Research Centre for Grid and Service Computing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080;

2. Graduate University of Chinese Academy of Sciences, Beijing 100039)

【Abstract】 A challenging problem of constructing Virtual Organization(VO) ontology is how to deal with the problem of member ontologies variations. This paper proposes a semi-automatic mechanism called Dynamically cOnstruct Domain-specific Ontology(DODO), with which VO ontology can be evolved and consummated by continuously capturing and absorbing updated domain knowledge of VO members. This mechanism also can facilitate domain experts to eliminate the impacts of changing VO ontology. The practical application shows that the DODO mechanism can provide a high average degree of covering member ontologies domain knowledge for VO ontology.

【Key words】 Virtual Organization(VO); ontology construction; ontology evolution

1 概述

为实现开放环境下自发的资源共享和业务协同, 可以通过按需集成不同组织提供的分布自治的Web服务来形成组织之间的动态联盟, 即虚拟组织。服务之间的语义异构问题是集成服务所面临的主要挑战。构建领域本体为服务的功能和参数提供统一表示是解决语义异构问题的有效手段。然而在虚拟组织运行过程中, 虚拟组织成员可以自主加入和退出, 还可以修改并撤销已经发布的服务。这就导致难以预先构建一个全局稳定的领域本体。例如在某实际应用项目中, 某科技部门要打造中央级的科技资源集成门户, 允许各地方科技厅的服务提供单位自主加入资源共享。本体构建是实现该门户的一个核心任务。依赖本体, 平台可以按统一的方式接入和描述分服务提供者提供的各类服务。然而, 由于事先并不能确定有哪些服务提供者加入共享, 无法事先构建一个能满足所有服务提供者需要的全局本体, 因此集成门户的本体需要随着服务提供者的变化而不断演化^[1]。

本文提出了一种本体动态构建机制 DODO(Dynamically cOnstruct Domain-specific Ontology)来为虚拟组织的本体构建过程提供动态演化的能力。它能够从虚拟组织成员(服务提供者)的本体中动态发现新的领域知识, 并通过吸收这些知识促进虚拟组织本体的不断演化完善。此外, 它还负责消除本体变化所带来的影响, 维护本体与依赖服务之间的一致性。

2 工作原理

DODO 机制首先要求领域专家构建一个初始本体作为整个构建过程的起点。该本体包含了虚拟组织所面对的应用领域中最重要概念和属性。初始本体是不完全的, 它将通过

吸收成员本体知识而不断演化完善。在初始本体构建完成后, 本体的构建将进入动态演化阶段。该阶段大致可以划分为图 1 所示的 3 个步骤。

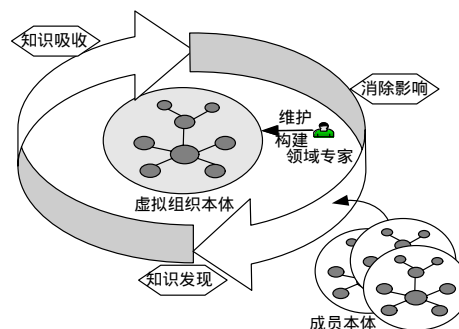


图1 DODO 机制工作原理

(1)知识发现。通过匹配成员本体和虚拟组织本体来发现它们之间的不同之处。在领域专家的辅助下, 不同之处将会以本体变更的形式表示出来。这些变更描述了虚拟组织本体为从成员本体中吸收领域知识所要经历的变化, 详见 3.1 节。

(2)知识吸收。顺序执行上一阶段所发现的变更以促进虚拟组织本体不断演化完善。但是变更的执行可能会破坏虚拟组织本体的正确性, 如删除一个正被其他实体引用的实体。

基金项目: 国家科技基础条件建设平台基金资助项目(2005DKA64201)

作者简介: 刘 晨(1980 -), 男, 博士研究生, 主研方向: 本体构建, 本体演化

收稿日期: 2007-11-15 **E-mail:** liuchen@software.ict.ac.cn

因此，本阶段的重点是如何保障变更前本体的正确性，详见 3.2 节。

(3)消除影响。本体的变更将被传播给依赖本体的服务以促使它们发生变化。变化的目的是使虚拟组织本体与依赖服务保持一致，详见 3.3 节。

3 核心技术

本节将以图 2 为例来介绍DODO机制的核心技术。该实例抽取自概述中所介绍的科技信息资源集成项目。此外，DODO机制采用KAON^[2]本体模型描述本体。

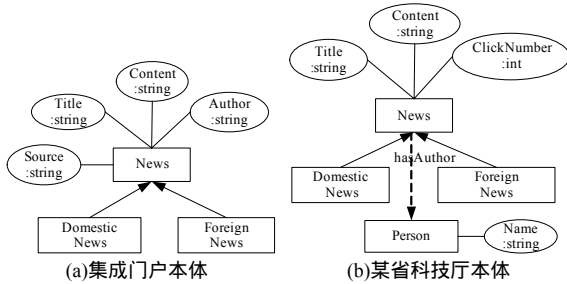


图 2 科技新闻本体片断

3.1 知识发现的方法与技巧

知识发现的目的是通过匹配成员本体和虚拟组织本体以发现应该被虚拟组织本体吸收的领域知识。以图 2 为例，通过匹配可以发现，成员本体比虚拟组织本体多提供了人员的详细信息(Person)。为了便于吸收，这些知识将被表示为本体变更，如 AddConcept(Person)。本体变更的详细定义如下：

定义1 一个本体变更操作被定义为一个函数 $OntCh(E)=E'$ ，其中 E 和 E' 分别是变更操作改变的本地实体集合及其改变后的结果。

知识的发现依赖于一个本体匹配算法。该算法以 2 个本体为输入，以本体实体之间的语义关系为输出。实体之间的语义关系包括相等(=)、包含(\subseteq)、包含于(\supseteq)、相交(\cap)以及无关(\perp)。参考全局词典WordNet^[1]，以下几种实体名称之间的词法关系被用来作为确定实体之间语义关系的依据：
 (1) Equal/Synonym：实体的名称完全相等或同义。这意味着相应实体间具有“=”关系。
 (2)Hypernymy/Hyponymy：实体的名称之间具有上位/下位关系。这意味着对应的实体间具有“ \subseteq ”/“ \supseteq ”关系。
 (3)Meronymy/Holonymy：实体的名称之间具有部分/整体关系。这意味着对应的实体间具有“ \cap ”关系。
 (4)Unknown：实体名称间不具有上述几种词法关系，意味着对应实体间具有“ \perp ”关系。根据实体间的语义关系，DODO将按照预定义的规则产生变更。部分产生规则如表 1 所示。其中，“ \circ ”表示 2 个变更之间的顺序执行关系。

表 1 变更产生规则(部分)

| 源实体类型 | 目标实体类型 | 语义关系 | 产生变更 |
|----------|----------|-------------|------------------------------------|
| Concept | Concept | \subseteq | AddSubConceptOf |
| Concept | Concept | \perp | AddConcept |
| Concept | Concept | \cap | RemoveConcept \circ AddConcept |
| Property | Property | \subseteq | AddSubPropertyOf |
| Property | Property | \perp | AddProperty |
| Property | Property | \cap | RemoveProperty \circ AddProperty |

3.2 知识吸收的方法与技巧

知识吸收阶段将依赖一个本体演化算法，通过逐个执行发现的变更来促进待构建本体的不断演化完善。实现本体演化的关键是在演化前后保持本体的一致性^[2]。该性质是指从本体的定义中无法推导出矛盾的结论。DODO机制首先把本体转化为一个图，然后把本体一致性定义为一个图约束的集

合。这样可以借助于已有的图算法对一致性进行判定与检测。本体的图模型被定义如下：

定义 2 给定本体 O ，它的概念集合为 C ，属性集合为 P ，实例集合为 I ，概念继承体系为 Hc ，属性继承体系为 Hp ，属性定义域为 $domain$ ，值域为 $range$ ，实例和概念之间关系为 $instconc$ 。那么该本体的图模型被定义为一个四元组 $\langle Nodes, Edges, src, tgt \rangle$ ，其中：

(1) $Nodes=C \cup P \cup I \cup L$ 。节点的名称被定义为实体的名称。 L 表示文本的集合。
 (2) $Edges=Hc \cup Hp \cup domain \cup range \cup instconc$ 。
 (3) src 和 tgt 定义了每条边的源节点和目标节点。它们的定义如下：如果 $e \in Hc$ ，那么 $src(e) \in C$ 且 $tgt(e) \in C$ ；如果 $e \in Hp$ ，那么 $src(e) \in P$ 且 $tgt(e) \in P$ ；如果 $e \in domain$ ，那么 $src(e) \in C$ 且 $tgt(e) \in P$ ；如果 $e \in range$ ，那么 $src(e) \in P$ 且 $tgt(e) \in C \cup L$ ；如果 $e \in instconc$ ，那么 $src(e) \in C$ 且 $tgt(e) \in I$ 。

根据文献[3]，本文把定义的图约束分为 3 类：(1)连接约束。这类约束是为了限制图中与某类节点相连的边的个数。例如，本体图中的任何一个除根概念(Thing)以外的概念节点都必须有一个父概念节点等。(2)模式约束。这类约束是为了限制本体图中不能出现某种模式，如路径和环。例如，概念和属性继承体系结构中均不能出现环。(3)禁止子图约束。这类约束是为了限制本体图中不能出现与预定义子图同构的子图。例如，数据属性节点的值域类型的边所连接的不能为概念节点等。

应用本体变更操作会导致旧的本体图消失以及新的本体图产生。因此，基于上述 3 类约束，本体演化过程被定义为一个图的变换过程，即从一个满足一致性约束的图开始并以一个满足变更需求和一致性约束的图终止。然而，单独执行一个变更操作可能会破坏本体的一致性。通过一致性检测，DODO 可以定位到违反不一致约束的实体集合。接着，在领域专家的帮助下，DODO 将产生附加变更操作来帮助本体回归到一个一致状态。

3.3 影响消除的方法与技巧

本体变化可能会对依赖本体的服务造成影响，引起它们的重新开发和重新部署。消除这种影响首先需要能够识别哪些服务会受影响，即确定变更的影响范围。一个服务是否受某个本体变更的影响是由变更的类型和服务所依赖的实体集合共同决定的。如果变更改变了服务所依赖的实体集合，那么服务将很有可能受到影响。此外，变更类型也是一个重要因素。并非所有的变更都会导致服务进行修改。在扩展型变更(在不影响其他实体的情况下对本体进行扩展的变更)和修改型变更(对本体的已有实体做出更新和修改的变更)中，只有修改型变更会导致服务的修改。

在确定变更影响范围后，就需要考虑消除影响的具体策略。这依赖于服务对本体的使用模式。本文区分 2 类使用模式：(1)外部标注模式，即采用本体实例的方式来为服务标注语义信息。对于这类模式可以把本体的实例和依赖的本体视为一个完整本体。然后参照 3.2 节中的演化技巧(产生针对实例的附加变更)加以解决。(2)内部使用模式，即服务依赖本体进行开发，在服务的内部实现代码中使用了本体信息。对于这种模式，在本体变化后系统将通知受到影响的服务，然后在人工的配合下消除影响。

4 相关工作

传统的本体构建方法可以分为 2 类：(1)从无到有的构建

方法。这类方法是由领域专家根据需求分析应用领域，然后从无到有构建本体。其代表项目有 CHEMICALS 及 Methontology 方法^[4]。(2)集成他人的构建方法。这类方法通过集成和修改已存在的本体来构建新的本体。其代表项目有 MOMIS^[5]。

这 2 类方法均没有为构建的本体提供足够的响应变化能力。例如，第(1)类方法需要领域专家手动分析并捕获变化。这导致变化响应慢且容易出错。而第(2)类方法则在有新的资源提供者出现时，需要重新对已有本体进行集成。这导致变化响应代价高。而且这 2 类方法均不能消除本体变化对依赖本体服务的影响。此外，现有本体演化工作^[2]关注于如何保障本体的一致性，而没有考虑如何通过分析他人的领域知识自动地捕获变更。这对虚拟组织的构建具有重要意义。因此，现有本体构建和演化工作均不能胜任为虚拟组织构建本体的任务。

5 评价

本文将结合前述的实际应用项目对 DODO 机制进行评价。如定义 3 所示，本文采用虚拟组织本体对每一个成员本体所包含领域知识的覆盖程度作为对 DODO 机制实际应用效果的评价指标。该指标本质上反映了所构建本体的完整性，程度越高，意味着越多的成员本体知识被发现并吸收。

定义 3 给定成员本体 O_s 和虚拟组织本体 O_t ，它们的实体集合分别为 E_s 和 E_t 。虚拟组织本体覆盖成员本体的程度(领域知识覆盖度)被定义为

$$\text{degree} = \frac{|E_s \cap E_t|}{|E_s|} \quad (1)$$

其中， E_s 和 E_t 的交集定义如下。注意，定义中相等关系的计算技巧已在 3.1 节中介绍。

$$E_s \cap E_t = \{e | e \in E_s \wedge \exists e' \in E_t \wedge e = e'\} \quad (2)$$

在当前项目的实现过程中，北京、山东以及四川省科技厅相继加入构建的科技资源集成门户。随着每个新部门的加入，在 DODO 机制的帮助下，科技资源集成门户的本体能够从这些部门的本体中吸收领域知识而不断演化完善。在这个

(上接第 25 页)

过程中，领域知识覆盖度的变化如表 2 所示。

表 2 领域知识覆盖度 (%)

| | DC(北京) | DC(山东) | DC(四川) | DC(平均) |
|------|--------|--------|--------|--------|
| 北京加入 | 100.0 | | | 100.0 |
| 山东加入 | 99.2 | 98.5 | | 98.9 |
| 四川加入 | 98.3 | 97.6 | 96.8 | 97.6 |

表 2 的实验结果表明，领域知识覆盖度一直维持在一个较高的水平(95%)。但是随着新部门的不断出现，领域知识覆盖度将会有所降低。这是因为不同资源提供者的领域知识之间存在矛盾，而虚拟组织本体仅能选择一方加以吸收，其他方可以通过建立映射来屏蔽异构问题。但从领域知识的平均覆盖程度来说，它的下降程度是非常有限的。这也就意味着，DODO 机制能够有效地动态集成来自不同资源提供者的领域知识，构建使能虚拟组织。

6 结束语

本文提出的领域本体动态构建机制 DODO 可为本体构建过程提供动态演化的能力。该机制不要求事先构建一个稳定的全局本体。它强调构建的本体在应用中通过不断吸收其他资源提供者的领域知识而不断演化完善，使得构建的本体是灵活可演化的，从而为虚拟组织的构建提供了必要保障。

参考文献

- [1] Miller A G. WordNet: A Lexical Database for English[J]. Communications of the ACM, 1995, 38(11): 39-41.
- [2] Stojanovic L. Methods and Tools for Ontology Evolution[D]. Karlsruhe: University of Karlsruhe, 2004.
- [3] Heckel R. Graph Transformation in a Nutshell[J]. Electronic Notes in Theoretical Computer Science, 2006, 148(1): 187-198.
- [4] Fernandez M, Gomez-Perez A, Juristo N. Methontology: from Ontological Art Towards Ontological Engineering[C]//Proc. of AAAI-97 Spring Symposium on Ontological Engineering. Palo Alto, CA, USA: [s. n.], 1997.
- [5] Beneventano D, Bergamaschi S. The MOMIS Methodology for Integrating Heterogeneous Data Sources[C]//Proc. of IFIP World Computer Congress. Toulouse, France: [s. n.], 2004.

4 结束语

GA 是进化计算研究的主要内容之一，它可以发现给定最优化问题的全局最优解。本文采用 GA 优化前馈神经网络，获得了较好的网络结构、参数及合适的激活函数和学习训练方法。由于个体编码长度较短，即解空间的维数较低，因此本文方法具有很好的操作性，适用于其他需要进行非线性映射问题求解的人工神经网络优化。

参考文献

- [1] Holland J H. Adaptation in Natural and Artificial Systems[M]. Boston, USA: MIT Press, 1975.
- [2] Goldberg D E. Genetic Algorithms: In Search Optimization & Machine Learning[M]. Massachusetts, USA: Addison-wesley Publishing Company, 1989.
- [3] 张讲社, 徐宗本, 梁怡. 整体退火遗传算法及其收敛充要条件[J]. 中国科学: E 辑, 1997, 27(2): 154-164.
- [4] 阎平凡, 张长水. 神经网络与模拟进化计算[M]. 北京: 清华大学出版社, 2005.
- [5] 陈小平, 石玉, 于盛林. 神经网络中克服局部最小的 BP-EP 混合算法[J]. 小型微型计算机系统, 2001, 22(12): 1460-1463.

表 1 GFNN 训练和 EFNN 训练的结果比较

| 网络 | MSE | Epoch | Time/s | E |
|------|----------|-------|--------|----------|
| GFNN | 9.740E-8 | 89 | 0.724 | 4.232E-6 |
| EFNN | 9.979E-8 | 117 | 0.941 | 4.815E-6 |

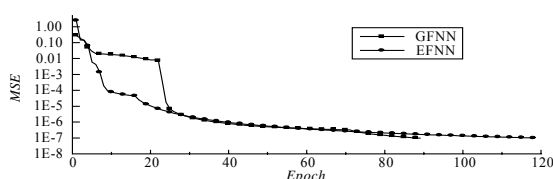


图 4 GFNN 训练和 EFNN 训练的收敛速度比较