

以语义规则刻画服务请求的 Top- k Web 服务发现方法

王海^{1,3}, 朱正东¹, 范琳², 李增智¹

(1. 西安交通大学计算机系统结构与网络研究所, 710049, 西安; 2. 西安邮电学院计算机系, 710061, 西安;
3. 北京邮电大学网络与交换技术国家重点实验室, 100876, 北京)

摘要: 为了更好地提供满足客户需求的 Web 服务, 在充分研究服务发布描述、服务请求描述等特点的基础上, 提出了以语义规则刻画的服务请求描述, 并以本体构建了服务发布描述. 通过对语义规则添加影响因子, 使得服务匹配精度可以通过匹配度来度量, 并按照给定的调节系数决定最终匹配是否成功. 在 OWL-S TC V2 测试服务集合中进行了对比实验, 结果表明, 所提方法可以有效提高查全率和查准率, 特别是在最大匹配数为 20 获得的 k 分别为 2、5、8、10、15 的情况下, Top- k 查准率比基于描述逻辑包含检测的算法分别提高了 15%、15%、9%、8%、2%, 比关键字算法分别提高了 5%、13%、30%、33%、30%, 而查准率稳定在 95% 以上.

关键词: Web 服务发现; 语义规则; 查准率

中图分类号: TP393 **文献标志码:** A **文章编号:** 0253-987X(2010)02-0035-04

A Method of Top- k Web Service Discovery Using Semantic Rule Characterizing Service Request

WANG Hai^{1,3}, ZHU Zhengdong¹, FAN Lin², LI Zengzhi¹

(1. Institute of Computer System Structure and Networks, Xi'an Jiaotong University, Xi'an 710049, China; 2. Computer Science and Technology Department, Xi'an Institute of Post and Telecommunications, Xi'an 710061, China; 3. State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: In order to better meet customers' demands for the provision of web services, the semantic rules are used as the means to describe the service requests and the OWL-S service description language is used as service advertisement description language. The contribution of the semantic rules to the matchmaking results is reflected and the calculation of matching degree is facilitated by adding an impact factor to the rules in the service request description. An accommodation coefficient is introduced to judge if a given matching degree is enough to have a matchmaking success. Experiments using OWL-S TC V2 testing service set show that the proposed method increases the effective rate of recall and accuracy. Comparisons with the sub-sumption and the keyword matching based method in cases of k setting 2, 5, 8, 10, 15 and the maximum possible matching being 20 show that the top- k precision is promoted 15%, 15%, 9%, 8%, 2% and 5%, 13%, 30%, 33%, 30%, respectively, while the accurate rate remains stable at 95% or more.

Keywords: Web service discovery; semantic rule; precision

针对以统一描述、发现和集成(UDDI)为代表、以关键字匹配为依据的服务发现技术查全率和查准

率不高的问题,学术界已有不少相关研究.文献[1-3]的方法是以假想的完整服务描述作为服务请求描述的,并通过与服务发布描述中对应部分做比对得到匹配结果.该方法未考虑服务请求描述与服务发布描述之间的不对称性,特别是未考虑服务请求描述的特点,虽然查全率和查准率较高,但难以在实际中得以应用.文献[4]考虑了服务请求描述与服务发布描述之间的不对称性,但未充分认识到服务请求的本质就是一组条件约束的情况,而其非全面的服务特性刻画,使得查全率和查准率难以保障.

本文提出以语义规则刻画的服务请求描述,以本体来构建服务发布描述,通过验证发布描述是否满足请求规则来进行有效的服务发现,这样在保证查全率、查准率的情况下,尽可能提升 Top- k 查准率,更好地完成合适的服务推荐.

1 形式化的服务描述

当前语义 Web 服务的基本做法是,以本体形式构建服务描述^[5-6].本体适用于表达具有分类特征的概念及个体,规则适用于建立概念以及个体之间的约束关系.本文在研究现有语义规则语言^[7-9]的基础上,结合以 Web 本体语言描述逻辑(OWLDL)为基础的本地知识库,提出语义规则的形式描述以及相应的服务请求描述.这样,验证服务发布描述是否满足服务请求描述的问题将转化为带有语义规则的本地知识库的可满足性检测问题.主要方式有 2 种:①结合本体与规则的异构方式^[10-11],其推理由 2 部分构成,即本体推理机和规则引擎协同工作;②将规则直接引入现有的本体语言中,形成本体与规则的同构系统^[12].

可以将服务请求看作一组语义规则,服务发布描述则使用现有的 OWL-S 形式的服务本体,并采用类似于 CARIN^[11]异构方式将本体与规则结合起来.

给定一个知识库 $B_K = (B_O, B_R)$,其中 B_O 表示本体库, B_R 表示规则库.为了实现 B_K 可满足性的判定,需要采用后向推理(自顶向下模式或目标驱动推理),即先对规则进行 SLD(Selective Linear Definite)消解,之后对基化的描述逻辑原子直接使用描述逻辑推理机来检查其可满足性.在具体实现上,本文采用 OO jDREW^[13]作为推理引擎,它支持前向、后向推理.

定义 1 给定 OWLDL 词汇 $V_O = (V_{cls}, V_{op}, V_{dp}, V_{ap}, V_{ind}, V_D, V_{lit})$,其中 V_{cls} 表示类名集合, V_{op} 表示对象属性名集合, V_{dp} 表示数据属性名集合, V_{ap} 表示

注释属性名集合, V_{ind} 表示个体名集合, V_D 表示数据类型名集合, V_{lit} 表示资源描述框架(RDF)的文字集合.令 V_{var} 表示变量名集合,则一个语义规则为

$$B_1 \cap B_2 \cap \cdots \cap B_n \rightarrow H_1 \cap H_2 \cap \cdots \cap H_m$$

式中: $H_1 \cap H_2 \cap \cdots \cap H_m$ 为规则头; $B_1 \cap B_2 \cap \cdots \cap B_n$ 为规则体. $H_1, \dots, H_m, B_1, \dots, B_n$ 形如 $C(x), P(x, y), C \in V_{cls}, P \in V_{op} \cup V_{dp} \cup V_{ap}, x, y \in V_{var} \cup V_{ind} \cup V_{lit}$.如果一个规则头包含多个原子,那么它可以等价地转换为多条规则,每条规则的头只含有一个原子.

语义规则的语义与通常的 Horn 规则类似.对于给定的解释 I ,假定所有的变量已经被指派.如果规则体中的每一个原子 B_i 都满足规则,在规则头中的每一个原子 H_i 就一定满足规则,那么称 I 满足这条语义规则.

定义 2 给定 $V_O = (V_{cls}, V_{op}, V_{dp}, V_{ap}, V_{ind}, V_D, V_{lit})$,一个有限的语义规则序列称为一个语义规则库,即

$$B_R = \{r_0, r_1, \dots, r_n\}$$

定义 3 给定 $V_O = (V_{cls}, V_{op}, V_{dp}, V_{ap}, V_{ind}, V_D, V_{lit})$,一个语义规则知识库由 Web 本体语言(OWL)知识库和语义规则库组成,记作 $B_K = (B_{Kowl-dl}, B_R)$.

对于给定的解释 $\mathcal{I}$,如果 $\mathcal{I}$ 是 $B_{Kowl-dl}$ 的模型,并且对于 B_R 中的规则都可以同时满足,则称 $\mathcal{I}$ 可以满足 B_K .

定义 4 给定 $V_O = (V_{cls}, V_{op}, V_{dp}, V_{ap}, V_{ind}, V_D, V_{lit})$,一个服务请求描述可以表示为一个 $B_K = (B_{Kowl-dl}, B_R)$.

2 服务发现算法

定义 5 给定 $V_O = (V_{cls}, V_{op}, V_{dp}, V_{ap}, V_{ind}, V_D, V_{lit})$,一个 K 扩展语义规则为

$$K: B_1 \cap B_2 \cap \cdots \cap B_n \rightarrow H_1 \cap H_2 \cap \cdots \cap H_m$$

式中: K 为影响因子,表示该规则对匹配结果的影响程度.

为规则添加了影响因子之后,相应的服务匹配结果计算需要做一些改动.给定服务请求描述(RD),不失一般性,假设 RD 中有 n 条 K 扩展语义规则,其相应的影响因子为 $K_{r_1}, K_{r_2}, \dots, K_{r_n}$.给定一个服务发布描述 d_a ,它可以满足 RD 中的 m 条规则,相应影响因子为 $K_{d_a 1}, K_{d_a 2}, \dots, K_{d_a m}$.

定义 6 服务发布描述对服务请求描述的满足程度称为匹配度,即

$$D_M = \frac{\sum_{i=1}^m K_{d_{a_i}}}{\sum_{j=1}^n K_{r_j}}$$

此定义使得 D_M 取值范围在 $0 \sim 1$ 之间,并且每条规则按影响因子的比例对匹配度起作用。相应地,在引入匹配度之后,对于给定的服务发布描述和服务请求描述,判定其是否匹配需要一个相应的阈值,称作调节系数 α 。当 $D_M < \alpha$ 时,匹配失败;当 $D_M > \alpha$ 时,匹配成功。

对于 2 个不同的服务发布描述 d_{a1} 、 d_{a2} ,给定服务请求描述,并假定相应的匹配度为 D_{M1} 、 D_{M2} 。如果 $D_{M1} \geq \alpha$, $D_{M2} \geq \alpha$,并且 $D_{M1} \leq D_{M2}$,称 d_{a2} 优于 d_{a1} ,记作 $d_{a1} \leq d_{a2}$,以此作为服务发现算法返回结果的排序依据。

根据上述分析,得到 Top- k 服务发现算法(TK-SDA)如图 1 所示。

```

输入: 服务请求描述, 服务发布描述集合, 调节系数  $\alpha$ 
输出: 匹配度大于  $\alpha$  的服务发布描述集合
流程: For 对每一个服务发布描述集合中的  $d_a$ 
      For 对每一个服务请求描述中的规则  $r$ ,
        其相应的影响因子为  $K_i$ 
          If !Sat( $d_a \cup r$ ) then
            continue
          Else
             $D_M = D_M + K_i / \sum_{j=1}^n K_j$ 
          End For
        If  $D_M \geq \alpha$  then
          将  $d_a$  添加到服务请求描述中
        End For
      End For
      对服务请求描述中的服务发布描述按  $D_M$  值进行降序排序
  
```

图 1 Top- k 服务发现算法

Top- k 服务发现算法可以很好地按照给定的服务请求描述、调节系数和影响因子,从服务发布描述集合中抽取出符合约束的服务发布描述,并且将其按照降序排列,以便于进行后续的服务选择。

3 服务发现算法评价指标

目前,广泛采用的服务发现算法评价指标主要包括查全率和查准率。由于本文旨在寻找满足用户需求的前 k 个服务描述,因此采用全局性的查准率与查全率不妥。作者在此提出 Top- k 查准率及 Top- k 稳定系数作为新的评价指标。

定义 7 假定满足用户需求的 Web 服务集合为 S_{Satis} ,服务发现算法返回的服务集合为 S_{Returk} ,则 Top- k 查准率为

$$p_k = \frac{|S_{Satis} \cap S_{Returk}|}{\min(k, |S_{Returk}|)}$$

式中: S_{Returk} 表示前 k 个返回结果。

定义 8 用 Top- k 查准率稳定系数 σ 表示 Top- k 查准率随 k 值变化的稳定程度,即

$$\sigma = 1 - \left(\frac{1}{N} \left(\sum_{i=1}^N p_i^2 \right) - \bar{p}_k^2 \right)^{1/2}$$

从定义 8 可以看出,Top- k 查准率是一个随 k 变化的函数,倘若作为衡量指标,该函数随 k 的变化需保持稳定。为此,本文用 σ 进行度量。由于 k 为离散取值,因此可用 p_k 的均值 $\bar{p}_k$ 作为特定 p_k 的基值。

4 实验及评价

为了对语义规则驱动的服务发现算法进行比较评价,本文使用了 OWL-S 服务发现测试集 OWL-S TC V2^[14],该测试集包含符合 OWL-S 1.1 推荐标准的 570 多个服务描述,覆盖 7 个应用领域,如教育、医疗保险、食品、旅行、通信、经济和军事。OWL-S TC V2 同时提供了 28 个测试查询集(服务请求),每个查询与 10~20 个不等的服务相关。另外,对这 28 个测试查询以语义规则的形式重新构建,并按照 Top- k 服务发现算法进行了实验。为了对比分析,将这 28 个测试查询按关键字进行提取,按照文献[5]方法以关键字相似度为依据执行发现算法,以文献[5]中的描述逻辑,即包含检测为基础的服务发现算法进行了测试。

在测试的返回结果集中,最大的匹配数为 20,因此 $k=2,5,8,10,15$,由此得到如图 2、图 3 所示的统计结果。从图 2 可以看出:Top- k 查准率在 $k=2,5,8,10,15$ 时,分别比基于描述逻辑包含检测的算法提高了 15%、15%、9%、8%和 2%,相对于关键字算法分别提高了 5%、13%、30%、33%和 30%。从图 2 的走势还可以看出,Top- k 服务发现算法针对不同的 k 值,其 Top- k 查准率均稳定在 95%以上,而基于包含检测的服务发现算法的 Top- k 查准率,在 k 较小时结果不理想,随着 k 的增大,结果逐渐逼近了包含检测的服务发现算法。

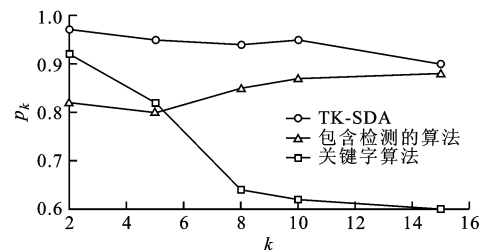
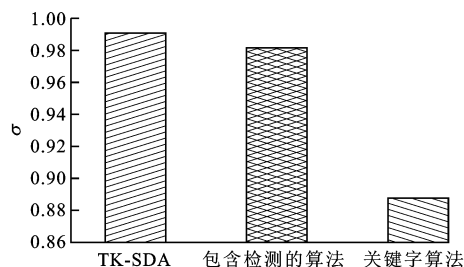


图 2 Top- k 查准率

图3 Top- k 查准率稳定系数

5 结束语

本文针对服务发布描述、服务请求描述的特点,设计了以语义规则为基础的服务请求描述.以此为出发点,对基本语义规则添加了影响因子,并给出了服务匹配度的计算方法.以调节系数作为判定匹配与否的标志,由此设计了完整的服务发现算法.对于以本体形式构建的服务发布描述起始于寻找满足特定请求的服务集合,该方法具有以下优点:①将服务请求描述的机制与服务发布描述的机制区分开来,使得服务请求描述可以仅表达用户关心的部分;②以影响因子来反映特定请求约束规则对匹配结果的影响程度,使得服务请求描述中的约束条件弹性更大,同时匹配度的计算充分考虑了影响因子,这样匹配算法的区分度更高.与基于描述逻辑包含检测的算法以及基于关键字的相似度算法进行了对比实验,证实了本文算法的 Top- k 查准率随着 k 的变化始终保持在较高水平,包含检测的算法和关键字算法的 Top- k 查准率随着 k 的变化波动非常大.

在目前的算法中,规则的影响因子与匹配度的调节系数是由作者根据实际服务请求的意图手工指定的,在未来的工作中,将通过对情景信息的处理,辅助得到特定规则的影响因子和匹配度的调节系数,使得方法更为简便、流畅.

参考文献:

- [1] PAOLUCCI M, KAWAMURA T, PAYNE T R. Semantic matching of web services capabilities[C]//1st International Semantic Web Conference Proceedings. Berlin, Germany: Springer, 2002: 333-347.
- [2] 朱正东, 缪相林, 胡亚红, 等. 可用性语义 Web 服务的通用发现机制[J]. 西安交通大学学报, 2008, 42(6): 659-663.
ZHU Zhengdong, MIAO Xianglin, HU Yahong, et al. Usability based general semantic Web services discovery mechanism[J]. Journal of Xi'an Jiaotong University, 2008, 42(6): 659-663.
- [3] GRIMM S, MOTIK B. Matching semantic service descriptions with local closed-world reasoning[C]//Proceedings of the 3rd European Semantic Web Conference on the Semantic Web. Berlin, Germany: Springer, 2006: 575-589.
- [4] 何玲娟, 刘连臣, 吴澄. 一种改进的基于 WSDL 描述的操作相似性度量方法[J]. 计算机学报, 2008, 31(8): 1331-1339.
HE Lingjuan, LIU Lianchen, WU Chen. Asymmetric web service operation matching in structural-level similarity measure[J]. Chinese Journal of Computers, 2008, 31(8): 1331-1339.
- [5] MARTIN D, BURSTEIN M. OWL-S: semantic markup for web services[EB/OL]. [2008-10-10]. <http://www.daml.org/services/owl-s/1.1/overview/>.
- [6] ROMAN D, LAUSEN H, KELLER U. D2 v1.1 web service modeling ontology (WSMO): WSMO final draft[EB/OL]. (2005-02-10). [2008-10-10]. <http://www.wsmo.org/>.
- [7] HORROCKS I, PATEL-SCHNEIDER P F, BOLEY H, et al. SWRL: a semantic web rule language combining OWL and RuleML[EB/OL]. [2008-10-10]. <http://www.w3.org/Submission/SWRL/>.
- [8] BOLEY H, TABET S, WAGNER G. Design rationale of RuleML: a markup language for semantic Web rules[C]//Proceedings of International of Semantic Web Working Symposium (SWWS). New York, USA: ACM, 2001: 381-401.
- [9] FRIEDMAN-HILL E. Jess in action: Java rule-based systems (in action series)[M]. Greenwich, USA: Manning Publications, 2002.
- [10] DONINI F M, LENZERINI M, NARDI D, et al. Al-Log: integrating datalog and description logics[J]. Journal of Intelligent Information Systems, 1998, 10(3): 227-252.
- [11] LEVY A Y, ROUSET M C. Combining horn rules and description logics in CARIN[J]. Artificial Intelligence, 1998, 104(1/2): 165-209.
- [12] 梅婧. 语义 Web 本体与规则[D]. 北京: 北京大学数学科学院, 2007.
- [13] BALL M, BOLEY H, HIRTLE D, et al. The OO jDREW reference implementation of RuleML[C]//Proceedings of Rules and Rule Markup Languages for the Semantic Web. Berlin, Germany: Springer, 2005: 218-223.
- [14] Sem Web Central. Owls-TC: an OWL-S service retrieval test collection[EB/OL]. [2008-10-10]. <http://projects.semwebcentral.org/projects/owls-tcV2>.

(编辑 苗凌)