

基于 Chord 的语义 Web 服务发现与合成的研究

赵红梅, 夏 阳, 赵 强, 李 琦

(中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116)

摘 要: 提出一种基于 Chord 的语义 Web 服务发现与合成模型. 将注册中心注册到分布式 Chord 网络中并基于领域本体对注册中心分类, 使用本体描述为 Web 服务加入语义信息, 语义 Web 弥补了传统 Web 语义不足的缺点. 实验分析表明, 该模型系统能提高 Web 服务发布、发现及合成的效率和准确率.

关键词: Web 服务; Chord; 领域本体; 服务合成; 服务发现

中图分类号: TP311

文献标识码: A

文章编号: 1000-7180(2010)08-0161-04

Research on Chord - Based Semantic Web Service Discovery and Composition

ZHAO Hong-mei, XIA Yang, ZHAO Qiang, LI Qi

(School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: This article presents a Chord - based Semantic web service discovery and composition model. It will be registered to a distributed registry Chord network, classified the registration center based on domain ontology, and using the ontology to describe web services for adding semantic information. Semantic web makes up the shortcomings of traditional web services for lack of semantics. Experimental analysis shows that the model improves the accuracy and efficiency of web service publishing, discovery and composition.

Key words: web services; Chord; domain ontology; services composition; services discovery

1 引言

目前, Web 服务合成技术主要分为四类: 申明性的 Web 服务合成、基于 workflow 模式、基于 QoS 模式和基于本体的 Web 服务合成^[1]. SELF - SERV 是采用对等服务协调模型的服务合成系统, Web 服务被申明性的合成, 并在一个动态的 P2P 环境中执行. 申明性的合成方法是一种静态的 Web 服务合成, 缺乏灵活性. 文献[2]提出一种基于静态 workflow 模式的服务复合平台 EFlow. 文献[3]提出一种静态 workflow 模式和动态 workflow 模式相结合的服务复合机制 PPM. 使用 workflow 技术合成 Web 服务, 通常需要预先设计好合成流程, 在实际应用过程中存在缺陷. 美国伊利诺大学的 Xiaohui Gu 等研究者^[4]提出一个 QoS 感知的服务合成框架 - SpiderNet. SpiderNet 服务合成模型可

以通过动态地合成和实例化服务路径来实现分布式应用合成. 基于 QoS 的 Web 服务合成由于缺乏对语义的利用, 降低了服务合成的准确率. 美国乔治亚大学的 METEOR - S 项目的 Web 服务合成框架^[5] (METER - S Web Service Composition Framework, MWSCF) 通过语义活动模板描述了一个活动, 然后在两个维度上 (语义匹配和 QoS 标准匹配) 对 Web 服务的总体排行进行了衡量. 在 Web 服务合成过程中, 使用本体可以提高服务合成的效率, 实现 Web 服务合成的自动化与智能化.

借鉴国内外 Web 服务发现与合成技术, 文中提出一种基于 Chord 的语义 Web 服务发现和合成模型, 将 Chord 和本体技术与 Web 服务技术有机结合, 以解决集中式架构中存在的问题并提高 Web 服务发现和合成的效率、准确度与智能化.

2 基于 Chord 的语义 Web 服务发现与合成原型

2.1 原型系统模型

如图 1 所示,基于 Chord 的语义 Web 服务发现与合成原型系统模型包含:服务提供者、服务请求者、服务社区、超节点.服务社区是提供相同或相似服务内容的一群服务提供者.超节点是某服务社区中功能较强大的成员^[6].

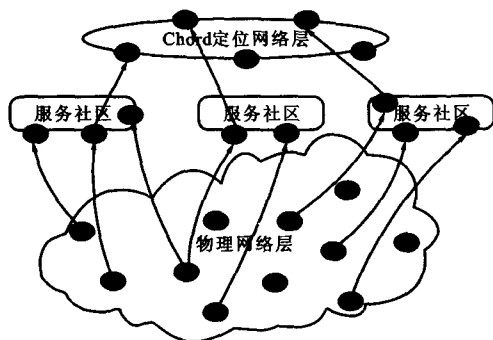


图 1 基于 Chord 的语义 Web 服务发现与合成原型

这个原型大致分为三层,从下往上依次是,物理网络层,服务社区及定位网络层.物理层是按照地理位置进行划分的实际的拓扑;基于领域本体构建了不同的服务社区,不同的服务社区提供不同的服务类型,共同构成结构化的 P2P 网络;所有的服务社区中的超节点以 Chord 结构为基础,构成了定位网络层.

在基于 Chord 的语义 Web 服务发现与合成原型系统中,每个节点可以提供一些服务,每个服务根据其服务类型归属于某个特定的领域(即服务社区).每个节点至少是一个服务社区的成员.每个服务社区都有一个超节点 MasterPeer 和一个备份节点 BackupPeer,MasterPeer 中有其他社区的 MasterPeer 和 BackupPeer 列表,而 BackupPeer 则存有这一列表的备份.

2.2 Chord 定位网络层

基于 Chord 的语义 Web 服务发现与合成原型系统的路由层采用 Chord 协议.注册中心的 URL 被散列到 Chord 关键字区间的一个整数. Chord 在 2001 年由麻省理工学院提出,是典型结构化 P2P 系统. Chord 实现了这样一种操作:给定一个关键字 K ,将 K 映射到某个节点. Chord 协议是基于分布式哈希表(DHT)的,其拓扑结构是一个闭合的 m -bit 标识环,编号范围为 $[0, 2^m - 1]$. Chord 中每个数据和节点都分别拥有一个 m 比特位的标识符.数据标识符 K 通过哈希数据本身得到,而节点标识符 N

通过哈希节点的 IP 地址得到.图 2 给出了一个 $m=3$ 的 Chord 定位网络层结构,标识空间是 $2^m=8$,环中分布了 0、3、6 三个实际存在的物理节点,每个节点的指针表如表 1 所示.

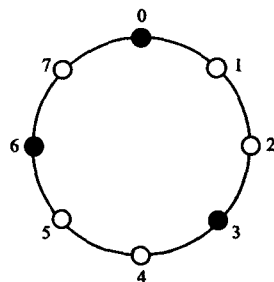


图 2 定位网络层的 Chord 环结构

表 1 各个节点的指针表

Node	Startnode	Interval	Nextnode	IP
节点 0	1	[1,2)	3	192.168.100.3
	2	[2,4)	3	192.168.100.3
	4	[4,0)	6	192.168.100.6
节点 3	4	[4,5)	6	192.168.100.6
	5	[5,7)	6	192.168.100.6
	7	[7,3)	0	192.168.100.10
节点 6	7	[7,0)	0	192.168.100.10
	0	[0,2)	0	192.168.100.10
	2	[2,6)	3	192.168.100.3

结构化 P2P 系统 Chord,解决了集中式架构中存在的单点故障等诸多问题,避免了非结构化 P2P 系统中使用的泛洪式查找,因此提高了信息搜索与查询的效率.

2.3 相关本体

领域本体提供了该领域内概念定义及概念之间的关系.采用本体对 Web 服务进行描述和理解,通过构造特定领域本体,从而建立领域本体与服务之间的关联.

定义 1 领域描述本体可以用五元组表示: $O = \{C, P, R, H, X\}$,其中 C 表示概念的集合; P 表示属性的集合,本体的属性具有作用域和值域,映射了概念间的二元关系; R 表示关系的集合; H 表示继承关系集,包括概念继承和属性继承,即 subClassOf 和 subPropertyOf ; X 为本体规则集,或称本体约束集,可以采用一阶谓词逻辑或者描述逻辑形式化地进行描述.

定义 2 一个 Web 服务用下面的表达式来描

述: $WS = \{S, Input, Output, QoS\}$, 其中 S 是基本描述, 即服务名称、文本描述、服务类别等; $Input$ 是服务输入参数; $Output$ 是服务输出参数; QoS 是属性描述, 主要包括代价(cost)、响应时间(time)、服务优先级(priorities)等非功能属性。

在基于 Chord 的语义 Web 服务发现原型系统中, 服务社区的结构与本体的结构一致(本体是层次化的, 因而服务社区结构也是层次化的)。原型系统中的相关本体主要是领域描述本体和服务描述本体。领域描述本体是对特定领域的总的拓扑结构, 其存储在 Chord 定位网路层(即服务社区的超节点内), 用于将服务定位到相关的服务社区。服务描述本体是对 Web 服务的语义描述, 包括功能性描述和非功能性描述(QoS), 其存储在发布或查找服务的节点里。功能性描述包括 Web 服务的输入、输出、前提和结果; 非功能性描述(QoS)主要有代价(cost)、响应时间(time)、服务优先级(priorities)等。

3 基于 Chord 的语义 Web 服务发布、发现与合成

服务发布、发现与合成的交互时序如图 3 所示。

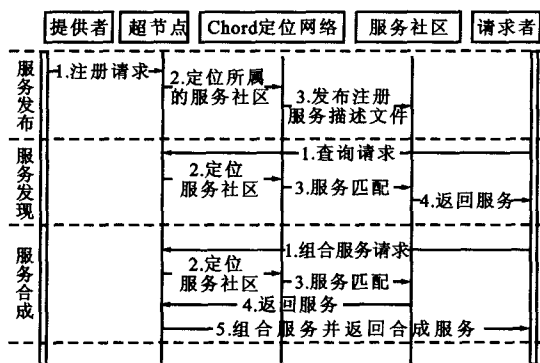


图 3 服务发布、发现与合成的交互时序图

3.1 服务发布

首先服务提供者向所在服务社区的超节点提交注册请求信息; 该超节点根据注册服务的类型, 判断是否是本服务社区的服务类型, 是则在该超节点注册, 不是则需将服务映射到其相关的服务社区, 根据 Chord 协议找到相关服务社区的地理位置(IP 地址); 在相关的服务社区内发布注册服务描述文件。

3.2 服务发现

首先服务请求者向所在服务社区的超节点提交查询请求信息; 该超节点根据请求服务类型, 判断是否是本服务社区的服务类型, 是则在本社区内查找, 不是则需将请求服务映射到其相关的服务社区, 根据

Chord 协议找到相关服务社区的地理位置(IP 地址); 在相关的服务社区内进行服务匹配操作; 返回服务。

文献[7]中提出了二阶段查询方法和两层搜索机制, 本文借鉴并进行改进, 提出了三阶段匹配算法用于服务匹配操作。该算法首先进行服务类型的匹配, 进而定位所属的服务社区, 这样可以大大缩小服务查找的范围, 提高效率; 再次进行基于接口的匹配, 服务社区里的目标子服务的输入和输出接口要与请求服务的输入和输出接口匹配, 提高查准率; 最后基于 QoS 匹配, 找到服务质量最优的 Web 服务并满足用户偏好。

3.3 服务合成

首先服务请求者向所在服务社区的超节点提交组合服务请求信息; 该超节点对组合服务请求信息进行解析和分解, 根据服务类型定位到服务社区, 根据三阶段匹配算法在服务社区中找到合适的目标子服务并返回到超节点; 在超节点内动态生成服务组合图, 并根据 Web 服务组合算法对目标子服务进行筛选与合成; 返回合成服务。

Web 服务组合算法的中心思想是在领域本体的支持下, 以概念相似度为基础, 计算服务之间的关联度, 根据服务接口的输入和输出形成一个服务组合图, 利用该服务组合图的特点, 进行高效的 Web 服务组合^[8]。在基于 Chord 的语义 Web 服务发现与合成原型系统中, Web 服务组合算法是一个难点, 需要解决服务之间的关联度计算问题, 以及各个服务接口的输入和输出的匹配和衔接操作。

4 实验与分析

针对文中提出的基于 Chord 的语义 Web 服务发现和合成模型, 对其进行了初步的模拟实验。本实验的实验环境是 P41.6 GHz CPU、512 MB 内存和 80 G 硬盘。软件开发工具为 Apache Axis、OWL 推理机 Racer、Jena 2.0 API、OWL-S API 和 Tomcat 5.0 服务器。领域本体的创建在 Protégé 工具中进行。本实验构造的服务均采用 WSDL 描述, 并发布成语义 Web 上的 OWL-S 服务。查全率 δ 的定义为 $\delta = \frac{|S \cap T|}{|S|}$, 查准率 η 的定义为 $\eta = \frac{|S \cap T|}{|T|}$, 其中 S 是标准结果集, T 是返回结果集^[7]。实验模拟服务数分别是 400、800、1200 和 1600 时, 计算并比较基于 Chord 的语义 Web 服务发现和合成方法和集中式服务发现方法的平均查全率及查准率。

图 4 和图 5 的实验结果分析表明, 基于 Chord

的语义 Web 服务发现和合成方法的查全率低于集中式服务发现方法,但查准率高于集中式服务发现方法,且查询效率明显高于集中式服务发现方法,这是因为基于 Chord 的语义 Web 服务发现和合成方法首先根据请求服务的类型定位到相关的服务社区中,减少了与不相关服务的查找匹配时间。

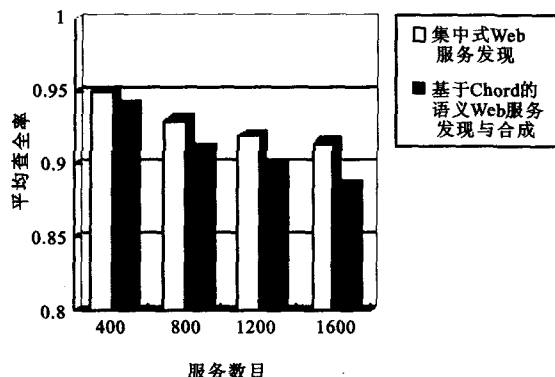


图4 查全率比较

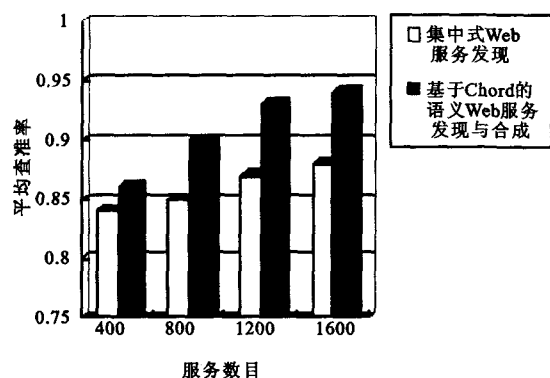


图5 查准率比较

5 结束语

文中提出一种基于 Chord 的语义 Web 服务发现和合成模型,将 Chord 和本体技术与 Web 服务技术有机结合。Chord 有效解决集中式架构存在的诸多问题;本体为服务描述加入语义信息,提高 Web 服务发现和合成的准确度与智能化;注册中心分布在 Chord 网络中并基于领域本体分类,增强了 Web 服务发现及合成的效率。目前语义 Web 服务合成的成功率和效率有待改善,计算 Web 服务概念相似度与优化 Web 服务组合算法是下一步研究工作的重点。

参考文献:

[1] Federica Mandreoli, Antonio M Perdicizzi, Wilma Penzo.

A P2P-based architecture for semantic web service automatic composition[C]// 18th International Workshop on Database and Expert Systems Applications. Germany, 2007:429-433.

- [2] Casati F, Ilnicki S, Jin L. Adaptive and dynamic service composition in EFlow[C]// Proceedings of 12th International Conference on Advanced Information Systems Engineering (CAiSE). Stockholm, Sweden: Springer Verlag, 2000.
- [3] Schuster H, Georgakopoulos D, Cichocki A, et al. Modeling and composing service-based and reference process-based multi-enterprise processes[C]// Proceeding of 12th International Conference on Advanced Information Systems Engineering (CAiSE). Stockholm, Sweden: Springer Verlag, 2000.
- [4] Gu Xiaohui, Nahrstedt K, Yu Bin. Spidernet: an integrated peer-to-peer service composition framework[C]// Proc. of IEEE Int. Symp. on High-performance Distributed Computing. Honolulu, USA: IEEE Press, 2004: 110-119.
- [5] Kunal Verma, Kaarthik Sivashanmugam, Amit Sheth, et al. METEOR-SWSDI: a scalable P2P infrastructure of registries for semantic publication and discovery of web services[J]. Journal of Information Technology and Management, 2005, Special Issue on Universal Global Integration: 17-39.
- [6] Mohamed Gharzouli, Mahmoud Boufaïda. A generic P2P collaborative strategy for discovering and composing semantic web services[C]// 2009 Fourth International Conference on Internet and Web Applications and Services. Italy: Venice, 2009:449-454.
- [7] 陈德伟,许斌,蔡月茹,等.服务部署与发布绑定的基于 P2P 网络的 Web 服务发现机制[J].计算机学报,2005, 28(4):615-626.
- [8] 李曼,王大治,杜小勇,等.基于领域本体的服务动态组合[J].计算机学报,2005,28(4):644-650.

作者简介:

赵红梅 女,(1985-),硕士研究生.研究方向为语义 Web 服务、电子商务。

夏阳 男,(1962-),博士,硕士生导师.研究方向为网络计算、电子商务。

赵强 男,(1985-),硕士研究生.研究方向为 Web 服务、语义 Web。

李琦 女,(1980-),硕士研究生.研究方向为 Web 服务、电子商务。