

基于语义的服务混合匹配方法研究

李迎秋^{1,2}, 温涛^{1,2}

LI Ying-qiu^{1,2}, WEN Tao^{1,2}

1.东北大学 软件中心,辽宁 大连 116023

2.大连东软信息学院 计算机科学与技术系,辽宁 大连 116023

1.Software Center, Northeastern University, Dalian, Liaoning 116023, China

2.Department of Computer Science and Technology, Dalian Neusoft Institute of Information, Dalian, Liaoning 116023, China

E-mail: liyingqiu@neusoft.edu.cn

LI Ying-qiu, WEN Tao. Research on service hybrid matching method based on semantics. Computer Engineering and Applications, 2010, 46(10): 223-225.

Abstract: In order to improve the efficiency and precision of Web service matchmaking, the paper presents a hybrid matching method based on semantics. The method extends the Web services functional description model, upon which hybrid matchmaking is carried out. The proposed hybrid method combines two matching strategies: A deductive strategy based on logical relationship between ontology concepts and a similarity-based strategy based on similarity computations between ontology concepts. During the matchmaking process, the deductive strategy is firstly used to distinguish candidate services into four matching types. Then, the similarity-based strategy is used to further match among the candidates. Experiment results show that the proposed method is feasible and effective.

Key words: Web service; service matchmaking; hybrid matchmaking; functional description model

摘要: 为了提高 Web 服务匹配的效率和灵活性, 提出了一种基于语义的 Web 服务混合匹配方法。该方法首先扩展了服务的功能描述模型, 并基于该扩展模型进行混合服务匹配。混合匹配方法联合使用了两种匹配策略: 基于推理的匹配策略, 通过本体概念的逻辑关系实现匹配; 基于相似性计算的策略, 通过本体概念间相似度计算进行服务的匹配。在服务匹配过程中, 首先使用基于推理的匹配策略将服务分成 4 个匹配等级, 然后, 采用基于相似性计算的匹配策略对候选服务进行进一步地匹配。实验结果表明该方法是可行而有效的。

关键词: Web 服务; 服务匹配; 混合匹配; 功能描述模型

DOI: 10.3778/j.issn.1002-8331.2010.10.069 **文章编号:** 1002-8331(2010)10-0223-03 **文献标识码:** A **中图分类号:** TP391.75

Web 服务是一种基于 Web 环境的自包含、自描述、模块化且具有良好互操作能力的新应用, 可以通过网络进行发布、定位和调用。由于 Web 服务通用标准的采用及需求的增加使得互联网上 Web 服务的数量不断增加, 这也使得如何快速、准确地发现目标服务成为一个迫切需要解决的问题。目前, 工业界关于服务注册与发现的标准是 UDDI, 其所支持的基于关键字的服务匹配机制只能实现语法层次上的匹配, 服务发现的性能低下。近年来, 作为 Web 服务语义扩展的语义 Web 服务在学术界引起了广泛关注, 语义 Web 服务在服务描述中包含了丰富的语义信息, 服务发现算法可以利用这些语义信息进行服务的匹配, 从而克服基于关键字的服务发现机制所存在的不足。对目前的基于语义的 Web 服务匹配策略进行了分析, 在此基

础上提出了一种 Web 服务混合匹配方法, 并通过实验对该方法的有效性进行了验证。

1 相关研究

基于语义的 Web 服务匹配方法主要有基于推理的匹配方法和基于相似性计算的匹配方法。

基于推理的服务匹配是建立在领域本体概念间的层次关系之上的, 使用描述逻辑进行推理匹配的方法^[1-3]。该方法根据服务的描述, 利用本体知识对服务的输入、输出参数进行匹配, 得出服务匹配的等级。基于推理的匹配方法有清晰的匹配结构, 运行效率较高, 但对服务的区分度不够。

基于相似性计算的服务匹配^[4-6]通过比较服务的描述, 评估

基金项目: 国家自然科学基金(the National Natural Science Foundation of China under Grant No.60602061); 国家高技术研究发展计划(863)(the National High-Tech Research and Development Plan of China under Grant No.2006AA01Z413); 大连市 2008 年度 IT 教师科研成果项目(2008 IT Instructors Scientific Research Achievements Project of Dalian Government)。

作者简介: 李迎秋(1972-), 女, 博士生, 副教授, 主要研究方向: 服务计算; 温涛(1962-), 男, 博士生导师, 教授, 主要研究方向: 网络安全、网格计算与应用研究、知识组织等。

收稿日期: 2008-09-25 **修回日期:** 2008-11-24

服务功能的相似度(一个介于 0 和 1 之间的数值)来得到更精确的匹配结果。基于相似性计算的服务匹配方法服务区分度高,但同时算法的复杂性也较高。

目前对语义 Web 服务匹配的研究大多针对服务的接口信息,即输入、输出参数,进行匹配,但是,仅依靠 Web 服务的接口信息,并不能完全满足语义匹配的需要。这一方面是由于服务的功能不能仅通过一系列的输入和输出参数表示,另一方面,接口信息中的语义信息也不够丰富。而 Web 服务的非功能性描述,例如文本描述,在表达 Web 服务内容时包含了比服务接口更为丰富的语义。两个服务的文本描述信息为:“Allow customers to order air tickets from flight company.”和“Allow developers to retrieve dates for specific US holidays.”。文献[7]结合 WordNet[®]对服务的文本描述信息采用基于文本的相似性比较,提高了服务匹配的精度。但该方法并没有充分利用领域本体信息,并且由于自然语言描述的随意性,把这些词汇等同比较显然是不恰当的。事实上,可以把以上两个服务的基本功能简要描述为“order tickets”、“retrieve dates”,这些动宾短语在表达该服务功能时起到了比其他词汇更重要的作用。由于 Web 服务包含了一系列操作和方法,可以将 Web 服务看成是面向特定领域完成某个特定任务的复杂应用程序,所以,在一定程度上 Web 服务的功能可以由表示动作的动词与表示动作施加对象的名词或动名词组成的动宾短语来描述,这样就可以利用本体对概念的相似性进行比较了。

基于以上的分析,在接口信息的基础上,对 Web 服务的功能描述模型进行了扩展,增加从服务描述中提取的动宾短语来补充描述 Web 服务的功能,并将两种服务发现策略,即基于推理和基于相似性计算的服务匹配策略相结合,提出 Web 服务混合匹配方法,该方法的基本思想是:将服务匹配过程分为两个阶段,第一阶段,首先采用简单高效的基于推理的匹配策略对 Web 服务扩展的描述信息进行匹配,得到 4 个匹配等级,即 Exact match、Plug-in match、Subsume match、mismatch;第二阶段,针对第一阶段匹配产生的服务候选集,进一步采用相似性计算的匹配策略基于接口信息进行匹配,得到介于[0-1]区间的相似度值。由于在大多数服务匹配情况下(Exact match 或 mismatch),不需要计算服务的相似度,从而提高了服务匹配的效率;另一方面,对 Web 服务功能扩展描述信息一动宾短语的匹配,又弥补了仅靠接口信息进行匹配的不足,提高了服务匹配的精确度。

2 Web 服务描述模型

首先定义扩展的 Web 服务的描述模型。按照 W3C Web Services Architecture 小组对 Web 服务的定义^[9],一个 Web 服务可以形式化地定义如下。

定义 1 Web 服务可以定义为:

$$WS = \langle S, C, Q \rangle$$

其中, S 是基本描述,即服务名称、服务的商业、实体以及服务的文本描述; C 是服务功能描述; Q 是对服务非功能属性的描述。目前,大多数的 Web 服务描述语言都自发地遵循 $\langle S, C, Q \rangle$ 。

定义 2 Web 服务功能定义为:

$$C = \langle I/O, F \rangle$$

其中, I/O 是服务接口,见定义 3; F 是扩展的功能描述,见定义 4。

定义 3 Web 服务的接口定义为:

$$I/O = \langle I, O \rangle$$

$I = \{In_1, In_2, \dots, In_n\}$ 是服务的输入集合,其中每个输入都用

本体进行标注;

$O = \{Out_1, Out_2, \dots, Out_m\}$ 是服务的输出集合,其中每个输出都用本体进行标注。

定义 4 Web 服务扩展的功能 F 定义为:

$$F = \langle Object, Action \rangle$$

其中, $Object$ 表示服务施加操作的对象,是来自于某领域本体的概念,如 *hotel*, *airline*, *car* 等。 $Action$ 表示服务动作,一般针对某一具体客体进行,如 *book*, *query*, *get* 等。

按照上述定义,对于一个飞机票预定服务(Book AirTicket-Service),它的基本描述可以表示为:

ServiceName: BookAirTicketService;

ServiceProvider: Northeastern University;

ServiceDiscription: Allows customers to order air tickets from Flight Company.

它的功能描述 C 可表示为:

$I = \{DepartureCity, ArrivalCity, SetOutDate, ReturnDate, CreditCard\}$

$O = \{Flight, ReservationID\}$

$F = \langle \text{airTicket}, \text{book} \rangle$

3 服务匹配

在服务匹配中,首先采用基于推理的方法对扩展的功能描述信息进行匹配,得出 4 个匹配等级,然后基于不同的匹配等级决定是否采用相似性计算方法进行服务接口信息的匹配。

3.1 扩展的功能描述信息匹配

由于 *object* 是服务施加操作的对象,只有满足 *object* 约束的服务才有可能满足服务请求,因此,依据 *object* 进行请求服务 R 与广告服务 S 之间的语义匹配,而动作部分的匹配则要求服务请求与服务广告的等价。

请求服务 R 的扩展功能描述: $\langle Ro, Ra \rangle$, 广告服务 S 的扩展功能描述: $\langle So, Sa \rangle$, 则扩展的功能描述信息的匹配等级如下:

(1) Exact match: 若服务 S 的操作对象和动作分别与请求服务 R 的操作对象和动作等价,则该匹配为等价匹配,可以形式地表示为: $Ro \equiv So \cap Ra \equiv Sa$ 。

(2) Plug-in match: 若服务 R 的操作对象是服务 S 的操作对象的子概念,则该匹配为插拔匹配,可以形式地表示为: $Ro \leq So \cap Ra \equiv Sa$ 。

(3) Subsume match: 若服务 S 的操作对象是服务 R 的操作对象的子概念,则该匹配为包含匹配,可以形式地表示为: $So \leq Ro \cap Ra \equiv Sa$ 。

(4) Mismatch: 若以上匹配都不满足,则匹配失败。

3.2 接口信息匹配

服务接口匹配即对输入/输出参数进行相似性比较以确定服务是否匹配。采用概念间语义相似性计算方法来度量请求与服务的接口匹配程度。概括地说,目前本体概念间的语义相似性计算方法主要有三种:基于属性、基于信息容量和基于几何距离的语义相似性计算方法。第一种方法通过概念的属性判断概念之间的区别和联系,认为两个概念间相同的属性越多,它们就越相似;第二种方法认为两个概念共享的信息越多,则两个概念的语义就越接近,语义相似度就越大,第三种方法的基本思想是用两个概念在本体中的几何距离来衡量语义相似度,也就是用两个概念在本体中的最短距离计算概念的语义相似度。文献[10]在综合考虑了多种因素之后提出了两个概念间相似度的计算公式,由于该方法实现简单,通用性好,因此选用该方法作为两个本体概念间的语义相似度的计算方法。两个概念

间相似度的计算公式:

$$SimCC(C_1, C_2) = \begin{cases} e^{-\alpha l} \cdot \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}}, & \text{if } (C_1 \neq C_2) \\ 1, & \text{if } (C_1 = C_2) \end{cases} \quad (1)$$

其中, l 表示两个概念之间最短路径的长度; h 表示两个概念在概念树中最近共同祖先的高度, $\alpha, \beta \geq 0$ 是调节因子, 根据文献[10]的实验, 当 $\alpha=0.2, \beta=0.6$ 时效果最好, 因此该文采用该值进行相似度的计算。

3.3 混合匹配算法

混合匹配算法如下。

- (1) MATCH
- (2) Inputs: a request R ;
a threshold ϕ ;
- (3) Outputs: a list results of services whose similarity with R is equal or greater than ϕ
- (4) begin
- (5) Load the list of advertisement services to AS
- (6) Results= Φ ;
- (7) for each $S \in AS$ {
- (8) TypeOfMatch:=logic_match_behavior(R, S);
- (9) if TypeOfMatch='exact'
- (10) GSim(R, S):=1;
- (11) if (TypeOfMatch='plug-in' OR TypeOfMatch='subsume')
- (12) GSim(R, S):=similarity_match_IO(R, S);
- (13) if TypeOfMatch='mismatch'
- (14) GSim(R, S):=0;
- (15) if GSim(R, S) $\geq \phi$
- (16) Results = Results \cup {($S, GSim(R, S)$)};
- (17) }
- (18) end for each
- (18) order services in results with DegreeOfMatch;
- (19) return Results;
- (20) end

首先, 运用基于逻辑推理的策略, 以服务动作的操作对象为核心, 将服务分为如 3.1 节所描述的 4 个等级(logic_match_behavior 函数), 然后, 根据以下规则, 使用基于相似性计算的匹配策略对服务的接口信息进行匹配(similarity_match_IO 函数), 得出具体的匹配度值。

(1) 如果 S 与 R 的扩展功能描述的语义匹配为 Exact match, 则将 $GSim(R, S)$ 设为 1 (full similarity), 将不再进行接口信息匹配;

(2) 如果 S 与 R 的扩展功能描述的语义匹配为 mismatch, 则将 $GSim(R, S)$ 设为 0, 将不再进行接口信息匹配;

(3) 如果 S 与 R 的扩展功能描述的语义匹配为 Plug-in match 或者 Subsume match, 则需进一步进行接口匹配。

当 $GSim(R, S)$ 大于或等于某个给定的阈值 ϕ , 则该服务为满足服务请求的服务, 最后这些服务将依据 $GSim(R, S)$ 进行排序。

4 实验与分析

首先, 在领域专家的指导下构建了包含 42 个概念的旅游本体, 从网上下载了 30 个与旅游相关的服务, 对这些服务采取人工方式进行了语义标注, 并添加了扩展的功能语义信息。

实验分别对基于接口的单一服务匹配方法和该文提出的基于扩展服务描述模型的混合匹配方法进行了测试, 需要说明

的是, 由于基于推理的匹配策略对服务的区分度低, 该实验中的单一服务匹配方法采用基于相似性计算的策略, 相似度计算算法均采用 3.2 节给出的算法。实验在 2.8 GHz Pentium 处理器和 512M 内存的微机上运行, 操作系统为 Windows XP Professional。通过实验对这两种匹配方法的查全率与查准率进行了比较, 查全率是指检索出的相关文档数和文档集中所有的相关文档数的比率, 查准率是指检索出的相关文档数与检索出的文档总数的比率^[1], 实验结果如图 1 所示。

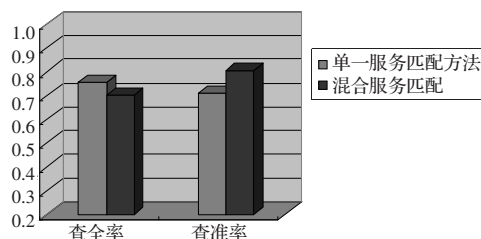


图 1 查全率与查准率的实验结果

实验结果表明基于接口的单一服务匹配方法和该文提出的方法发现平均查全率分别为 75.60% 和 70.11%, 平均查准率分别为 70.97% 和 80.62%, 实验采用的相似度阈值为 0.6, 该文方法的查全率略低, 这主要是由于服务扩展的功能描述信息中动作部分要求必须等价的强匹配造成的, 但在服务发现中认为查准率比查全率更重要。此外, 对这两种方法进行匹配的平均执行时间进行了对比, 其中单一服务匹配方法的平均执行时间为 32.96 ms, 而该文方法的平均执行时间为 24.74 ms, 显然基于混合服务匹配策略的服务发现方法更为简单、灵活。

5 结论和下一步工作

提出了一种基于扩展服务描述模型的语义 Web 服务混合匹配方法, 该方法首先对 Web 服务描述模型进行了扩展, 增强了 Web 服务模型的功能语义描述能力, 并基于此, 对扩展的功能语义信息和接口信息分别采用逻辑推理和相似性计算策略进行匹配。实验表明, 较之基于接口的单一服务匹配方法, 该方法提高了服务匹配的精确度和灵活性。

对于扩展的功能语义信息— $\langle \text{object}, \text{action} \rangle$ 中的 action 采取的是强匹配(exact), 如何深入挖掘动作之间的语义关系, 如何从服务的文本描述中自动提取更为丰富的功能语义信息, 使其更好地应用于服务发现, 以及如何实现自动语义标注, 这些问题是下一步的工作重点。

参考文献:

- [1] Sycara K, Paolucci M, Anolekar A, et al. Automated discovery, interaction and composition of semantic Web services[J]. Web Semantics: Science, Services and Agents on the World Wide Web, 2003, 1(1).
- [2] Li L, Horrocks IA. software framework for matchmaking based on semantic Web technology[C]//Proc of the 12th International World Wide Web Conference Workshop on E-Services and the Semantic Web (ESSW 2003), Budapest, Hungary, 2003: 331-339.
- [3] Colucci S, Noia T D, Sciascio E D, et al. Concept abduction and contraction for semantic-based discovery of matches and negotiation spaces in an e-marketplace[C]//Proc of the 6th International Conference on Electronic Commerce (ICEC 2004), Delft, The Netherlands, 2004, 60: 41-50.