

QoS 本体驱动的 Web 服务发现

孔华云, 杨庚, 陈蕾, 王伟

(南京邮电大学计算机学院, 江苏南京 210003)

摘要:当前 Web 服务发现过程中, 为满足不同用户的个性化选择需求, 一种具有可靠服务质量保证的 Web 服务匹配方法是迫切需要的。基于 WordNet 本体, 文中提出一种考虑 QoS 语义和数值综合匹配的 QoS 匹配模型。该匹配模型包含了语义匹配、数值匹配、多属性决策矩阵、领域专家意见、以及获得 Top-k 候选服务等过程, 文中讨论了这些匹配过程, 并进行了仿真实验。结果表明, 与传统方法相比, 采用文中提出的 QoS 匹配模型在一定程度上提高了 Web 服务发现的查准率。

关键词:服务质量; WordNet; 多属性决策矩阵; Web 服务

中图分类号: TP31

文献标识码: A

文章编号: 1673-629X(2011)12-0201-04

Web Service Discovery Based on QoS Ontology

KONG Hua-yun, YANG Geng, CHEN Lei, WANG Wei

(College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: In the process of Web service discovery, in order to satisfy various kinds of users personalized need, Web service matching based on reliable QoS is necessary. Based on WordNet, it presents a QoS matching model considering compositing QoS ontology and QoS numerical value. The matching model contains ontology matching, numerical value matching, multi-attribute decision matrix, domain experts' suggestion, get Top-k advertising services. And a set of experiments demonstrate the benefits and effectiveness of the approach.

Key words: QoS; WordNet; multi-attribute decision matrix; Web service

0 引言

Web 服务作为一种新的技术, 成功地在 Internet 上实现了跨平台、语言独立、松散耦合的异构应用的交互和集成, 构成了面向服务的分布式计算模式。然而, 随着 Web 服务的日益增多, 查找符合要求的服务变得越来越困难。基于 QoS (Quality of Service) 服务质量保证的服务查询成为一种发展必然趋势。传统的 QoS 模型定义了响应时间、安全性、可靠性、可用性、吞吐率和价格参数。其结构简单明了, 但其表示能力有限。实际服务多种多样, 仅凭几个有限的 QoS 参数难以对 QoS 做全面描述。基于语义 QoS 的 Web 服务发现近年来获得了诸多学者的重视。

文献[1]提出一个 QoSOnt 本体模型, 分为通用层和领域相关层, 通用层定义 QoS 属性、QoS 度量和其

他基本 QoS 概念; 领域相关层则定义了特定服务领域、适应不同服务领域的特定 QoS 需求, 满足 QoS 模型稳定性和可扩展性的需求。

文献[2]设计一种 DAML-QoS 模型。DAML-QoS 本体将 QoS 本体分为三层: QoS 配置层、QoS 参数定义层和 QoS 度量方法。DAML-QoS 本体关注的是模型整体结构的划分, 着重阐述了 QoS 参数之间的关系。没有具体描述 QoS 参数的衡量标准。

文献[3]提出一种 WSMO-QoS 本体模型。WSMO-QoS 本体关注的是每个 QoS 参数, 具体描述 QoS 参数的每个属性, 认为每个度量值都是单值。没有关注 QoS 参数间的依赖关系。

文献[4]设计一个 Web 服务匹配框架, 包含一个 QoS 本体模型和基于 ARL 语言的自定义规则。该框架实现了更加灵活准确的服务匹配。通过引入描述逻辑推理引擎和定义相关的推理规则。

文中主要研究基于服务非功能属性 (QoS) 的服务发现^[5-8]。针对以上提出的经典模型的不足, 提出了一种基于 WordNet 本体, 用于简单便捷的描述 QoS, 具有多种 QoS 度量方式, 以及考虑领域专家意见的 QoS 服务发现模型 SN-QoS。

收稿日期: 2011-05-22; 修回日期: 2011-08-25

基金项目: 国家重点基础研究发展规划 (973) (2011CB302903); 国家自然科学基金资助项目 (60873231); 江苏省自然科学基金 (BK200942)

作者简介: 孔华云 (1986-), 女, 硕士研究生, 研究方向为计算机通信与网间互联技术; 杨庚, 教授, 博士生导师, 研究方向为计算机网络与安全、计算机通信与网间互联技术。

1 QoS

QoS (Quality of Service) 服务质量, 采用例如响应时间、可靠性、可用性、可访问性、吞吐量、安全性等众多质量属性, 用于度量 Web 服务的服务质量。文中将 QoS 匹配^[9-11]分为两部分, 本体匹配与数值匹配。对 QoS 本体和 QoS 数值作如下设计。

1.1 QoS 本体

在服务选择过程中, 作为本体概念的 QoS 参数之间存在着语义相似关系。然而由于 QoS 参数匹配是“对象”之间的匹配。单纯的语义匹配并不适应 QoS 匹配。如“响应时间”和“价格”作为 QoS 参数是没有比较意义的, 但是“响应时间”和“价格”作为本体概念, 二者之间可能存在语义距离, 相似度不为 0。因此文中通过设计 QoS 相容性 $C(q1, q2)$ 来描述该特性。当概念间的语义距离小于某个临界值 T 时, 二者相容, 即是语义可比的。反之, 如果大于该临界值 T , 二者不相容, 语义不可比。相容性见公式(1):

$$C(q1, q2) = \begin{cases} 1 & \text{dis}(q1, q2) \\ 0 & \text{dis}(q1, q2) \end{cases} \quad (1)$$

其中 $C(q1, q2)$ 表示 $q1, q2$ 两个 QoS 参数之间的相容性。 $\text{dis}(q1, q2)$ 表示 $q1, q2$ 两个 QoS 参数之间的语义距离。 T 代表某个临界值, 视具体情况而定。

因此, 综合 QoS 语义距离的特性, 计算两个 QoS 参数的语义相似度 $\text{DSim}(q1, q2)$ 取两者之间语义相似度与相容性的乘积, 即公式(2):

$$\text{DSim}(q1, q2) = \text{sim}(q1, q2) * C(q1, q2) \quad (2)$$

其中, $\text{sim}(q1, q2)$ 代表 $q1, q2$ 两 QoS 参数之间的语义相似度。

1.2 QoS 数值设计

传统的 QoS 模型都是基于 QoS 数值设计的, 说明 QoS 数值差异也是极其重要的。文中支持三种 QoS 数值度量方式: 单值、模糊单值和区间值。如响应时间是一个固定的值, 为单值。如安全性, 不能具体确定某一固定值, 描述为 {很高, 高, 一般, 低, 很低}, 对应于数值 {5, 4, 3, 2, 1}, 为模糊单值。如价格区间, 600 ~ 900, 为区间值。如果只单独针对三种度量方式互相匹配, 将有单值与单值、单值与模糊单值等 6 种匹配形式, 过于复杂, 难以处理。所以将所有度量方式统一为区间值, 统一度量, 为后文 QoS 相似度计算提供方便。很明显, 单值 q 即相当于区间 $[q, q]$ 。模糊单值转化为对应的单值后, 再转化为相应的区间值。到此, 所有的 QoS 数值都被标准化为统一的区间值。

2 综合 QoS 匹配过程

文中设计了一综合 QoS 语义与 QoS 数值的综合匹配模型 SN-QoS, 如图 1 所示。该匹配模型包含了语

义匹配、数值匹配、多属性决策矩阵、领域专家意见、以及获得 Top-k 候选服务五个部分。以下分别对五个部分作详细介绍。

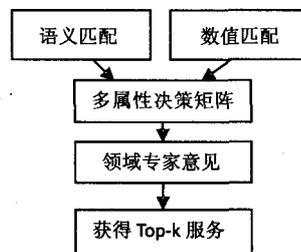


图 1 QoS 综合匹配模型 SN-QoS

2.1 QoS 本体匹配

进行 QoS 本体对象匹配, 归根到底是要研究文字语义间的匹配^[12,13], 因此文中引入一种比较公认的文字语义本体 WordNet, 可以直接将 WordNet 用做词典。每个单词都有一系列“词意”(Sense), 每个词意对应一个“同义词组”(Synset)。每个词之间有各种关系将其连接, 有上位关系(Hypernym)、下位关系(Hyponym)等。考虑本体分类树中两本体的语义距离主要考虑继承关系和二元关系, 并且当两个类在分类树上的深度越深, 它们之间的关系就越紧密, 因此它们间的距离就会越小。GSCM 距离算法即这样一个算法。引入最小共同祖先概念(LCA, Lowest Common Ancestor), 它是指两个本体概念节点共同祖先中深度最深的祖先节点。文中基于 GSCM 距离算法和 WordNet 本身“同义词组”特性提出 QoS 本体语义距离计算公式如下:

$$\text{Dis} = \frac{\text{size}(\text{synset1}) \times (\text{depth1} + 1) + \text{size}(\text{synset2}) \times (\text{depth2} + 1)}{(\text{size}(\text{synset1}) + \text{size}(\text{synset2})) \times (\text{depth}(\text{LCA}) + 1)} \quad (3)$$

对应的 QoS 语义相似度 Sim 为公式(4):

$$\text{Sim} = \frac{1}{\text{Dis}} = \frac{(\text{size}(\text{synset1}) + \text{size}(\text{synset2})) \times (\text{depth}(\text{LCA}) + 1)}{\text{size}(\text{synset1}) \times (\text{depth1} + 1) + \text{size}(\text{synset2}) \times (\text{depth2} + 1)} \quad (4)$$

对以上公式做如下说明:

1) $\text{size}(\text{synset})$ 代表 WordNet 中同义词组的大小。在传统 GSCM 距离算法中, 往往采用 $\log p$ 进行信息量的度量, 但是 p 概率需要通过文档集统计。文中讨论的算法不涉及文档集。因此, 假设所有单词出现的频率和其所处的 Synset 大小成正比, 这样即可直接采用 Synset 的大小作为权重。

2) 为避免分母出现 0, 采用平滑处理。将 $\text{depth}(\text{LCA})$ 、 depth1 、 depth2 加 1 进行平滑。

3) 显而易见, 保证 sim 值在 $[0, 1]$ 区间内。

根据 1.1 节 QoS 本体设计, QoS 综合语义相似度等于 QoS 对象相容性与 QoS 本体相似度的乘积, 即公式(2):

$$\text{DSim}(q1, q2) = \text{sim}(q1, q2) * C(q1, q2)$$

2.2 QoS 数值匹配

根据 1.2 节 QoS 数值设计,已经将所有服务请求和候选服务的 QoS 数值统一为数值区间 $[qr_{min}, qr_{max}]$ 。采用公式(5)对 QoS 数值区间进行匹配。假设存在服务 QoS 请求 $R[qr_{min}, qr_{max}]$, 候选服务 QoS 描述为 $A[qa_{min}, qa_{max}]$ 。QoS 数值相似度 SSim 为公式(5):

$$SSim = \begin{cases} 0 & \text{当 } qr_{max} < qa_{min} \text{ 或 } qr_{min} > qa_{max} \\ 1 & \text{当 } qr_{max} \leq qa_{max} \text{ 且 } qr_{min} \geq qa_{min} \\ \frac{qr_{max} - qa_{min}}{qa_{max} - qr_{min}}, & \text{当 } qr_{min} < qa_{min} < qr_{max} < qa_{max} \\ \frac{qa_{max} - qr_{min}}{qr_{max} - qa_{min}}, & \text{当 } qa_{min} < qr_{min} < qa_{max} < qr_{max} \end{cases} \quad (5)$$

对公式(5)作如下说明:

1)文中所计算 QoS 参数均假设候选服务中一项 QoS 属性都为一个区间值。基于实际情况中很难确定某个服务的 QoS 值为一固定值,取其为一个范围是合情合理的。

2)当请求服务 QoS 数值范围在候选服务 QoS 数值范围之外,认为其相似度为 0;当请求服务 QoS 数值范围完全在候选服务 QoS 数值范围内,认为其完全匹配,相似度为 1;当请求服务 QoS 数值范围部分在候选服务 QoS 数值范围内,取其共同的取值范围与请求服务和候选服务区间最大间距值的比值为其相似度。

3)明显的,保证 SSim 的取值区间为 $[0, 1]$ 。

2.3 QoS 多属性决策矩阵

2.1 与 2.2 节分别计算出每个请求服务 QoS 和每个候选服务 QoS 的语义和数值相似度^[14]。假设一个请求服务有 m 个候选服务,每个服务有 n 个 QoS 属性。构成 $m * n$ 多属性决策矩阵 $M_{m \times n}$ 。

$$M_{m \times n} = \begin{bmatrix} q_{11} & \dots & q_{1n} \\ \vdots & \dots & \vdots \\ q_{m1} & \dots & q_{mn} \end{bmatrix}$$

矩阵 $M_{m \times n}$ 说明如下:

1) 矩阵行代表候选服务,列代表候选服务的所有 QoS 属性。

2) 矩阵元素取值为 $DSim(q1, q2) * SSim(q1, q2)$ 。当 QoS 语义可比时,该 QoS 属性数值匹配才有意义。当 QoS 语义相似度 $DSim(q1, q2)$ 不变, QoS 数值相似度 $SSim(q1, q2)$ 越大,该 QoS 属性越匹配。当 QoS 数值相似度 $SSim(q1, q2)$ 不变,当 QoS 语义相似度 $DSim(q1, q2)$ 越大,该 QoS 属性越匹配。所以综合 QoS 相似度向量取两者乘积,即公式(6):

$$QSim(q1, q2) = DSim(q1, q2) \times SSim(q1, q2) \quad (6)$$

2.4 领域专家意见与获得合适 top-k 服务

领域专家意见是指领域专家根据经验给出的在保证服务可执行的基本条件下各个 QoS 参数门限值。该意见是为了避免服务请求者忽略重要的 QoS 参数而导致的服务无法执行的情况发生。对于矩阵 $M_{m \times n}$ 每个向量不在领域专家意见范围内的 QoS 相似度向量置为 0。

考虑通常情况下,多属性决策问题要获得合适的候选服务,利用经典门槛算法(Threshold 算法),获得多属性决策矩阵 $M_{m \times n}$ 的 Top-k 候选服务。步骤如下:

1. 对矩阵中每个 QoS 属性按照相似度从大到小快速排序。
2. 并行访问各个属性列,并记录各个访问到的表项,随机访问未访问到的表项,计算质量评分,维护一个临时结果集 R ,记录当前得分最高的前 K 个表项。
3. 给出门限值,取当前访问到的最下面位置表的各个位置上得分的总和。
4. 当存在 K 个表项的质量评分大于当前门槛值时就停止访问,算法收敛。
5. 获得临时结果集,即匹配合适的 Top-k 候选服务供用户选择使用。

3 实验结果

文中实验在 pc 机上完成,配置是 CPU Intel Pentium T3200,内存 2G,操作系统为 Windows XP。实验工具采用 MyEclipse7.0,Java jdk 1.6.0_14,WordNet 2.1。比较考虑语义与数值综合相似度得 QoS 匹配模型 SN-QoS 和仅仅考虑数值相似度的 QoS 匹配模型 N-QoS 的匹配度。通过查全率(召回率)和查准率来这两个重要指标来标识。查全率=(检索匹配成功的服务数量/候选服务中相关服务的总数量) * 100%。衡量文中匹配算法从候选服务中检索出相关服务成功度得一项指标。查准率=(检索匹配成功的服务数量/检索匹配出的总服务数量) * 100%。衡量文中匹配算法成功匹配度得一项指标。分别计算不同数量的 Web 服务发现,计算多次查询的平均查全率和查准率,并绘制成图以作比较。

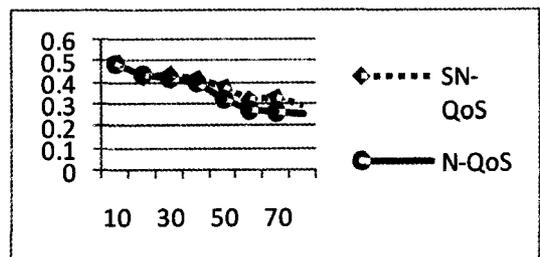


图 2 综合 SN-QoS 与数值 N-QoS 匹配的查准率比较
图 2 比较的是综合 SN-QoS 与数值 N-QoS 匹配的

查准率。横坐标轴代表候选服务数量,纵坐标代表查准率。虚线标识综合 SN-QoS 的查准率,实线标识只考虑数值 N-QoS 的查准率。从图 2 中可以得出结论,文中采取的综合 QoS 相似度匹配算法可以更准确的检索出匹配的服务。随着候选服务数量的增加,查准率有所下降。

图 3 比较的是综合 SN-QoS 与数值 N-QoS 匹配的查全率。横坐标轴代表候选服务数量,纵坐标代表查全率。虚线标识综合 SN-QoS 的查全率,实线标识只考虑数值 N-QoS 的查全率。从图 3 中可以得出结论,大致上,文中采取的综合 QoS 相似度匹配算法的查全率却低于只考虑数值的 QoS。这是因为仅仅考虑数值 QoS 的搜索范围更广泛,查全率比较高是合乎情理的。

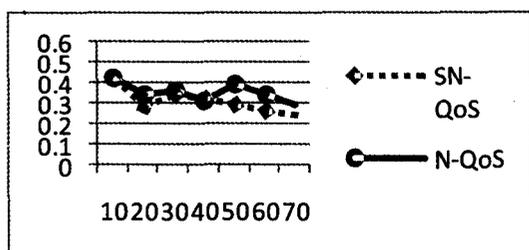


图 3 综合 SN-QoS 与数值 N-QoS 匹配的查全率比较

4 结束语

考虑语义服务质量的 Web 服务发现在信息量急剧膨胀社会中极具探讨价值。文中提出了一个考虑基于 WordNet 的 QoS 语义相似度和支持三种度量方式的 QoS 数值相似度相综合的 QoS 匹配算法。获得更高的查准率。并且考虑领域专家意见,避免用户请求忽略某些 QoS 参数。标准化为多属性决策矩阵,采用传统阈值算法得出符合用户需求的 Top-k 候选服务。构成一个比较完整的 Web 服务 QoS 匹配模型。QoS 参数在实际情况中有很多种形式,文中所采取的区间形式并不能覆盖所有情况,这需要进一步考虑。在实际情况中,用户的个人偏好也是值得提供商探讨考虑的。

参考文献:

[1] Dobson G, Lock R, Sommerville I. QoSOnt: an Ontology for QoS

in Service-Centric Systems[C]//31st EUROMICRO Conference on Software Engineering and Advanced Applications. Porto, Portugal: [s. n.], 2005.

[2] Zhou C, Chia L T, Lee B S. DAML-QoS Ontology for Web Service [C]//Proceedings of IEEE International Conference on Web Services(ICWS'04): [s. l.]: [s. n.], 2004.

[3] Wang X, Vitvar T, Kerrigan M, et al. A QoS-aware Selection Model for Semantic Web Service [C]//4th International Conference on Service Oriented Computing (ICSOC2006). [s. l.]: [s. n.], 2006: 390-401.

[4] Oldham N, Verma K, Sheth A, et al. Semantic Partner Selection [C]//15th International World Wide Web Conference. Edinburgh, Scotland, UK: [s. n.], 2006.

[5] Yang F C, Su S, Li Z. Hybrid QoS-aware semantic web service composition strategies [J]. Science in China Series F-information Sciences, 2008, 51(11): 1822-1840.

[6] Zeng L Z, Benattallah B, Ngu A H H. QoS-aware middleware for Web services composition [J]. IEEE Transactions on Software Engineering, 2004, 30(5): 311-327.

[7] Ran S. A Model for Web Services Discovery with QoS [J]. ACM Sigecom Exchanges, 2003, 4(1): 1-10.

[8] Al-Masri E, Mahmoud Q H. Discovering the Best Web Service [C]//Proc of the 16th International Conference on World Wide Web. Banff, Alberta, Canada: [s. n.], 2007: 1257-1258.

[9] 蒋哲远, 韩江洪, 王 钊. 动态的 QoS 感知 Web 服务选择和组合优化模型 [J]. 计算机学报, 2009, 32(5): 1014-1025.

[10] 郭得科, 任 彦. 一种 QoS 有保障的 Web 服务分布式发现模型 [J]. 软件学报, 2006, 17(11): 2324-2334.

[11] 张龙昌, 邹 华, 杨放春. 一种基于多 QoS 注册中心和模型异构的 WEB 服务选择算法 [J]. 电子与信息学报, 2011, 33(1): 168-174.

[12] 王芝虎, 葛 声, 张力军. 企业级 Java Web 服务的研究与实现 [J]. 计算机应用研究, 2005, 22(1): 128-133.

[13] 王东睿, 杨 庚, 陈 蕾, 等. 基于 WordNet 和 Kernel 方法的 Web 服务发现机制研究 [J]. 计算机技术与发展, 2010, 20(12): 69-76.

[14] 高 洋, 黄映辉. 基于三次匹配的语义 Web 服务发现模型 [J]. 计算机技术与发展, 2009, 19(9): 122-124.

(上接第 174 页)

[13] Law L, Menezes A, Qu M, et al. An Efficient Protocol for Authenticated Key Agreement [J]. Designs, Codes and Cryptography, 2003, 28: 119-134.

[14] Smart N P. An Identity-based Authenticated Key Agreement Protocol Based on the Weil Pairing [J]. Electronic Letters, 2002, 38: 630-632.

[15] Scott M. Authenticated ID-based Key Exchange and Remote Log-in with Insecure Token and PIN Number [R/OL]. 2002.

http://eprint.iacr.org/2002/164.

[16] Chen L, Kudla C. Identity Based Authenticated Key Agreement Protocols from Pairings [R/OL]. 2002. http://eprint.iacr.org/2002/184.

[17] McCullagh N, Barreto P S L M. A New Two-Party Identity-Based Authenticated Key Agreement [C]//Proceedings of CT-RSA 2005. [s. l.]: Springer-Verlag, 2005: 262-274.