# 基于语义网格的OWL-S服务本体的语义匹配研究*)

蒲国林[1,2] 杨清平[1,2] 邱玉辉[1] 王 刚[1] 葛继科[1]

(西南大学语义网格实验室 重庆 400715)[1] (四川文理学院 达州 635000)[2]

**摘 要** 研究了本体、本体匹配、NBC文本分类和OWL-S。OWL-S把网格中的资源组织为服务,并用服务本体来表示和描述,不但可以描述服务的语义,而且还能够进行适当的推理。针对OWL-S服务本体的异构性,利用OWL、OWL-S的元素值和文本内容,从本体结构、功能和文本信息等多个维度分析本体间的语义匹配问题,并给出了相应的语义等价匹配规则和基于NBC的文本分类式语义相似匹配算法,为语义网格中的服务本体共享、交互和集成等技术的实现提供了基础。

**关键词** 语义网格,朴素贝叶斯分类器(NBC),OWL,OWL-S,本体匹配

## Semantic Matching of OWL-S Service Ontology Based on Semantic Grid

PU Guo-Lin[1,2] YANG Qing-Ping[1,2] QIU Yu-Hui[1] WANG Gang[1] GE Ji-Ke[1]

(Semantic Grid Laboratory of Southwest University, Chongqing 400715)[1] (Sichuan University of Arts and Science, Dazhou 635000)[2]

**Abstract** Ontology, ontology matching, the NBC method and OWL-S are studied in paper. OWL-S organizes resources to be services in the grid, and it expresses and describes services with service ontology, not only the semantics of service, but also the suitable inference. In view of the heterogeneity of service ontology in OWL-S, we analyze semantic match problem between ontologies, with the element values and the text of OWL and OWL-S, from the dimensions of structure, function and the text information. It has given the corresponding semantic equivalent match rules and the approximate NBC-based text classification-type semantics match algorithm, and provides the foundation for the realizations of sharing, interaction and integration to the ontology in semantic grid.

**Keywords** Semantic grid, Naive Bayes Classifier(NBC), OWL, OWL-S, Ontology match

## 1 引言

在动态、异构和分布式的语义网格环境中,本体是语义知识表示的一个重要工具,它能捕获信息环境的结构和语义[10]。OWL-S中以服务本体的形式表示和描述服务,不但可以描述服务的语义,而且还能够进行适当的推理[9]。网格环境中本体的异构性使本体匹配成了本体共享、交互和集成的基础。同理,服务本体的匹配问题也就成了服务本体中语义共享、集成和聚融的关键。本文第2小节对本体、本体语义匹配和服务本体进行探讨;第3小节概述了朴素贝叶斯文本分类方法;第4小节研究OWL-S服务本体的语义匹配,给出相应的匹配规则或算法,并计算了本体匹配度。

## 2 本体、本体匹配和OWL-S服务本体

### 2.1 本体

#### 2.1.1 本体的定义和形式化

本体是一种知识表示,对本体(ontology)的定义还无统一标准,其中"本体是关于可共享概念化体系的明确的形式化规格说明"[1]的定义得到了普遍认同。该定义明确给出了本体的四个特征,即"共享","概念化","形式化"和"明确化"。

对本体的形式化表示也很多,其中应用较多的主要有四元素表示法和六元组表示法(Myo Myo Naing, 2002)。文[2~4]中的四元素表示形式的基本思想是:一个本体主要包括概念(concepts)、关系(relations)、实例(instances)和公理(axioms)四个元素,即:

$$O=\{C,R,I,A^O\} \quad (1)$$

其中,$C$表示概念集合,$c$表示概念,且$c\in C$;$R$表示关系集合,$r$表示关系,且$r\in R$;$I$表示实例集合,$i$表示实例,且$i\in I$;$A^O$表示公理的集合。

概念表示特定领域中一类实体或事物的集合,每个概念都有一到多个属性;实例是概念所表示的具体事物,实例继承了概念的属性,实例是领域概念化的结果,概念和实例是相对的;公理是对概念或实例的约束,以具体的约束规则的形式出现;关系是概念间的或属性间的关系,关系也有属性。本体描述的关系有part-of、kind-of、instance-of和attribute-of等。part-of表达概念间部分与整体的关系;kind-of表达概念间的继承关系;instance-of表示概念间是实例关系;attribute-of表示概念间是属性关系。在OWL中,用Class来描述这里的概念,用保留的RDF Schema特性rdfs:subClassOf来陈述类间的层次结构,用Individual来陈述个体是某个类的实例,用oneOf来枚举类的所有个体,用rdf:Property来陈述属性,用equivalentClass和disjointWith等描述概念间的等价和不相交关系,用inverseOf、TransitiveProperty和SymmetricProperty来描述属性间的逆、传递和对称关系等。

#### 2.1.2 本体的异构性

本体是领域知识概念化和模型化的方法,因此本体是本

*)受西南大学研究生科技创新基金资助(2006011)。**蒲国林** 博士研究生,主要研究领域为语义网格、服务计算;**杨清平** 副教授,访问学者,主要研究领域为人工智能;**邱玉辉** 教授,博导,主要研究领域为人工智能、语义网格、服务计算;**王 刚、葛继科** 博士研究生,主要研究领域为语义网格、服务计算。

体专家在领域知识专家的参与下共同建立的。由于词汇本身的一词多义和异形同义现象的存在，不同领域甚至同一领域使用了不完全相同的词汇集。这可借助语义词典（如 wordnet）在一定程度上实现词汇的语义共享。同时，对不同领域知识建立开放的权威的词汇集作为本体建立时的用词规范，这将有利于语义信息的共享。如网络资源上用的都柏林元数据元素集 Dublin core[5] 和出版业中的一个元数据规范 PRISM[6] 以及得到公认的 XML 命名空间（QNAME）等都促进了语义信息在领域内的共享。

本体异构性的另一方面就是表示形式的异构。这主要源于不同国家不同民族的历史文化背景。最典型的是日期时间和姓名等，它们都可用多种表示形式表达同一语义。形式上的异构可以通过形式转换的方法来消除，而不必对用户作特殊的要求。

### 2.2 本体语义匹配

本体匹配又称本体映射，是源本体（Source Ontology）元素集和目标本体（Target Ontology）元素集之间的一个有向对应关系[7]。用 $V_s$ 和 $V_d$ 分别表示源本体 $O_s$ 和目标本体 $O_d$ 中的具体元素，用 $\langle V,\prec\rangle$ 表示偏序集，它由集合 $V$ 和 $V$ 中的一个严格的偏序关系组成[10]。偏序集 $\langle V_s,\prec_s\rangle$ 到 $\langle V_d,\prec_d\rangle$ 的匹配记为 $m_{sd}$，是从 $V_s$ 到 $V_d$ 的特定函数。$m_{sd}$ 的逆匹配记为 $m_{ds}$。用 $\vdash$ 表示语义匹配关系或进行语义匹配的运算符[11,12]。本体间的语义关系匹配 $m_{sd}$ 是偏序集 $\langle V_s,\prec_s\rangle$ 到 $\langle V_d,\prec_d\rangle$ 的语义匹配，是在指定语义关系 $\vdash$ 下从 $V_s$ 到 $V_d$ 的函数。本体语义匹配模型图如图 1 所示。

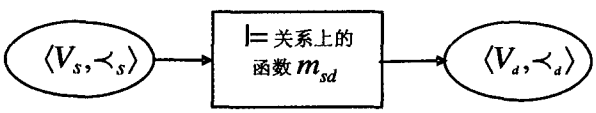


图 1 本体语义匹配模型图

### 2.3 OWL-S 服务本体介绍[9]

OWL-S 提供描述 Web 服务的词汇表，具有显式语义和机器可理解的特点，可以描述 Web 服务的属性和功能。OWL-S 的层次结构由一整套本体构成，所有的服务本体元素构成了一个以服务类为根结点的树。最上层是服务（Service）类，它有三个属性，即表示（presents）、被描述（describedBy）和支持（supports）。每个属性值对应着一个次高层的类，分别是服务简档（ServiceProfile）、服务模型（ServiceModel）和服务基点（ServiceGrounding）。

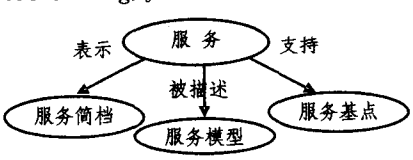


图 2 清晰地表示出了它们之间的关系

OWL-S 规范对图示关系的基数约束是：一个服务最多和一个服务模型相关联；一个服务基点必须和一个服务相关联；一个服务可有多个服务简档或者服务基点。

服务简档（ServiceProfile）主要提供与该服务相关的基本信息，具体包括以下三种基本类型的信息：服务提供者的相关信息；服务的功能及其他属性；描述服务特性的其他信息，如服务的名称、服务实例、服务模型的实例、服务的文字描述、服务提供者的联系信息、服务的分类信息、服务的一个输入输出、服务结果以及可扩展的属性列表等等。

## 3 朴素贝叶斯文本分类

贝叶斯方法的新实例分类目标是在给定描述实例的属性值 $\langle a_1,a_2,\cdots,a_n\rangle$ 下，从有限集合 $V$ 中得到最可能的目标值 $v_{MPA}$。

$$v_{MAP}=\arg\max_{v_j\in V}P(a_1,a_2,\cdots,a_n|v_j)P(v_j) \tag{2}$$

朴素贝叶斯分类器（Naïve Bayes Classifier，NBC）又叫朴素贝叶斯学习器，它以公式（2）为基础，并基于以下假设：

**假设 1** 条件独立性：指在给定目标值时属性值之间相互条件独立[8]。

在假设 1 下，公式（2）被化为更易于计算的公式（3）：

$$v_{NB}=\arg\max_{v_j\in V}P(v_j)\prod_i P(a_i|v_j) \tag{3}$$

使用假设 1 的贝叶斯学习方法就是朴素贝叶斯分类器（NBC）。NBC 可通过公式（3）计算出该实例是属于 $V$ 集中的哪一类。只要条件独立性能得到满足，NBC 的分类结果 $v_{NB}$ 等于 MAP 分类的结果[8]。特别地，当训练样例和新实例都是文本信息（如单词集）时，该分类就是文本分类。所谓文本分类是指根据文档的内容或属性将大量的文档归到一个或多个类别中的过程。用 $w_k$ 代表英文词典中的第 $k$ 个单词，则 NBC 文本分类用公式（4）等价计算。

$$v_{NB}=\arg\max_{v_j\in V}P(v_j)\prod_i P(a_i=w_k|v_j) \tag{4}$$

其中，$P(v_j)$ 可基于每一类在训练数据中的比例很容易地得到，而 $P(a_i=w_k|v_j)$ 难于计算。Lewis（1991）、Lang（1995）和 Joachims（1996）首先将 NBC 用于文本分类，并引入了以下三个假设以简化计算：

**假设 2** 位置无关性。指遇到一个特定单词的概率独立于单词所在位置。

**假设 3** 均匀的先验概率。如果某属性有 $k$ 个可能值，那么该属性取某个值的先验概率 $p=1/k$。

**假设 4** 新文本无关性。是指忽略在新文档中出现而在目标文档中没有出现的新单词对文本分类的影响。

在这些假设下，$P(a_i=w_k|v_j)$ 的 $m$-估计为：

$$\frac{n_k+1}{n+|Vocabulary|} \tag{5}$$

其中，$n$ 为所有目标值为 $v_j$ 的训练样例中单词位置的总数，$n_k$ 是在 $n$ 个单词中找到 $w_k$ 的次数，而 $|Vocabulary|$ 为训练数据中的不同单词的总数。

## 4 OWL-S 服务本体的语义匹配

### 4.1 语义等价匹配规则

语义匹配不仅是文本的匹配，结构和功能上的匹配也是语义匹配的具体方式。充分利用 OWL 和 OWL-S 本体中的一些属性能找到一些快速有效的匹配规则。语义匹配有语义等价匹配和语义相似匹配。文[13]中定义了一些等价匹配和相似匹配启发式规则，参考这些规则并结合 OWL-S 服务本体特性，特给出如下语义等价匹配规则。

#### 4.1.1 通过 OWL 的本体导入进行匹配

在 OWL 本体中，通过 owl:imports 能在该本体中导入其它本体，这增强了本体的复用性和共享性。

**规则 1** 如果两个本体相互导入，则它们拥有相同的元素集，所以它们是等价的[10]。

在 $O_s$ 中导入 $O_d$，显然有 $V_d\subseteq V_s$；同理，在 $O_d$ 中导入 $O_s$，则 $V_s\subseteq V_d$。如果 $O_s$ 和 $O_d$ 相互导入，显然 $V_s=V_d$，所以 $O_s\vdash O_d$ 且 $O_d\vdash O_s$，记为：$O_s\models O_d\models$，表示语义等价。

规则 1 用 IF - THEN 规则描述为：

IF importsEachOther($O_s$,$O_d$)

THEN $O_s \models O_d$

4.1.2 通过 OWL-S 的输入输出进行匹配

源服务本体 $O_s$ 和目标服务本体 $O_d$ 都可以看成是一个黑箱信息系统。它们从 Internet 环境中感知并服务于 Internet 环境。两个服务本体的输入和输出完全相同,从系统的观点看,显然这两个本体具有相同的服务功能。OWL-S 服务本体的服务简档 serviceProfile 中有两个特殊的元素 hasInput 和 hasOutput,它们对服务的输入输出进行描述。

**规则 2** 当两个服务本体具有完全相同的输入输出时,这两个服务本体语义匹配,即:

IF hasInput($O_s$)=hasInput($O_d$) and

hasOutput($O_s$)=hasOutput($O_d$)

THEN $O_s \models O_d$

4.1.3 通过 OWL-S 的元素在本体树中的节点位置进行匹配

**规则 3** 具有完全相同的父元素、子元素和兄弟元素的两个元素是语义匹配的,即:

IF hasFather($V_s$)=hasFather($V_d$) and

hasSons($V_s$)=hasSons($V_d$) and

hasBrothers($V_s$)=hasBrothers($V_d$)

THEN $V_s \models V_d$

4.1.4 通过 OWL-S 元素的 URI 进行匹配

**规则 4** 如果两个元素的 URI 相同,那么这两个元素是等价的[13],即:

IF hasURI($O_s$)=hasURI($O_d$)

THEN $O_s \models O_d$ 或者

IF hasURI($V_s$)=hasURI($V_d$)

THEN $V_s \models V_d$

### 4.2 基于 NBC 的 OWL-S 文本分类式语义匹配

对语义相似匹配,只能使用统计学方法找到一个相对较佳的次优匹配。文本语义匹配通过服务本体中的一些文本信息利用 NBC 文本分类方法实现。OWL-S 目标服务本体 $O_d$ 中的所有元素集被看作分类类别,OWL-S 源服务本体 $O_s$ 中的所有元素及其文本信息被看作新实例。把目标元素中的所有文本信息(包括类、个体、实例、属性、属性值和描述等信息)进行文本抽取,主要包括形式转换、分词、过滤和提取等技术,最终形成由单词集 Vocabulary 构成的文本集。于是 OWL-S 服务本体的语义匹配问题就转换为文本分类问题,将前文所述的 NBC 文本分类方法用于此处就能实现服务本体元素的语义匹配。其算法描述如下:

Step1 收集文本信息并形成单词集。

Vocabulary←$\{w_k \mid \forall \text{word} \in \text{Examples}\}$

Step2 计算 $P(V_d)$ 和 $P(w_k \mid V_d)$

对 $O_d$ 中的每个 $V_d$

$docs_j$←Exaxples 中目标值为 $V_d$ 的文档子集

$$P(V_d) \leftarrow \frac{|docs_j|}{|Examples|}$$

对每个 $w_k \in$ Vocabulary

计算 $P(w_k \mid V_d)$ //依据公式(5)

Step3 用 NBC 分类 $V_s$

pos←在 $V_s$ 中的所有单词的位置

//$V_s$ 中每个单词的位置对应一个属性

$$v_{NB} = \arg\max_{v_j \in \{V_d\}} P(v_j) \prod_{i \in pos} P(\alpha_i \mid v_j)$$

Step4 Return $v_{NB}$ //即 $V_s \models v_{NB}$

Step5 END

### 4.3 语义匹配度

语义匹配的程度用相似度描述,显然语义等价匹配的相似度为 1,而语义相似匹配的相似度介于 0 与 1 之间。

假设 $O_s$ 的所有元素是集合$\{V_{s1}, V_{s2}, \cdots, V_{sn}\}$,定义 $O_s$ 的度为 $n$;假设 $O_d$ 的所有元素是集合$\{V_{d1}, V_{d2}, \cdots, V_{dm}\}$,定义 $O_d$ 的度为 $m$。

则 $V_{si}$ 和与之语义匹配的 $v_{NBj}$ 的相似度记为 $Sim(V_{si}, v_{NBj})$。

在基于 NBC 的 OWL-S 文本分类式语义匹配算法中,$Sim(V_{si}, v_{NBj})$用公式(6)计算:

$$Sim(V_{si}, v_{NBj}) = \max_{v_j \in \{V_d\}} P(v_j) \prod_{i \in pos} P(\alpha_i \mid v_j) \quad (6)$$

本体 $O_s$ 和 $O_d$ 的相似度记为 $sim(O_s, O_d)$,且:

$$Sim(O_s, O_d) = \frac{2\sum_{i=1}^{n} Sim(V_{si}, v_{NBj})}{n + \max(n, m)} \quad (7)$$

其中 $1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m$ 且 $V_{si} \models v_{NBj}$,显然 $0 \leqslant Sim(O_s, O_d) \leqslant 1$。该式刻画了源本体 $O_s$ 和目标本体 $O_d$ 之间的语义匹配程度。

**总结** 网格环境中的 OWL-S 服务,将分布式的异构的资源以服务本体的形式进行表示和描述,增强了服务的语义性。为了实现服务本体的语义共享,本文给出了几条快速有效的 OWL-S 服务本体的语义等价匹配规则和基于 NBC 文本分类的服务本体语义相似匹配算法,并在相似匹配中给出了语义匹配程度的计算公式。文中充分利用本体中的元素信息,从功能、结构、系统和文本内容等多个维度分析和解决本体间的语义匹配问题,增强了服务本体的语义匹配能力,并对基于 NBC 的 OWL-S 文本分类式语义匹配算法中的语义匹配程度进行了定量刻画。

## 参 考 文 献


1 Gruber T R. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In: N. Grarino & R. Poli, eds. Formal Ontology in Conceptual Analysis and Knowledge Representation. The Netherlands: Kluwer Academic Publishers, 1993. 199~220

2 Dean M, Schreiber G, Bechhofer S, et al. OWL web ontology language reference. W3C recommendation, 2004

3 Stevens R, Carole A. Goble and Sean Bechhofer. What is an ontology? Available at: http://www.cs.man.a.c.uk/~stevens/onto/node3.html

4 Ting K M, Witten I H. Issues in stacked generalization. Journal of Artificial Intelligence Research, 1999, 10: 271~289

5 http://dublincore.org/

6 http://www.prismstandard.org/

7 唐杰,等. 语义 Web 中的本体自动映射[J]. 计算机学报, 2006, 29(11):1958

8 Mitchell T M. 机器学习(英文版). 北京:机械工业出版社, 2003, 3:176~183

9 喻坚,等. 面向服务的计算——原理和应用. 清华大学出版社, 2006, 12:211~217

10 Singh M P, Huhns M N. Service-Oriented Computing——Semantics, Processes, Agents. John Wiley & Sons, Ltd, 2005

11 Solomon M, Raman R, Linvy M. Resource management through multilateral matchmaking[A]. In: Proceedings of the Ninth IEEE Symposium on High Performance Distributed Computing [C], Pittsburgh, August 2000. 290~291

12 Bianchini D. Hybrid Ontology Based Matchmaking For Service Discovery[A]. In: Symposium on Applied Computing Proceedings of the 2006 ACM symposium on Applied computing[C], 2006. 1707 ~ 1708

13 Ehrig M, Sure Y. Ontology Mapping — An Integrated Approach: [Technical report]. University of Karlsruhe, Institute AIFB. http://www.aifb.uni-karlsruhe.de/WBS/meh/mapping/, 2004