

基于 QoS 和 OWL-S 的 Web 服务发现研究

杨清平^{1,2} 邱玉辉² 蒲国林^{1,2}

(四川文理学院计算机科学系 达州 635000)¹ (西南大学语义网络实验室 重庆 400715)²

摘 要 建立了原子服务发现模型,探讨了原子级服务匹配的 4 个过程,并重点对输入输出匹配和 QoS 匹配进行了研究。在 QoS 匹配中,给出了服务性价比概念,并提出了高性价比的服务选择原则。

关键词 原子服务,服务发现,OWL-S,QoS

Research on Web Service Discovery Based on OWL-S and QoS

YANG Qing-ping^{1,2} QIU Yu-hui² PU Guo-lin^{1,2}

(Computer Science Department, Sichuan University of Arts and Science, Dazhou 635000, China)¹

(Semantic Grid Laboratory, Southwest University, Chongqing 400715, China)²

Abstract We built an atomic service discovery model, discussed four processes of atom-level service matching, and mainly studied I/O matching and QoS matching. In the QoS matching, proposed performance-price ratio, and advised that user discovered services on the basis of high performance-price ratio principle.

Keywords Atomic service, Service discovery, OWL-S, QoS

Web 服务^[1]是当前主要的服务实现技术。W3C 定义 Web 服务是由 URI 标识的软件应用程序,其接口和绑定可以通过 XML 构件进行定义、描述和发现。Web 服务的体系结构中定义了 3 种角色:服务提供者、服务注册中心和服务请求者。Web 服务使得通过内容而不仅是关键字访问 Web 资源成为可能。OWL-S 本体用本体描述 Web 服务,增强了服务的语义描述和语义处理能力。它提供了描述 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。基于语义的 Web 服务发现已成为目前学术界的研究热点^[2,3],而 OWL-S 使基于语义的 Web 服务发现成为可能。本文在 OWL-S 本体描述的基础上增加了 QoS 约束,不但可以描述服务的语义,还能描述服务的质量,使语义 Web 服务发现有质量保证。

1 服务描述

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 3 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(Service-Model)和服务基点(ServiceGrounding)。在 ServiceProfile 中,就描述了 IOPE 等内容。IOPE 分别指 Web 服务的输入(input)、输出(output)、执行的前提条件(precondition)和执行后的结果(effect)。

根据 OWL-S^[4]规范,Web 服务可被划分为原子服务和组

合服务。原子服务对应于通过一次交互完成的服务,它提供不需要再分解的服务;组合服务对应于需要多个动作完成的服务,它由原子服务或粒度更小的组合服务动态组成。特别地,当组合服务由一个原子服务组成时,该原子服务就是一个组合服务。因此,原子服务是一个特殊的组合服务,即粒度最小的组合服务。因此,原子服务发现是服务发现的基础,原子服务发现的质量直接关系到用户能否获得满意的服务。本文重点讨论原子服务发现的模型和过程,因此后文的服务在未说明的情况下都是指原子服务。

定义 1(服务) 服务可以表示为多元组,记为:

$WS = \langle WSName, WSID, WSInput, WSOOutput, WSContext, WSDomain, WSQoS \rangle$

其中,

- (1) WSName 是服务的名称;
- (2) WSID 是服务的唯一标识;
- (3) WSInput 是 WS 的输入集合;
- (4) WSOOutput 是 WS 的输出集合;
- (5) WSContext 是 WS 的语境信息,它提供 WS 向上向下组合的语义信息和成功组合的案例;
- (6) WSDomain 是 WS 的应用领域信息;
- (7) WSQoS 是 WS 的服务质量描述,包括响应时间、费用开销和安全级别与可靠性等。

并设 Name(), ID(), Input(), Output(), Context() 和 Domain() 为特征提取函数。

本文使用 OWL-S 描述 Web 服务,前 6 元可从其描述中

到稿日期:2008-06-29 本文受西南大学研究生科技创新基金(2006011)资助。

杨清平(1954—),男,副教授,主要研究方向为语义 Web、人工智能,E-mail: yqp_1@163.com;邱玉辉(1938—),男,教授,博士生导师,主要研究方向为人工智能、服务计算、语义网络;蒲国林(1971—),男,副教授,博士研究生,主要研究方向为语义网络、服务计算。

获取,而 WSQoS 是增加的服务质量(QoS)描述。

2 原子服务发现模型

原子服务发现是服务发现的基础。当原子服务发现系统接收到用户的服务请求时,首先判断是否存在成功案例,并判断这些成功案例能否被调用成功。如果存在一个案例能成功调用,则跳过原子级匹配过程,并将该原子服务作为系统的原子服务发现结果;如果没有成功案例或成功案例中的所有服务都不能被成功调用,则进入原子级匹配过程。成功案例的使用主要是避免系统中重复的匹配过程,以提高系统的效率。

定义 2(原子级匹配) 原子级匹配是原子服务和用户的服务请求之间的匹配,它是从注册服务库中搜索并发现能满足用户服务请求的原子服务的过程。

图 1 是原子服务发现模型,其输入主要包括用户的一个服务请求和系统的注册服务库两部份。其输出是一个能满足该服务请求的服务。WS_r 表示用户的服务请求,WS_i 表示注册服务库中的一个服务。

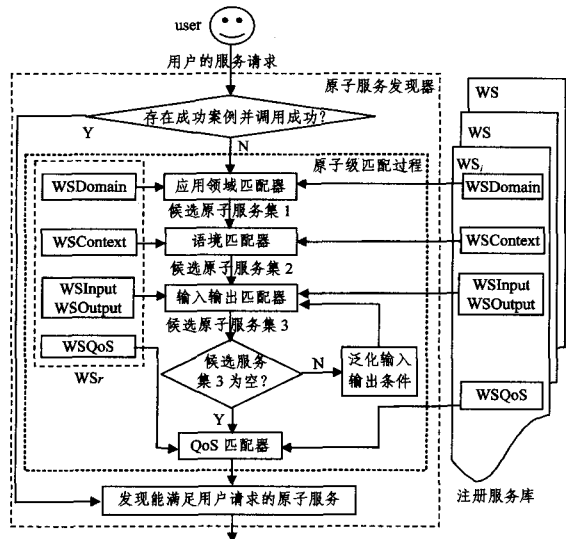


图 1 原子服务发现模型

3 原子级匹配过程

原子级匹配包括应用领域匹配、语境匹配、输入输出匹配和 QoS 匹配。

3.1 应用领域匹配

在服务发现过程中,应用领域信息非常重要。一个单词在不同的应用领域有不同的语义。在进行应用领域匹配时,原子服务和用户请求应使用同一个分类标准。现有的分类标准有 NAICS(North American Industry Classification System)和 UNSPSC(Universal Standard Products and Services Classification)等。应用领域匹配过程是一个大粒度的过滤过程,它由应用领域匹配器完成。但通过应用领域匹配可以过滤掉大量的与用户服务请求领域无关的原子服务。其匹配规则为:如果 $Domain(WS_i) = Domain(WS_r)$,则服务 WS_i 和服务请求 WS_r 是应用领域匹配的。将满足该匹配规则的所有原子服务作为候选原子服务集 1。

3.2 语境匹配

语境匹配是比应用领域匹配粒度更小的过滤过程。语境

匹配的输入是候选原子服务集 1,它的输出是对语境匹配过滤得到的候选原子服务集 2。语境匹配是语义上的匹配,它有效地解决了同义词和一词多义的现象,并得到与当前语境有关的原子服务集。

在文献[5]中介绍了当前语境信息的语义抽取过程。其过程简述如下:

首先建立基于 WordNet^[6] 的与当前语境相关的词法空间,然后根据词法空间建立本体的概念集,再根据概念集生成词法本体,该词法本体表示了当前语境的语义抽取结果。于是语境匹配问题就转换为词法本体的匹配问题。而本体的匹配技术已经相对成熟。

3.3 输入输出匹配

输入输出匹配的输入是候选原子服务集 2,输出候选原子服务集 3,这些原子服务都能在一定程度上满足该服务请求。根据输入输出匹配的程度,将匹配分成如下几个等级:Exact, Plugin, Subsume, Overlap 和 Fail 匹配^[7],显然它们的匹配度是递减的。FAIL 匹配表示匹配失败,所以它不包含在输入输出匹配器中。

从服务输入输出满足用户需求的程度来看,有如下定义:

定义 3(Exact 匹配) 服务请求 WS_r 和原子服务 WS_i 是 Exact 匹配,表示为:

$$Exact(WS_r, WS_i) = \{ (Input(WS_r) \text{ equivalentClass } Input(WS_i)) \wedge (Output(WS_r) \text{ equivalentClass } Output(WS_i)) \}$$

定义 4(Plugin 匹配) 服务请求 WS_r 和原子服务 WS_i 是 Plugin 匹配,表示为:

$$Plugin(WS_r, WS_i) = \{ (Input(WS_r) \text{ subclassOf } Input(WS_i)) \wedge (Output(WS_r) \text{ subclassOf } Output(WS_i)) \}$$

定义 5(Subsume 匹配) 服务请求 WS_r 和原子服务 WS_i 是 Subsume 匹配,表示为:

$$Subsume(WS_r, WS_i) = \{ (Input(WS_r) \text{ subclassOf } Input(WS_i)) \wedge (Output(WS_r) \text{ subclassOf } Output(WS_i)) \}$$

定义 6(Overlap 匹配) 服务请求 WS_r 和原子服务 WS_i 是 Overlap 匹配,表示为:

$$Overlap(WS_r, WS_i) = \{ (Input(WS_r) \text{ intersectionOf } Input(WS_i)) \wedge (Output(WS_r) \text{ intersectionOf } Output(WS_i)) \}$$

定义 7(Fail 匹配) 服务请求 WS_r 和原子服务 WS_i 是 Fail 匹配,表示为:

$$Fail(WS_r, WS_i) = \{ (Input(WS_r) \text{ disjointWith } Input(WS_i)) \wedge (Output(WS_r) \text{ disjointWith } Output(WS_i)) \}$$

Exact 和 Plugin 匹配发现的服务能满足服务请求的所有要求,Subsume 匹配只发现了满足部份服务请求的服务,而 Fail 匹配发现的服务不能满足服务请求。在输入输出匹配器中,首先进行 Exact 匹配,如果没有发现原子服务,则降低匹配要求进行 Plugin 匹配,以此类推。这个匹配过程是条件泛化继续匹配直至发现原子服务的过程。Overlap 匹配的程度可以通过相似度或语义距离的计算得到,本文不作研究。一般情况下,通过 Exact 和 Plugin 匹配就能发现多个服务,它们是由不同公司开发的能满足同一服务请求的服务。正因为如此,必须还要对这些原子服务进行 QoS 匹配。

3.4 QoS 匹配

定义 8(QoS) 原子服务的 QoS 可以表示成一个四元组

(下转第 166 页)

$$D^\alpha = \left[\begin{array}{cccccc} \emptyset & & & & & \\ a_1, a_2, a_3 & \emptyset & & & & \\ a_2 & a_1, a_3, a_4 & \emptyset & & & \\ a_1, a_4 & A & a_3, a_4 & \emptyset & & \\ a_1, a_2 & a_1, a_2, a_4 & A & a_1, a_2, a_4 & \emptyset & \\ a_1, a_2, a_3 & a_2 & a_1, a_2, a_3 & A & \emptyset & \emptyset \end{array} \right]$$

$$\begin{aligned} \Delta^\alpha(K) &= a_2 \wedge (a_1 \vee a_4) \wedge (a_3 \vee a_4) \\ &= (a_2 \wedge a_4) \vee (a_1 \wedge a_2 \wedge a_3) \end{aligned}$$

此时, $\{a_2, a_4\}$ 与 $\{a_1, a_2, a_3\}$ 是区间值信息系统的 α 分类约简。明显地, 例 1 中在水平 $\alpha=0.5$ 时的分类约简 $\{a_1, a_4\}$ 不是水平 $\alpha=0.6$ 时的分类协调集。

参考文献

[1] Pawlak Z. Rough set. International Journal of Computer and Information Science, 1982(11): 341-356
 [2] Pawlak Z. Rough set; Theoretical Aspects of Reasoning about Data. Boston; Kluwer Academic Publishers, 1991
 [3] 王钰, 苗夺谦, 等. 关于 Rough Set 理论与应用的综述[J]. 模式识别与人工智能, 1996(6): 337-344
 [4] 刘清. Rough 集及 Rough 推理[M]. 北京: 科学出版社, 2001
 [5] Han J, Cai Y, Cercone N. Data-driven discovery of quantitative rules in relational databases. IEEE Transactions on Knowledge and Data Engineering, 1993, 5 (1): 29-40

[6] Mrozek A. Rough sets and dependency analysis among attributes in computer implementations of expert's inference models [J]. International Journal of Man-Machine Studies, 1989, 30 (4): 457-473
 [7] Yasdi R. Combining rough sets learning and neural learning: method to deal with uncertain and imprecise information, Neuralcomputing, 1996, 7 (1): 61-84
 [8] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001: 80-100
 [9] 张文修, 梁怡, 吴伟志. 信息系统与知识发现[M]. 北京: 科学出版社, 2003
 [10] Greco S, Matarazzo B, Slowinski R. Rough sets theory for multi-criteria decision analysis. European Journal of Operational Research, 2001, 129: 1-47
 [11] Kim D. Data classification based on tolerant rough set. Pattern Recognition, 2001, 34: 1613-1624
 [12] Yao Y Y. Two views of the theory of rough sets in finite universes. International Journal of Approximate Reasoning, 1996, 15 (4): 291-317
 [13] Yao Y Y. Information granulation and rough set approximation. International Journal of Intelligent Systems, 2001, 16 (1): 87-104
 [14] Skowron A, Rauszer C. The Discernibility Matrices and Functions in Information Systems [M]. Boston; Kluwer Academic Publishers, 1992: 331-362

(上接第 147 页)

$\langle time, cost, security, reliability \rangle$ 。其中, $time$ 是原子服务的响应时间; $cost$ 是获取该原子服务所必须支付的费用; $security$ 是原子服务的安全等级; $reliability$ 是原子服务的可靠性。并设 $time()$, $cost()$, $security()$ 和 $reliability()$ 为 QoS 提取函数。

WS_i 和 WS_r 是 QoS 匹配的, 则必须满足如下条件:

- I: $security(WS_i) \geq security(WS_r)$;
- II: $reliability(WS_i) \geq reliability(WS_r)$;
- III: $time(WS_i) \leq time(WS_r)$;
- IV: $cost(WS_i) \leq cost(WS_r)$ 。

I 式表示服务的安全等级不能低于用户要求的安全等级; II 式表示服务的可靠性级别不能低于用户要求的可靠性级别; III 式表示服务的响应时间不能高于用户给定的最长响应时间; IV 式表示服务的收费不能高于用户愿承担的最高费用。

当满足上述条件的服务不止一个时, 怎样选择服务呢? 肯定不是 $security(WS_i)$ 和 $reliability(WS_i)$ 越大越好, 也不是 $time(WS_i)$ 和 $cost(WS_i)$ 越小越好。因为这样的选择或许是不经济实用的, 或许是资源浪费的, 虽然它的其它性能很好。在此, 给出基于 QoS 的服务性价比的定义:

定义 9(服务性价比) 用 $\langle time, cost, security, reliability \rangle$ 描述服务的 QoS, 则其性价比 λ 定义如下:

$$\lambda = \frac{security \cdot reliability}{time \cdot cost} \quad (3)$$

其中, $\frac{security \cdot reliability}{time}$ 是对 QoS 性能的刻画, 它与安全性和可靠性成正比, 与响应时间成反比; $cost$ 是价格。

不同的用户群对 QoS 的要求不同, 但追求高的性价比是用户的普遍要求。

高性价比选择策略。在满足 QoS 匹配条件的服务中, 选

择具有最高性价比的服务推荐给用户。

该策略不仅能为用户提供满意的服务, 而且还把高性能的服务留给那些有特殊要求的用户。如航空航天业、国家安全机关和金融业等。

结束语 原子服务发现就是从原子服务库中搜索并发现能满足服务请求的原子服务的过程。文中提出的 QoS 匹配策略回答了“怎样从已发现的能满足用户服务请求的原子服务集中找到一个用户满意的服务”。原子服务发现是组合服务发现的基石, 对该模型的细化深入和在该模型下的组合服务发现将是下一步研究的重点。

参考文献

[1] <http://www.w3.org/2002/ws/>
 [2] Agarwal S, Handschuh S, Staab S. Annotation, Composition and Invocation of Semantic Web Service[J]. Journal on Web Semantics, 2005, 2(1): 1-24
 [3] Berardi D, Calvanese D, Giacomo G D, et al. Automatic Composition of E-services that Export Their Behavior[J]//Proc. of 1st Intl. Conf. on Service Oriented Computing. Trento, Italy, 2003
 [4] Martin D, Burstein M, Hobbs J, et al. OWL-S: Semantic markup for Web services. 2004. <http://www.w3.org/Submission/OWL-S/>
 [5] Chen Wang - Hu, Liu Chen, Li Hou - Fu, et al. An Approach to Dynamically Forming Semantic Infrastructure for Virtual Organizations[J]. Chinese Journal of Computers, 2006(7): 1130-1132
 [6] Kornilakis H, Grigoriadou M, et al. Using WordNet to support interactive concept map construction[J]// Proceedings of the IEEE International Conference on Advanced Learning Technologies. Joensuu, Finland, 2004: 600-604
 [7] Paolucci M, Kawamura T, Payne T R, et al. Semantic Matching of Web Services Capabilities//International Semantic Web Conference, 2002: 333-347