

基于 Hadoop 的 Web 服务语义可信 QoS 发现模型

何小霞¹ 谭良^{1,2}

(四川师范大学计算机科学学院 成都 610068)¹ (中国科学院计算技术研究所 北京 100190)²

摘要 随着 Web 服务应用的快速增长,用户如何在众多功能相似的 Web 服务中更加准确地选择出满足自己 QoS 需求的 Web 服务是一个亟需解决的问题。目前已有的研究工作存在两个问题,其一是对 QoS 的量化存在客观性和准确性问题,其二是对服务质量缺乏语义描述,QoS 匹配时也仅限于数值上的匹配。针对这两个问题,利用 Hadoop 的分布式注册、查找服务模式,提出了基于 Hadoop 的 Web 服务语义可信 QoS 发现模型。本模型从数值匹配和 QoS 语义匹配两方面考虑,首先,用主客观赋权模式为多维 QoS 属性赋权值,提高 QoS 属性的客观性和准确性,并加入信誉度参数来提高 Web 服务 QoS 属性的可信性;其次,对 OWL-S 进行了 QoS 本体扩展,以满足客户对服务质量的语义匹配需求。通过验证表明,本模型能有效解决查询瓶颈和单点失效问题,且能根据用户 QoS 需求偏好更加准确地找到满足用户需要的 Web 服务。

关键词 Web 服务, Hadoop, 语义, QoS 本体, QoS 权重, 可信

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.8.045

Web Service Discovery Model of Semantic and Trust QoS Based on Hadoop

HE Xiao-xia¹ TAN Liang^{1,2}

(College of Computer Science, Sichuan Normal University, Chengdu 610068, China)¹

(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)²

Abstract With the rapid growth of the Web service applications, how to select Web service meeting users' QoS requirements more accurately is an emergent problem from many Web services with similar functions. There are two problems in the research work; one is the QoS quantitative problems of objectivity and accuracy, and the other is that the QoS match is also limited to the matching of numerical value and the lack of semantic description for quality of service. To solve the two problems, using the Hadoop distributed registration and service-seeking mode, we designed the Web service discovery model of semantic and trust QoS based on Hadoop. In the pattern, considering two aspects of QoS semantic match and numerical match, first, we used subjective and objective weighting mode to empower multi-dimension QoS properties, improving the objectivity and accuracy of QoS properties, and added credibility parameters to increase the trust of the Web service QoS attributes. Second, we extended the QoS ontology on the OWL-S to meet customer demand for the semantic matching of QoS. Experimental results show that this model can effectively solve the problems of query bottleneck and single point of failure, and accurately find Web services meeting the needs of users according to users' QoS requirements.

Keywords Web service, Hadoop, Semantic, QoS ontology, QoS weights, Credibility

1 引言

Web Service 是一种基于 XML 和 HTTPS 的面向服务架构 SOA 的分布式计算技术,其通信协议主要基于 SOAP (Simple Object Access Protocol),以 WSDL (Web Services Description Language) 作为其服务描述语言,并通过 UDDI (Universal Description, Discovery and Integration) 来发现和获得服务的元数据,是一种自描述的、模块化的、由 URI 标识的应用程序^[1-3]。

随着 Web 服务商业应用的快速增长,用户在使用 Web

服务时不仅注重服务的功能,而且更加注重服务的非功能需求,即 Web 服务的服务质量(Quality of Service, QoS),包括 Web 服务的可靠性、可用性、可扩展性、安全性、响应时间、服务价格等^[11,13,17]。用户如何在众多功能相似的 Web 服务中更加准确地选择出满足自己 QoS 需求的 Web 服务是一个亟需解决的问题。

云计算是在宽带网络、互联网应用服务、并行计算与分布式计算以及负载均衡、虚拟化技术日趋成熟的基础上发展而来的^[5],当前比较有代表性的是 Google 的 Hadoop 项目, Hadoop 是 Apache 开源组织的一种可靠、高效、可伸缩的分布式

到稿日期:2014-08-11 返修日期:2014-10-21 本文受国家自然科学基金(61373162),四川省科技支撑项目(2014GZ007)资助。

何小霞(1989—),女,硕士生,主要研究领域为中间件,E-mail:hxx_mail@163.com;谭良(1972—),男,博士,教授,CCF 高级会员,主要研究领域为可信计算、网络安全,E-mail:tanliang2008cn@126.com。

并行编程框架,它主要包括 MapReduce、HDFS(Hadoop Distributed File System)分布式文件系统和 HBase(Hadoop Database)分布式存储系统^[6]。Hadoop 不仅能够提供一种易扩展、容错性高的方式来存储服务,而且可以通过分布式并行计算框架来提高服务查找速度,能够很好地满足海量 Web 服务管理的要求^[4]。但随着各种商业应用对 Web 服务性能要求的提高,用户对 Web 服务 QoS 查找准确性要求也在提高。目前也有学者开展了这方面的研究工作,其提出采用多维 QoS 属性的全局最优或局部最优选择模式^[13],从众多功能相似的 Web 服务中选择满足用户非功能性需求的服务。采用此方法存在两个方面的问题:1)对 QoS 的量化存在客观性和准确性问题,容易造成服务欺骗,给用户的业务功能造成不良影响,如果采用此方法选择的服务应用于一些关键的业务系统,如电子商务、金融或税务,会给用户造成巨大的损失;2)对服务缺乏语义描述,QoS 匹配时也仅限于数值上的匹配,容易盲目排除匹配不一致的候选服务,使得一些高质量的同类服务不能及时提供给用户。

本文针对这两个问题,利用 Hadoop 的分布式注册、查找服务模式提出了基于 Hadoop 的 Web 服务语义可信 QoS 发现模型。本模型从数值匹配和 QoS 语义匹配两方面考虑,首先,用主客观赋权模式为多维 QoS 属性赋权值,提高 QoS 属性的客观性和准确性,并加入信誉度参数来提高 Web 服务 QoS 属性的可信性;其次,对 OWL-S 进行了 QoS 本体扩展,以满足客户对 QoS 服务质量的语义匹配需求。通过验证表明,本模型能有效解决查询瓶颈和单点失效问题,且能根据用户 QoS 需求偏好更加准确地找到满足用户需要的 Web 服务。

本文第 2 节介绍相关工作;第 3 节详细介绍了基于 Hadoop 的语义可信 Web 服务发现模型;第 4 节用实验验证了本文提出的模型能有效提高 Web 服务查找的速率和准确性,避免了服务注册瓶颈和单点失效的问题;最后总结全文。

2 相关工作

文献[7]提出一种基于 Hadoop 的 Web 服务管理框架,该框架中 HDFS 作为支持 HBase 和 MapReduce 的基础,HBase 用于管理 Web 服务功能性和非功能性信息,且使用 QoS 索引机制来实现非功能性信息的管理和查找,使用 MapReduce 编程实现 Web 服务查找;不足之处是没有对 QoS 权重选择作说明,无法根据客户的选择偏好提供可供选择的服务。文献[9]提出基于 QoS 选择的云计算 Web 服务发现模型,该模型使用云技术框架,改善了传统集中式 Web 服务存储方式引起的单点失效和瓶颈问题,并引入 QoS 属性从众多相似的 Web 服务中过滤掉不满足用户非功能性需求的 Web 服务。但其不足之处是没有考虑 QoS 的真实性和动态性,导致反馈给客户的服务中有很多的不确定性。文献[12]提出一种基于云模型的不确定性 QoS 感知的 Skyline 服务选择方法,该方法通过云模型计算 QoS 的不确定性,然后采用 Skyline 计算提取 Web 服务中的 Skyline 服务,剔除冗余服务,最后采用混合整数规划在 Skyline 服务中进行服务选择。其满足了用户全局 QoS 约束的最优服务组合,但不能反映用户的选择偏好,无法满足现在消费者的个性需求。文献[13]在 QoS 本体构建的基础上提出一种语义 Web 服务发现模型,在

传统 UDDI 集中式管理模式上使用 QoS 本体、QoS 数据测评机制和 QoS 选择机制对 QoS 属性选择方法进行研究。但是其不足之处是传统的 UDDI 方式容易出现瓶颈和单点失效的问题,也没有对 QoS 信息的可信度做评估,从而无法满足用户对 QoS 属性的偏好需求。

上述文献都对 Web 服务的发现效率和准确性有一定的贡献,但是都没有很好地综合解决 Web 服务发现存在的两大问题:1)对 QoS 的量化存在客观性和准确性问题;2)对服务质量缺乏语义描述,QoS 匹配时也仅限于数值上的匹配,容易盲目排除匹配不一致的候选服务。所以本文针对这些问题提出了基于 Hadoop 的 Web 服务语义可信 QoS 发现模型,利用 Hadoop 的存储优势解决单点失效和瓶颈等问题,且从 QoS 数值匹配和 QoS 语义匹配两方面考虑。首先,用主客观赋权模式为多维 QoS 属性赋权值,提高 QoS 属性的客观性和准确性,并加入信誉度参数来提高 Web 服务 QoS 属性的可信性;其次,对 OWL-S 进行了 QoS 本体扩展,以满足客户对于服务质量的语义匹配需求。

3 基于 Hadoop 的语义可信 Web 服务发现模型

为解决以上问题,本文以基于 Hadoop 的 Web 服务管理框架^[7]为基础,提出了一种基于 Hadoop 的 Web 服务语义可信 QoS 发现模型。为了提高 Web 服务的查找准确性和可信性,本文在 Web 服务功能语义描述相似的前提下使用信誉度来控制 Web 服务 QoS 属性的真实性;并且构建 QoS 本体,从 QoS 语义匹配和数值匹配两方面考虑,使用 3.2 节中的主客观赋权模式方法为 QoS 属性赋权值,提高服务 QoS 属性查找准确率,避免盲目排除匹配不一致的候选服务,同时采用 Hadoop 平台也可提高 Web 服务的发现速率,解决单点失效和瓶颈等问题。基于 Hadoop 的语义可信 Web 服务发现模型如图 1 所示。

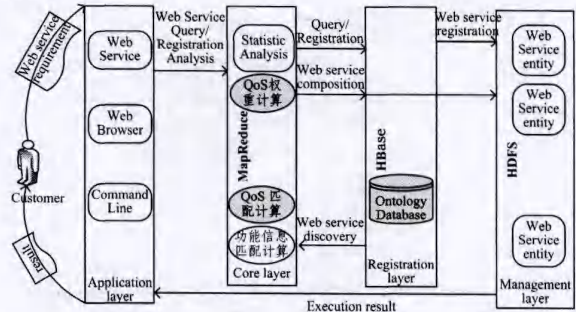


图 1 系统模型

本模型分为 4 层:应用层(Application Layer)、中心层(Core Layer)、注册层(Registration Layer)和管理层(Management Layer)^[10],各层的主要作用描述如下:

应用层:主要作为展示 Web 服务描述信息的接口,方便客户进行 Web 服务的发现和选择操作。

中心层:该层的主要组件是 MapReduce,用于处理 Web 服务功能属性和 QoS 属性的语义信息转换和匹配计算,相比文献[7]该层增加了 QoS 权重计算模式和 QoS 匹配计算,而 QoS 匹配计算又分为 QoS 语义匹配和数值匹配,用于接收从注册层中返回的 QoS 属性,并计算满足用户需求的 Web 服务对应的 QoS 属性权值。

注册层:该层的主要组件是 HBase,用于存储 Web 服务

的功能性信息和非功能信息及其对应的索引。HBase 表由于是疏松地存储的,一个数据行有一个可选择的键和任意数量的列,因此很适合用于管理这些松散的数据。相比文献[7],本文使用 QoS 本体的匹配索引,并采用主客观赋权模式加入了 QoS 属性的可信权重和 QoS 属性信誉度,可有效提高 Web 服务 QoS 属性的可信性和查找准确性,避免了盲目排除匹配不一致的候选服务。

管理层:该层主要由 HDFS 构成,由于分布式环境的高可扩展性和可靠性,它能为 Web 服务的执行提供保证,除了提供 Web 服务的部署外,同时还包括相关的算法的部署以及处理过程中产生的数据和元数据信息的管理。

该模型的 Web 服务发现工作流程如图 2 所示:

(1)用户通过应用接口层向服务注册端发送 Web 服务选择要求。

(2)经过 MapReduce 中心层,为用户的选择要求做分析,并将任务分配给分布式中的任务服务器进行分布式运算,从注册层中寻找满足用户功能性需求的 Web 服务信息描述并返回给中心层。

(3)中心层进行服务功能信息的语义匹配计算,并得到一组满足用户功能需求的候选服务组一,再从注册层中找出这组候选服务的 QoS 信息,根据 3.3 节中的 QoS 语义匹配法排除一些 QoS 语义不相关的 Web 服务,得到了候选服务组二,再根据 3.2 节中的主客观赋权模式和信誉度算法计算该服务组的 QoS 属性权重和信誉值,从而得到 QoS 的语义相似度匹配,即可对候选服务组二进行排序。

(4)根据优先顺序从管理层中将 Web 服务经应用接口层返回给用户,自此就完成了 Web 服务发现的基本流程。

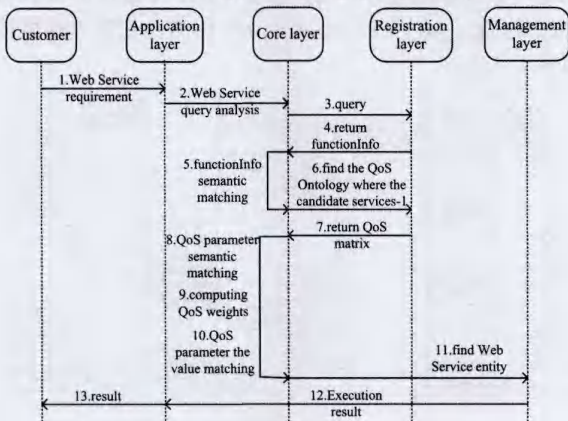


图 2 Web 服务发现交互时序图

在如图 2 所示的流程中,最为关键的是实现基于主客观赋权模式的可信 QoS 权重、QoS 本体和 QoS 语义相似度匹配。本文将在后面一一介绍。

3.1 QoS 本体

本文采用 OWL-S 来描述服务的语义信息。但 OWL-S 缺乏对服务质量的语义描述,不同服务参与方对质量的概念描述与度量方式不统一,因此本文基于 OWL-S Profile 扩展了 QoS 本体,统一了服务质量的描述和度量方式。

定义 1 服务质量模型 $QoS(WS) = \langle RT, A, R \rangle$ 。RT (Response Time)代表服务的响应时间,A (Availability)代表服务的可用性,R (Reliability)代表服务的可靠性。本文在参

考文献[8]的基础上,利用 OWL-S Profile 定义中 ServiceParameter 扩充机制引入 QoS 本体的定义,来描述服务质量。其 QoS 本体模型如图 3 所示。

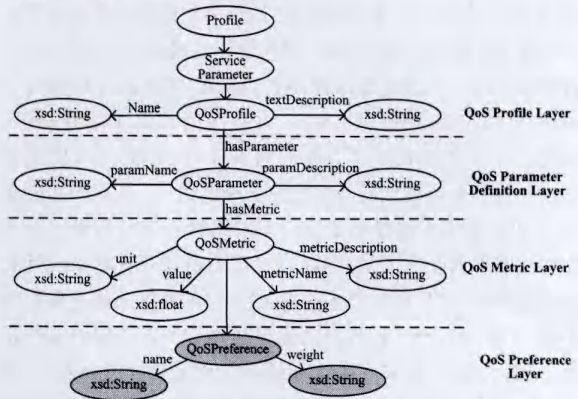


图 3 QoS 本体模型

根据服务质量模型的定义,引入的 QoS 本体应该具有 RT、A 和 R 3 个属性,以及根据客户的需求偏好而加入的 QoS Preference Layer,后者主要用于 QoS 参数权重计算并根据偏好做出选择。3.2 节主要对 QoS 本体中 QoS Preference Layer 的 QoS 权重和信誉度计算做详细说明,3.3 节主要从 QoS 参数的语义匹配和数值匹配两方面分别对 QoS 本体中 QoS Parameter Definition Layer 和 QoS Metric Layer 做出说明。

3.2 基于主客观赋权模式的可信 QoS 属性匹配

定义 2 n 维服务质量属性可信模型 $QoS = \{q_1, q_2, \dots, q_n\} \cup Rep(S)$,其中 $Rep(S)$ 表示服务 S 的信誉度。

定义 3(信誉度定义) 设多维服务质量模型是一个 n 维可扩充的向量: $QoS = \{q_1, q_2, \dots, q_n\}$,其服务 S 的信誉度 Rep (Reputation)是用户对该服务的 n 维服务质量属性的评价,表示为

$$Rep(S) = \frac{\sum_{i=1}^n Rep_{q_i}(S)}{n} \quad (1)$$

$$Rep_{q_i}(S) = \frac{\sum_{j=1}^{m_i} r_{ij}(S)}{m_i} \quad (2)$$

式(2)表示用户对第 i 个 QoS 属性的评价, $r_{ij}(S)$ 表示第 i 个 QoS 属性被第 j 次访问给出的评价, m_i 表示第 i 个 QoS 属性被评价的总次数。引入信誉度的主要目的是控制服务质量属性的可信性。

在多维 QoS 属性中选择优质服务的关键是确定每个 QoS 属性的权重^[14,15]。QoS 属性的权重可采用主观赋权模式和客观赋权模式,主观赋权模式根据用户自己的需求确定各属性的权重,虽然能体现用户的偏好,但其评价结果具有较大的主观随意性;客观赋权模式通过较强的数学理论依据确定各属性间的权重,其结果不依赖于人的主观判断,但没有体现用户偏好,有时候确定的权重与属性的实际重要程度相差较大^[16,17]。因此,本模型结合主观赋权模式和客观赋权模式两者的长处,采用主客观赋权模式来确定属性的权重,以提高服务质量的查找准确性。

定义 4(QoS 主观赋权模式权重) 是由用户自己确定的

每个 QoS 属性的权重,使用 ω_j 表示,如式(3)所示:

$$\omega_j, \sum_{j=1}^n \omega_j = 1, \omega_j \geq 0, j=1, 2, \dots, n \quad (3)$$

定义 5(QoS 客观赋权模式权重) 由系统给出推荐值,该推荐值是系统根据该服务被用户使用后记录的经验值的平均值来作为权重值,使用 ω_j' 表示,如式(4)所示:

$$\omega_j' = \frac{\sum_{i=1}^k \omega_{ij}}{k}, \omega_j' \geq 0, j=1, 2, \dots, k \quad (4)$$

其中, k 表示该 QoS 属性被访问的经验次数, ω_{ij} 表示该属性每次被访问时记录下来的权重需求值。

定义 6(QoS 可信权重) 表示为 ω_j^* , 如式(5)所示:

$$\omega_j^* = \alpha\omega_j + \beta\omega_j', j=1, 2, \dots, n \quad (5)$$

其中, α, β 表示主观和客观赋权模式的相对重要程度,且满足 $\alpha + \beta = 1, 0 < \alpha, \beta < 1, \omega_j^*$ 即为主客观赋权模式表示下该服务 QoS 属性 q_j 对应的权重值。使用 QoS 本体模式存储 QoS 信息,并经过以上算法精选出满足用户偏好且可信的 Web 服务质量属性。其中使用 3.3 节中的 QoS 数值匹配将所有的 QoS 属性值转换为 $[0, 1]$ 之间的值,如响应时间为 5s,需要将此数值相应地除以 10,变为 0.5,这样该数值就在 $[0, 1]$ 范围之内;QoS 属性分为积极属性和消极属性两类^[18],积极属性考虑 QoS 的最大化,如吞吐量、可靠性、可用性、可扩展性等;消极属性考虑 QoS 的最小化,如响应时间、服务价格等^[19]。当 QoS 为消极属性时将其属性值变为负数,通过 3.3 节式(7)得到的 QoS 语义相似度最高的服务即为最满足客户需要的最好的 Web 服务。

3.3 QoS 语义相似度匹配

Web 选择宗旨是在满足用户服务功能需求的基础上,再从服务质量角度对服务进一步区分选择。本文通过 3.2 节中关于 QoS 选择权重的叙述,从服务发布和请求双方质量参数的语义匹配和参数的数值匹配两方面进行 QoS 语义匹配和选择,双方最终的 QoS 语义相似度计算方式如下:

$$QoSSim(Req, Adv) = \frac{\sum_{i=1}^n Q_iSim(Req, Adv) \times \omega_i \times q_iSim(Req, Adv)}{\sum_{i=1}^n Q_iSim(Req, Adv)} \quad (6)$$

$$\max_{i=1}^n QoSSim_i(Req, Adv) \quad (7)$$

其中, $QoSSim(Req, Adv)$ 为总体语义相似度,用户可根据式(7)选择值最大的为最优服务, $Q_iSim(Req, Adv)$ 为匹配双方第 i 个 QoS 参数的语义相似度, $q_iSim(Req, Adv)$ 为第 i 个参数的数值相似度,具体描述如下:

(1) QoS 参数的语义匹配:由 Massimo Paolucci^[11] 经典语义区分,可将 QoS 参数匹配等级做如下 4 种划分:

- 1) Adv 和 Req 的 QoS 参数相同或语义等价,则 $Q_iSim(Req, Adv) = 1$;
- 2) Adv 是 Req 的 QoS 参数的直接父类,则 $Q_iSim(Req, Adv) = 0.9$;
- 3) Req 是 Adv 的 QoS 参数的直接父类,则 $Q_iSim(Req, Adv) = 0.8$;
- 4) 其他情况,则 $Q_iSim(Req, Adv) = 0$ 。

过滤掉 $Q_iSim(Req, Adv) < 0.8$ 的候选服务,提高服务查找效率。

(2) QoS 参数的数值匹配:根据服务请求和发布的 QoS

数值构造数值匹配矩阵 $QM(q_{ij})_{(n+1)m}$:

$$QM = \begin{bmatrix} q_{11} & q_{12} & \dots & q_{1m} \\ q_{21} & q_{22} & \dots & q_{2m} \\ \dots & \dots & \dots & \dots \\ q_{n1} & q_{n2} & \dots & q_{nm} \\ q_{(n+1)1} & q_{(n+1)2} & \dots & q_{(n+1)m} \end{bmatrix} \quad (8)$$

其中,第一行为服务请求者数值,第 2 至 $(n+1)$ 行分别为 Web 服务经过功能语义筛选和 QoS 语义匹配之后得到的 n 个候选服务 QoS 参数值, m 为服务参数个数,匹配矩阵的标准化包括 3 方面的转换:1)取值类型转换,将 QoS 参数的不同取值类型进行相应的转换,如将布尔型取值 TRUE/FALSE 转换为 1/0;2)数值取向转换,将其压缩至区间 $[0, 1]$;3)单位转换,如吞吐量可能采用 MB/s 或 kB/s 度量,需进行统一换算。这样得到一个标准的 QoS 数值匹配矩阵,并根据 3.2 节中主客观赋权模式得到的 QoS 相应权重 ω_i ,可以得到这 n 个候选服务的 $QoSSim_i(Req, Adv)$ 值,用户可根据其值大小选择适合自己需求的 Web 服务。

4 实验验证

本节将通过实验对所提出的基于 Hadoop 的 Web 服务语义可信 QoS 发现模型进行验证。在满足用户功能性语义需求的前提下,基于 QoS 本体将 QoS 匹配分成两大步骤完成:基于 QoS 参数的语义匹配和基于 QoS 参数的数值匹配,最终可根据式(7)选择 $QoSSim_i(Req, Adv)$ 最大或较大的服务为最佳的满足用户功能性需求和服务质量需求的 Web 服务,且该服务还通过 QoS 参数权重及信誉度,满足了客户的选择偏好,提高了 QoS 参数的可信度。设使用 3 个常用的 QoS 属性:响应时间(RT)、可用性(A)和可靠性(R),其中响应时间以秒为单位计算,可用性和可靠性以百分数来计量。该实验证明了本方案的有效性和可行性。

4.1 实验环境

实验所用环境: Intel 酷睿 i3 2120 处理器、4GB 内存、Linux 操作系统、Hadoop 云平台、JDK 1.7。

4.2 实验分析及实验结果

本实验准备了 10000 个功能性相似的 Web 服务,其各自的 QoS 属性向量均不相同,使用 QoS 本体方式将这些 QoS 属性向量和信誉度存储在 HBase 中,当用户请求某一服务时,在满足服务功能性需求的基础上通过 3.3 节中所述的 QoS 语义匹配方式选出一批候选服务,再根据 3.2 节所述的主客观赋权模式得到 QoS 对应的权值,并计算信誉度,提高 QoS 可信度,结合 QoS 数值匹配和式(6)得到最终的候选服务组,最后根据式(7)选出满足用户需求的服务。

实验中取 $\alpha = 0.3, \beta = 0.7$, 且设用户的 QoS 需求 $QoS_{user} = (0.9, 0.95, -0.6)$, 要求信誉度高于 0.6。根据上述步骤,本实验得到 5 个候选 Web 服务,表示为 $WS_i (i=1, 2, \dots, 5)$, 如表 1 所列, $QoS_i (i=1, 2, \dots, 5)$ 表示 5 个候选服务对应的 QoS 值信息; $(A, R, -RT)$ 分别表示可用性、可靠性和响应时间,并将响应时间除以 10 的相应指数倍数,转换为 $[0, 1]$ 区间再取反,其中 0 表示最好;使用 3.2 节的定义 3 得出信誉度 $Rep(S)$ 的实际值,使用 $[0, 1]$ 区间表示,值越大表示信誉度越高;表中主客观权值 ω^* 的 4 个向量分别表示了对可用性、可靠性、响应时间和信誉度的权重值;另外最终得到的 $QoSSim_i(Req, Adv)$ 值将作为 Web 服务 QoS 属性排名的参数值,值越

高表示该 QoS_i 对应的 Web 服务的非功能属性值匹配度越高,越可信。

表 1 QoS 向量对应的权重值和 QoS_{Sim_i} 值

QoS	信誉度 语义匹配	信誉度 Rep(S)	主客观 权重 w*	QoS 数值匹配	QoSSim _i (Req, Adv)
QoS ₁	0.8	0.6	(0.171, 0.3, 0.215, 0.314)	(0.95, 0.96, -0.6)	0.40788
QoS ₂	1	0.85	(0.206, 0.244, 0.271, 0.279)	(0.93, 0.95, -0.5)	0.52503
QoS ₃	0.9	0.7	(0.185, 0.265, 0.215, 0.335)	(0.95, 0.95, -0.6)	0.47970
QoS ₄	0.9	0.8	(0.185, 0.286, 0.250, 0.279)	(0.94, 0.95, -0.6)	0.46435
QoS ₅	1	0.9	(0.199, 0.244, 0.257, 0.3)	(0.92, 0.96, -0.6)	0.53312

图 4 是根据表 1 得到的 QoS_i 分别对应的 QoSSim_i(Req, Adv) 值,并绘制了图 5,从中可比较得到返回的 Web 服务顺序,由此可见,经过了 QoSSim_i(Req, Adv) 计算得到的返回顺序是按照从大到小排序,依次是 WS5、WS2、WS3、WS4、WS1,得到的该顺序既保证了 QoS 的可信性,也满足了用户需求偏好,经过 QoS 的语义相似度匹配其属性更能满足用户的需求,用户可根据该顺序择优选择服务。

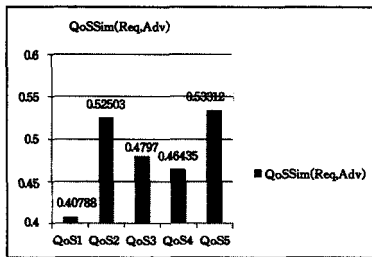


图 4 QoS_i 对应的 QoSSim_i(Req, Adv)

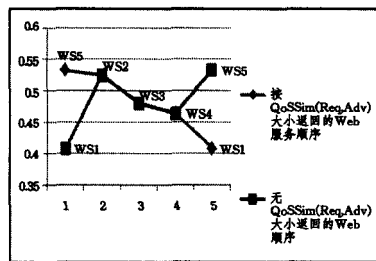


图 5 有无 QoSSim_i(Req, Adv) 时服务返回的顺序

图 6 和图 7 分别对比了有无 QoS 语义相似度匹配时的查找准确率和查找时间,从图中可以明显看出,加了 QoS 语义相似度匹配之后,随着 Web 服务的数量增加,其查找准确率也相应地增加,且比没有加入 QoS 语义相似度匹配的高,而查找时间并没有很大变化。

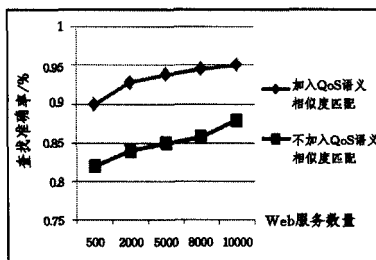


图 6 有无 QoS 语义相似度匹配时的查找准确率

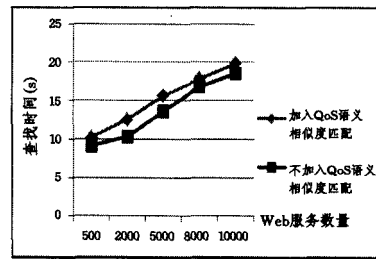


图 7 有无 QoS 语义相似度匹配时的查找时间对比

总之,在 Hadoop 平台中使用 QoS 本体,加入 QoS 语义匹配和数值匹配,最终根据式(7)选择 QoS_{Sim_i}(Req, Adv) 最大或较大的服务作为最佳的满足用户功能性需求和服务质量需求的 Web 服务,且该服务还通过 QoS 参数权重及信誉度,满足了客户的选择偏好,提高了 QoS 参数的可信度。并且 Hadoop 的分布式存储和计算的优点解决了传统 UDDI 集中式存储出现的瓶颈和单点失效的问题。实验证明本方案能有效提高 Web 服务在 Hadoop 平台中的查找准确性和 QoS 的可信性,使得优先返回的候选服务更能满足用户的 QoS 属性的语义准确性和需求偏好。

结束语 本文针对 Web 服务 QoS 查找准确性和可信性的问题,以及缺乏对服务质量的语义描述的问题,提出了一种基于 Hadoop 的 Web 服务语义可信 QoS 发现模型。其支持服务功能语义描述和 QoS 属性语义描述,并将 QoS 属性语义匹配分为 QoS 属性语义匹配和 QoS 数值匹配。首先,在满足用户功能属性语义需求的前提下进一步对 QoS 属性进行语义匹配,得到一组候选服务;其次使用主客观赋权模式为候选服务的 QoS 属性赋权值,提高 QoS 属性权重的客观性和准确性,再加入信誉度参数来提高 Web 服务 QoS 属性的可信性,同时采用 Hadoop 平台也可提高 Web 服务的发现速率,以满足客户对服务语义和服务质量的查找需求。实验证明本文提出的方法具有可行性和有效性。

参考文献

- [1] Funk C, Schultheis A, Linnhoff-Popien C, et al. Adaptation of composite services in Pervasive Computing environments[C]// Proceedings of the IEEE International Conference on Pervasive Services(ICPS 2007). 2007;242-249
- [2] 王广正,王喜凤,夏敏.基于本体的 Web 服务可靠性动态评估方法[J]. 计算机科学,2012,39(11):98-102
Wang Guang-zheng, Wang Xi-feng, Xia Min. The Web Service Dynamic Reliability Assessment Method based on Ontology[J]. Computer Science,2012,39(11):98-102
- [3] 石琳. Web 服务中的 WSDL 文档结构分析[J]. 软件,2012,33(10):142-144
Shi Lin. WSDL Document Structure Analysis in Web Service [J]. Software,2012,33(10):142-144
- [4] 李淑芝,刘锋,杨书新.基于云仿真的 Web 服务选择研究[J]. 计算机应用研究,2013,30(4):1069-1071
Li Shu-zhi, Liu Feng, Yang Shu-xin. Web Services Selection Research based on CloudSim[J]. Application Research of Computers,2013,30(4):1069-1071

(下转第 243 页)

- 1992;331-362
- [4] Marzena K. Comparative study of alternative types of knowledge reduction in inconsistent systems [J]. *International Journal of Intelligent Systems*, 2001, 16: 105-120
- [5] 秦克云, 赵华, 裴峥. 基于广义不可区分关系的决策表约简[J]. *西华大学学报(自然科学版)*, 2013, 32(4): 1-4
Qin Ke-yun, Zhao Hua, Pei Zheng. The reduction of decision table based on generalized indiscernibility relation [J]. *Journal of Xihua University(Natural Science Edition)*, 2013, 32(4): 1-4
- [6] Kryszkiewicz M. Rough set approach to incomplete information systems [J]. *Information Sciences*, 1998, 112: 39-49
- [7] Kryszkiewicz M. Rules in incomplete information Systems [J]. *Information Sciences*, 1999, 113: 271-292
- [8] 王国胤. Rough 集理论在不完备信息系统中的扩充[J]. *计算机研究与发展*, 2002, 39(10): 1238-1243
Wang Guo-yin. Extension of rough set under incomplete information systems [J]. *Journal of Computer Research and Development*, 2002, 39(10): 1238-1243
- [9] Bonikowski Z, Bryniarski E, Wybraniec U. Extensions and intentions in the rough set theory [J]. *Information Sciences*, 1998, 107: 149-167
- [10] Yao Y Y. Constructive and algebraic methods of theory of rough sets [J]. *Information Sciences*, 1998, 109: 21-47
- [11] Dubois D, Prade H. Rough fuzzy set and fuzzy rough sets [J]. *International Journal of General Systems*, 1990, 17: 191-209
- [12] Zhang X H, Zhou B, Li P. A general frame for intuitionistic fuzzy rough sets [J]. *Information Sciences*, 2012, 216: 34-49
- [13] Ziarko W. Variable precision rough set model [J]. *Journal of Computer and System Sciences*, 1993, 46(1): 39-59
- [14] 李德毅, 杜鹤. 不确定性人工智能[M]. 国防工业出版社, 2005
Li De-yi, Du Yi. *Artificial intelligence with uncertainty* [M]. National Defence Industry Press, 2005
- [15] Yao Y Y, Zhao Y. Attribute reduction in decision theoretic rough set models [J]. *Information Sciences*, 2008, 178: 3356-3373
- [16] Yao Y Y. Three way decisions with probabilistic rough sets [J]. *Information Sciences*, 2010, 180: 341-353
- [17] Zadeh L A. Fuzzy sets [J]. *Information and Control*, 1965, 8: 338-353
-
- (上接第 224 页)
- [5] 刘越. 云计算综述与移动云计算的应用研究[J]. *信息通信技术*, 2010, 4(2): 14-20
Liu Yue. *Reviews of Cloud Computing and Research on the Application of Mobile Cloud Computing* [J]. *Information and Communication Technology*, 2010, 4(2): 14-20
- [6] 侯建, 帅仁俊, 侯文. 基于云计算的海量数据存储模型[J]. *通信技术*, 2011, 44(5): 163-165
Hou Jian, Shuai Ren-jun, Hou Wen. *Massive Data Storage Model based on Cloud Computing* [J]. *Communications Technology*, 2011, 44(5): 163-165
- [7] Zhu X, Wang B. *Web Service Management Based on Hadoop* [C] // 2011 8th International Conference on Service Systems and Service Management (ICSSSM). Tianjin, 2011: 1-6
- [8] 任波. 基于语义的 Web 服务发现研究[D]. 杭州: 浙江工业大学, 2005
Ren Bo. *Based on the Semantic Web Service Discovery Research* [D]. Hangzhou: Zhejiang University of Technology, 2005
- [9] Mukhopadhyay D, Chathly F J, Jadhav N N. QoS Based Framework for Effective Web Services in Cloud Computing [J]. *Journal of Software Engineering and Applications*, 2012(5): 952-960
- [10] 祝希路. 基于 QoS 的可信 Web 服务关键技术研究[D]. 北京: 北京邮电大学, 2011
Zhu Xi-lu. *Research on Key Issues of Trustworthy Web Service Based on QoS* [D]. Beijing: Beijing University of Posts and Telecommunications, 2011
- [11] Srinivasan N, Paolucci M, Sycara K. *Semantic Web Service Discovery in the OWL-S IDE* [C] // Proceedings of the 39th Hawaii International Conference on System Sciences. 2006
- [12] 王尚广, 孙其博, 张光卫, 等. 基于云模型的不确定性 QoS 感知的 Skyline 服务选择[J]. *软件学报*, 2012, 23(6): 1397-1412
Wang Shang-guang, Sun Qi-bo, Zhang Guang-wei, et al. *Uncertain QoS-Aware Skyline Service Based on Cloud Model* [J]. *Journal of Software*, 2012, 23(6): 1397-1412
- [13] 陈洪磊, 刘东苏. 基于 QoS 本体的语义 Web 服务发现模型研究 [C] // 工程和商业管理国际会议. 2012: 3167-3171
Chen Hong-lei, Liu Dong-su. *The Semantic Web Service Discovery Model based on QoS Ontology Research* [C] // International Conference on Engineering and Business Management. 2012: 3167-3171
- [14] 方其庆, 刘庆华, 彭晓明, 等. QoS 全局最优的多目标 Web 服务选择算法[J]. *计算机应用研究*, 2009, 26(12): 4442-4448
Fang Qi-qing, Liu Qing-hua, Peng Xiao-ming, et al. *Global QoS Optimizing and Multi-Objective Web Service Selection Algorithm* [J]. *Application Research of Computers*, 2009, 26(12): 4442-4448
- [15] 康国胜, 刘建勋, 唐明董, 等. QoS 全局最优动态 Web 服务选择算法[J]. *小型微型计算机系统*, 2013, 34(1): 73-76
Kang Guo-sheng, Liu Jian-xun, Tang Ming-dong, et al. *Dynamic Web Services Selection Algorithm with Globally Optimal QoS* [J]. *Journal of Chinese Computer Systems*, 2013, 34(1): 73-76
- [16] 胡建强, 李涓子, 廖桂平. 一种基于多维服务质量的局部最优服务选择模型[J]. *计算机学报*, 2010, 33(3): 526-534
Hu Jian-qiang, Li Juan-zi, Liao Gui-ping. *A Multi-QoS Based Local Optimal Model of Service Selection* [J]. 2010, 33(3): 526-534
- [17] Ardagna D, Pernici B. *Global and Local QoS Guarantee in Web Service Selection* [C] // Proc. Business Process Management Workshop (BPM'05). 2005: 32-46
- [18] Altifai M, Risse T. *Combining global optimization with local selection for efficient QoS-aware service composition* [C] // Proceeding of 18th International Conference on World Wide Web (WWW 2009). 2009: 881-890
- [19] Chen H, Yu Tao, Lin K-J. *QCWS: an implementation of QoS-capable multimedia web services* [C] // IEEE Fifth International Symposium on Multimedia Software Engineering. 2003: 165-187