

DOI:10.13196/j.cims.2013.10.SHENLimin.20131032

一种考虑 QoS 数据不确定性的服务选取方法

申利民^{1,2}, 陈 真¹, 李 峰³

- (1. 燕山大学 信息科学与工程学院, 河北 秦皇岛 066004;
2. 河北省计算机虚拟技术与系统集成重点实验室, 河北 秦皇岛 066004;
3. 东北大学秦皇岛分校 计算机与通信工程学院, 河北 秦皇岛 066004)

摘 要:针对 Web 服务随机性、模糊性和不一致性变化导致的 QoS 数据波动问题,提出一种考虑 QoS 数据不确定性的服务选取方法。该方法采用 QoS 属性云的期望、熵和超熵三个数字特征预测 Web 服务的 QoS 数据分布,分析 QoS 数据的不确定性,剔除 QoS 数据波动比较大的 Web 服务。在此基础上采用 QoS 属性反馈相似度作为计算 QoS 聚合值的权重,依据服务使用者的 QoS 评价及其数量计算服务的推荐度。通过一个原型系统和一组实验说明该方法能够有效减少 QoS 数据的不确定性对服务选取的影响,提高 Web 服务选取结果的准确性,使选取的服务更可靠。

关键词: Web 服务; 服务选取; 服务质量; 数据不确定性; 属性云

中图分类号: TP393 **文献标志码:** A

Service selection approach considering the uncertainty of QoS data

SHEN Li-min^{1,2}, CHEN Zhen¹, LI Feng³

- (1. College of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China;
2. The Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, Qinhuangdao 066004, China;
3. College of Computer and Communication Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China)

Abstract: To deal with the Quality of Services (QoS) data fluctuation caused by random, fuzzy and inconsistent changes of Web services, a service selection approach considering the uncertainty of QoS data was proposed. Three eigenvalues of expectation, entropy and hyper entropy were used to predict QoS data distribution of Web service and to analyze the uncertainty of QoS data. Thus the Web services with wider QoS data fluctuation were rejected. On this basis, QoS attributes feedback similarities were calculated as the weights for QoS aggregation, and the service recommendation degree was computed according to QoS evaluations and the numbers of service users. A prototype system and a set of experiments were implemented to show that this approach could effectively reduce the influence of QoS data uncertainty on the service selection, thus the accuracy of service selection result could be improved, and the selected service would be more reliable.

Key words: Web service; service selection; quality of service; data uncertainty; attribute cloud

0 引言

面向服务计算 (Service-Oriented Computing,

SOC) 及其主要应用——Web 服务已经成为 Internet 环境下构建分布式应用系统的主要理念^[1]。Web 服务因其可动态绑定、松耦合和跨平台等特

收稿日期: 2012-07-04; 修订日期: 2012-11-07。Received 04 July 2012; accepted 07 Nov. 2012.

基金项目: 国家自然科学基金资助项目 (61272125, 61300193); 河北省自然科学基金资助项目 (F2011203234); 河北省高等学校科学技术研究重点资助项目 (ZH2011115)。Foundation items: Project supported by the National Natural Science Foundation, China (No. 61272125, 61300193), the Natural Science Foundation of Hebei Province, China (No. F2011203234), and the Higher Education Science & Technology Research Foundation of Hebei Province, China (No. ZH2011115).

点,允许软件开发者从网络中选取满足相应功能需求的 Web 服务进行组合、协作,以完成既定的任务。随着部署在 Internet 上的 Web 服务不断丰富,具有相同或相似功能的 Web 服务越来越多,如何及时、准确地从众多候选服务中选择满足用户要求的 Web 服务,成为一个亟待解决的问题。

文献[2]将服务质量(Quality of Service, QoS)定义为一组服务非功能属性的集合,其中每个属性表征服务在某一侧面的质量信息,具有一个属性值,如响应时间、可靠性和信誉度等。由于 QoS 能很好地区分服务之间的差异,目前学者针对 QoS 提出了多种 Web 服务选取方法。分析现有的服务选取方法,一般可以分为基于 QoS 本体的语义 Web 服务选取方法和基于 QoS 属性计算的选取方法^[3-4]两类。前者通过 Web 服务本体语言(Web Ontology Language for Service, OWL-S)描述服务 QoS 属性和功能的本体,基于服务 QoS 语义的距离或者相似性进行服务匹配和选取;后者通过对 Web 服务的 QoS 建模,将每个服务的 QoS 数据表示为一个向量,然后量化并规范化用户对服务的 QoS 反馈数据,基于多属性综合评价理论将 QoS 数据加权求和,计算服务的推荐度,从而选取最好的服务。文献[5-6]从可信推荐者的角度,基于协同过滤思想收集用户关于服务的 QoS 数据,使每个用户在进行服务选取时拥有尽可能大量全面的信息来权衡并决策。然而更多时候,Web 服务选取被看作是一个优选问题,用户在选取服务前需要定义其感兴趣 QoS 属性的优先权和对服务的使用偏好等约束条件。文献[7-8]分别采用传输控制协议(Transmission Control Protocol, TCP)和用户数据包协议(User Datagram Protocol, UDP)网络图描述用户的服务使用偏好,这些网络图不但可以呈现不同 QoS 属性的重要性关系,而且建立了它们之间的依赖关系,然而这些网络图比较复杂,普通用户很难理解。文献[9-10]提出了一个上下文感知的服务选取方法,它采用 Web 本体描述语言(Web Ontology Language, OWL)/资源描述框架(Resource Description Framework, RDF)模型,将请求者的上下文转换为基于服务类别和用户领域相关的约束值,但是偏好权重是由用户根据一定语义主观设定的。实践表明,这些方法在很大程度上提高了 Web 服务选取的准确性和可靠性,但仍存在以下不足:

(1)没有考虑 Web 服务 QoS 数据的不确定性
Web 服务的提供者、实现语言和运行环境可

能发生变化,Internet 环境中与 Web 服务有关的任何变化,如位置、网络条件、时间及其他因素都将影响 Web 服务的质量,而这些变化都可以通过用户对服务 QoS 属性的反馈数据反映出来。同一个 Web 服务会有大量来自各方面的 QoS 数据,而这些数据具有不确定性,这将直接影响对 QoS 数据反馈的可靠性和服务性能计算的准确性,以及最终的 Web 服务选取结果。因此,QoS 数据的不确定性应该作为 Web 服务选取的重要条件。值得注意的是,一贯具有比较稳定 QoS 性能的 Web 服务通常比 QoS 数据波动较大的 Web 服务更可取。然而,已有服务选取方法在其 QoS 模型中大都没有考虑 QoS 数据的不确定性。

(2)不提供 QoS 数据的分布情况 在实际服务选取场景中,用户难以定义合理准确的 QoS 约束条件,而服务推荐结果准确与否,与 QoS 约束条件的设置息息相关。例如,用户在请求服务前定义 QoS 约束条件:平均服务响应时间(期望)为 $87\text{ ms} < rt < 93\text{ ms}$ 。若存在服务 w_{s1} 和 w_{s2} 的响应时间期望分别为 88 ms 和 89 ms 都满足用户要求,则传统的方法通常会把 w_{s1} 作为推荐服务,但进一步分析发现 w_{s1} 的响应时间波动范围(方差)为 2.5 ,而服务 w_{s2} 的方差为 0.5 ,即大部分情况下 w_{s2} 的响应时间小于 w_{s1} ,应该把 w_{s2} 作为推荐服务而非 w_{s1} 。出现这种情况的主要原因是用户不了解服务 QoS 数据的分布情况,定义了与实际有较大偏差的约束条件。因此,在服务选取前向用户提供 QoS 数据的分布情况是十分必要的,这可以使用户定义更符合实际 QoS 需求,提高服务选取的准确性。

针对一个服务会有多个来自多方面的 QoS 数据,本文将云模型引入到 QoS 数据具有不确定性的服务选取中,建立 QoS 属性云模型,基于其数字特征分析候选服务 QoS 数据的波动情况,并预测服务 QoS 数据的分布情况,为用户提供定义 QoS 约束条件的依据,应用 QoS 属性云模型实现候选服务的 QoS 反馈数据从定量到定性的转换,过滤掉不满足用户需求和 QoS 数据不稳定的 Web 服务,削弱 QoS 数据的不确定性对服务选取的影响,提高服务选取结果的准确性和可靠性。

1 QoS 数据不确定分析

现有 Web 服务的 QoS 数据一般通过可信第三方监控、服务使用者反馈和服务提供者声明三种方式获取。可信第三方基于监控代理和认证中心收集

注册表所有可用服务的各类信息,并基于此向用户提供服务的 QoS 信息;服务使用者在使用服务后提交它对服务 QoS 的反馈信息,Web 服务选取平台会记录每个用户的反馈数据,并将这些信息共享作为其他用户选取服务的参考;服务提供者在注册 Web 服务时,该服务的 QoS 信息会与其功能属性一起发布到统一描述、发现和集成(Universal Description, Discovery and Integration, UDDI)。

在开放的 Internet 环境中,Web 服务自身或外部环境的变化都将引起服务性能的不稳定,而这种不稳定主要以 QoS 数据不确定性的形式表现出来。文献[11]指出,人类在完成对客观世界的主观认知过程中,随机性、模糊性和不一致性是不确定性认知的重要特征,而 QoS 数据的不确定性也主要体现在这三个方面:① Web 服务的更新、服务器端硬件的变动以及网络负载的变化都将引起可信第三方监控到的 Web 服务 QoS 数据发生随机性的波动。② Web 服务 QoS 的一些属性具有主观性,如信誉度。本身是一个模糊的概念,没有明确的评价界限,而且不同使用者的评价尺度也有所不同。文献[12]将用户对服务的信誉评价划分为{恶意、非常不信任、不信任、稍微不信任、不确定、稍微信任、信任和非常信任}八个等级,然而相邻两个等级的转变在微观上是一个渐进的连续过程,它们之间不存在一个明确的、突变的界限,这也将导致 Web 服务的 QoS 数据发生模糊性的波动。③ Web 服务的 QoS 数据可能无法准确反映服务的实际性能。例如 Web 服务提供者为了短时间内吸引大批用户获取利益,可能会故意欺骗用户,不能保证总是提供他们许诺的服务质量。可见,QoS 数据的随机性、模糊性和不一致性是引起 QoS 数据不确定性的主要原因。

通过上述分析可知,QoS 数据随机性、模糊性和不一致性的变化是 QoS 数据不确定性的主要内涵。若忽略 QoS 数据的不确定性对服务选取的影响,则很难保证服务选取结果的准确性和可靠性。将云模型引入 Web 服务的选取中,能够有效地解决 QoS 数据的不确定性问题,下面给出云模型的定义。

2 QoS 属性云模型

2.1 云模型

云模型是在概率理论和模糊集合理论交叉渗透的基础上,用于统一刻画语言值随机性和模糊性的一种方法,实现语言值表述的定性概念与其定量表

示之间的不确定转换^[11,13]。其中定性信息表示的概念到定量数据的范围和分布情况采用正向云发生器(forward Cloud Generator, CG)实现,定量数据值到定性概念的转换采用逆向云发生器(backward Cloud Generator, CG^{-1})实现^[14]。文献[15-17]基于云模型解决了不同领域中定性概念和定量数据的转换问题。正态云模型是应用最广泛的一种云模型,因其具有良好的数学性质而可以处理自然科学和社会科学中大量不确定现象^[18],因此在本文的服务选取方法中采用正态云模型解决 QoS 数据的不确定性问题。

定义 1 云和云滴。设 U 是一个用数值表示的定量论域, C 是 U 上的定性概念,若定量值 $x \in U$ 是定性概念 C 的一次随机实现, x 对 C 的确定度 $\mu(x) \in [0,1]$ 是有稳定倾向的随机数, $\mu:U \rightarrow [0,1], \forall x \in U, x \rightarrow \mu(x)$,则 x 在论域 U 上的分布称为云,记为云 $C(x)$,每一个 x 称为云滴^[13]。

正态云模型的整体特征可以通过期望 Ex 、熵 En 和超熵 He 三个数字特征来表示。期望 Ex 是云滴在论域空间分布的期望,位于云模型的中心位置,是最能够代表定性概念的点;熵 En 代表定性概念的可度量粒度,是定性概念的不确定性度量,反映能够代表这个定性概念云滴的离散程度,由定性概念的随机性和模糊性共同决定;超熵 He 是熵的不确定性度量,即熵的熵,由熵的随机性和模糊性决定。用这三个特征数字表示的定性概念整体记作 $C(Ex, En, He)$,称为云模型的特征向量。

2.2 QoS 属性云模型定义

面对众多 QoS 数据不确定的 Web 服务,如何选择可用、可信和可靠的 Web 服务是用户面临的一个关键问题。本文根据云模型将随机性和模糊性相结合的知识表达特点描述 QoS 的定性概念,为用户选取可靠的 Web 服务提供更合理有效的支持。这里给出 QoS 属性云的相关定义。

定义 2 QoS 属性取值空间。QoS 属性取值空间 QAVS 是定量论域区间 $[0, v]$ 的一个有序数值集合,该集合可以由连续或离散的单调数值构成。其中有序集合中的每个数值表示用户对一个 QoS 属性的反馈值,0 和 v 分别为该 QoS 属性取值的下限和上限。

定义 3 QoS 属性等级空间。QoS 属性等级空间 QALS 是定性表示一个 QoS 属性的有序概念集合,可以使用定性概念表达用户对该 QoS 属性的期望程度。可以为 QALS 预设 1 个或多个期望度等

级,也可以忽略期望度等级的设定。

定义 4 QoS 属性云。设 QoS 属性取值空间 QAVS 是云上的定量论域 $U, e(e \in QALS)$ 是 QoS 属性等级空间上的 QoS 属性定性概念, $x(x \in QAVS)$ 是定性概念 e 的一次定量 QoS 属性数据反馈, x 对 e 的确定度 $\mu(x) \in [0, 1]$ 是有稳定倾向的随机数: $\mu: QAVS \rightarrow [0, 1], \forall x \in QAVS, x \rightarrow \mu(x), x$ 在论域 U 上的分布称为 QoS 属性云, 记为 $QAC(x)$, 每一个 x 称为一个云滴。

例如,应用 QoS 属性云模型表示响应时间属性云。设响应时间云模型的属性取值空间 $U = [0, 1]$, 其对应的属性等级空间定义为 $QALS = \{\text{快}, \text{中}, \text{慢}\}$ 。用响应时间云模型表示定性概念“中”时, 根据用户体验普遍认为, 将响应时间 0.5 s 定为完全属于“中”的概念即期望 Ex 。图 1 所示为利用正向云发生器^[13]生成表示定性概念“中”的定量数据的范围和分布情况。每一个云滴是定性概念映射到属性取值空间的一个点, 即实现了一次量化, 这种实现带有不确定性, 而 QoS 属性云模型同时给出了这个点能够代表定性概念的确定度。云滴出现的概率越大, 云滴的确定度也就越大。图 1 中云滴的数据范围主要落在概念“中”的属性等级区间内, 且期望值 $Ex = 0.5$, 因此该服务的响应时间分布对应的响应时间属性等级为“中”。

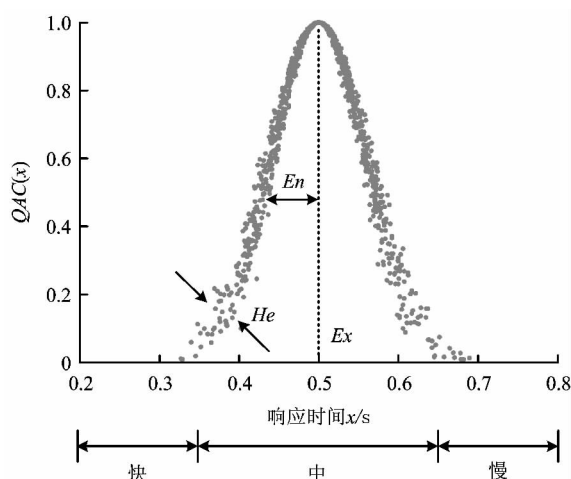


图1 响应时间云模型

QoS 属性云是应用云模型表示的 QoS 属性定性概念。QoS 属性云定量取值可以是 $[0, n]$ 区间任意数值构成的有序数值, QAVS 的有序数值可以由表示 QoS 取值的一组连续或离散数值构成, 适用于不同 QoS 属性数据反馈评价机制, 能够有效体现 QoS 属性云的定性、定量转换及定性概念提升的特

点。当一个 QoS 属性取值为连续数值空间时, QoS 属性云表示的是定性 QoS 属性概念与定量论域间的定性、定量转换关系; 当 QoS 属性取值为离散数值空间时, 可以认为 QoS 属性云表达的定性概念是对论域空间内细粒度概念的一种提升, 即定性概念与论域空间的数值形成概念层次。QoS 属性云不强制要求对 QoS 属性等级空间内的定性概念进行排序, 它表达的是定性 QoS 属性概念, 仅使用云的数字特征对概念进行表述。但可以通过 QoS 属性云数字特征的比较, 支持 QoS 数据的不确定分析和决策, 即通过比较 $QAC(Ex, En, He)$ 对服务进行不确定评价, $QAC(Ex, En, He)$ 称为 QoS 属性云特征向量。在使用 QoS 属性云模型进行定性知识表示时, 需要根据不同的 QoS 属性为其数字特征赋予合理而有意义的物理含义。

QoS 属性云的期望 Ex 位于云的重心, 是用户对服务属性度评价的典型值, 即服务 QoS 属性反馈数据的平均水平; 熵 En 是云滴距离 Ex 的距离, 表示用户反馈数据与平均水平的偏离程度, 即用户可接受的服务 QoS 属性数据的波动范围; He 表现为云的厚度, 表示云滴确定度的随机性大小, 即用户对 QoS 属性评价标度的不确定程度。 En 和 He 体现了 QoS 属性数据的稳定程度, 因为 He 是 En 的熵, 所以在服务选取不确定决策过程中, He 的优先级高于 En 。当 Ex 较大时, 对于正 QoS 属性(如可靠性、信誉度), 认为服务对用户的可满足性较大; 对于负属性(如响应时间、费用), 认为服务对用户的可满足性较小, 反之亦然。当 En 和 He 较小时, 认为服务对用户满足性比较稳定, 反之稳定性较差。

3 考虑 QoS 数据不确定性的服务选取方法

图 2 所示为考虑 QoS 数据不确定性的服务选取处理流程。用户得到一组功能类似的候选 Web 服务后, 首先, 根据服务 QoS 数据的分布情况自定义合理准确的 QoS 约束条件; 其次, 应用 QoS 属性云对 Web 服务的 QoS 数据进行不确定决策, 过滤掉 QoS 数据表现波动比较大的 Web 服务, 在保证服务可靠性的基础上缩小 Web 服务选取的搜索空间, 提高服务选取效率; 再次, 对过滤后服务的 QoS 数据进行规范化处理, 计算用户对服务评价的 QoS 聚合值; 最后, 依据服务使用者的 QoS 评价及其数量计算服务的推荐度, 并依此对服务排序完成服务推荐。

3.1 基于 QoS 数据分布的约束条件定义

约束条件是区分服务 QoS 差异性的准则, 本文

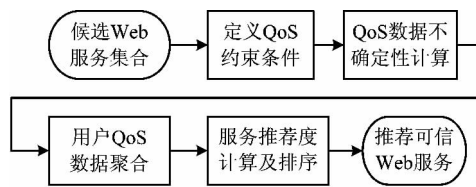


图2 考虑QoS数据不确定性的服务选取流程

依据 QoS 属性云模型实现定量 QoS 数据到定性概念的转换,使用 QoS 属性云刻画不确定决策的约束条件,辨别多个候选服务 QoS 数据平均水平和变化方式的差异。当服务请求者面对一组功能相同的候选服务集合 $WSC = \{WS_1, WS_2, \dots, WS_m\}$ 时,在不了解 QoS 数据分布的情况下,难以给出合理准确的 QoS 约束条件。因此,为用户提供服务 QoS 数据的分布情况是必要的。服务的 QoS 数据分布不能只是简单的用户反馈 QoS 数据的罗列,否则用户很难直观理解服务的分布情况,而且这些未经处理的 QoS 数据不能反映 QoS 数据的波动情况,不利于用户 QoS 约束条件的定义。为此,依据所有候选服务的 QoS 反馈数据,计算候选服务关于 QoS 属性 i 的数字特征 $QAC^i(Ex, En, He)$,并将其作为正向云发生器的输入,预测 QoS 属性 i 的数据分布。其中,若 QoS 有 l 个属性,则相应地会生成 l 朵 QoS 属性云。

用户可以根据生成的 QoS 属性正态云直观理解服务的 QoS 数据分布情况,并根据 QoS 属性云三个数字特征的物理含义定义合理准确的约束条件。约束条件包括用户关注的服务 QoS 属性集合 $SAC = \{A_1, A_2, \dots, A_r\}$ 和相应 QoS 属性的约束阈值集合 $threshold\{(ex^i, en^i, he^i)\}$,其中 $0 < i \leq r$ 。

3.2 QoS 数据不确定决策

得到 QoS 约束条件后,即可进行 QoS 数据不确定决策(QoS Data Uncertainty Decision-making, QDUD)。在候选服务 $WSC = \{WS_1, WS_2, \dots, WS_m\}$ 中,根据用户定义的 QoS 约束条件和过滤规则,将 QoS 数据不满足用户期望的和波动比较大的服务过滤掉,生成规模较小的候选服务集合 $WSC' = \{WS'_1, WS'_2, \dots, WS'_t\}$, $|WSC'| \leq |WSC|$,完成择优选取的问题。式(1)给出了 QDUD 决策过程的形式化描述:

$$QDUD(WSC) \xrightarrow{\text{constraints}} WSC'. \quad (1)$$

为了减少由于 QoS 数据的不确定性对服务选取准确性、可靠性和计算效率的影响,采用无确定度的逆向云生成算法,将定量的 QoS 属性数据转换为

定性的 QoS 属性概念,计算服务 QoS 属性云的特征向量;再依据 QoS 约束条件过滤掉不满足条件的服务,得到 QoS 数据表现比较稳定可靠的候选 Web 服务集合。在给出具体的算法前定义服务的用户反馈。

定义 5 服务用户反馈。 WS_i 是一个 Web 服务,其用户反馈是一个二元组 $UF_i = (U_i, QF_i)$,式中: U_i 是提供服务 WS_i 的 QoS 数据的有限用户集合, $U_i = \{u_{i1}, u_{i2}, \dots, u_{im}\}$,包括服务提供者、可信第三方和服务使用者三类用户; QF_i 是 U_i 对服务 WS_i 的 QoS 反馈数据集合, $QF_i = \{qf_{i1}, qf_{i2}, \dots, qf_{im}\}$,其中服务 WS_i 的用户 u_{ij} 反馈的 QoS 数据为向量 $qf_{ij} = (f_{ij}^1, f_{ij}^2, \dots, f_{ij}^l)$,表示服务用户 u_{ij} 从 l 个 QoS 属性角度对服务 WS_i 进行反馈评价。

考虑到不同用户在选取服务时具有个性化的 QoS 属性期望,用户在选取服务时需要自定义其关注的 QoS 属性,其中 QoS 约束条件可以设置一个 QoS 属性或多个,从而提高算法的灵活性和扩展性。当服务请求者在获得一组功能相似的候选 Web 服务 WSC 后,设置其关注的服务 QoS 属性集合 $SAC = \{A_1, A_2, \dots, A_r\}$ 及相应的阈值 $threshold\{(ex^i, en^i, he^i)\}$ 等约束条件,完成 QoS 数据的不确定决策。具体算法如下:

算法 1 过滤 QoS 数据不确定 Web 服务的决策过程。

输入: 一组功能相似的候选 Web 服务集合 $WSC = \{WS_1, WS_2, \dots, WS_m\}$; 用户关注的 r 个 QoS 属性集合 $SAC = \{A_1, A_2, \dots, A_r\}$,其中 $r \leq l$; 约束阈值集合 $threshold\{(ex^i, en^i, he^i)\}$,其中 $0 < i \leq r$ 。

输出: 过滤后的候选 Web 服务集合 $WSC' = \{WS'_1, WS'_2, \dots, WS'_t\}$, $|WSC'| \leq |WSC|$ 。

步骤 1 获取候选服务 WS_i 的用户反馈 UF_i 。

步骤 2 根据服务请求者关注的 QoS 属性 A_k ,从 UF_i 获得使用者关于属性 A_k 的反馈数据 $F_i^k = \{f_{i1}^k, f_{i2}^k, \dots, f_{im}^k\}$ 。

步骤 3 根据 F_i^k 计算属性 A_k 的样本均值 $\bar{F}_i^k = \frac{1}{n} \sum_{j=1}^n f_{ij}^k$ 和样本方差 $S_i^{k^2} = \frac{1}{n-1} \sum_{j=1}^n (f_{ij}^k - \bar{F}_i^k)^2$ 。

步骤 4 计算服务 WS_i 关于属性 A_k 的 QoS 反馈数据的云特征值:期望 $Ex_i^k = \bar{F}_i^k$,熵 $En_i^k = \frac{1}{n} \sqrt{\frac{\pi}{2} \sum_{j=1}^n |f_{ij}^k - Ex_i^k|}$,超熵 $He_i^k = \sqrt{(S_i^k)^2 - (En_i^k)^2}$ 。

步骤 5 服务过滤规则: ex^k 判断服务属性 k 是否满足用户的期望, en^k 和 he^k 判断服务的不确定程度, 其中 en^k 判断服务 QoS 数据的波动情况, he^k 判断用户对服务 QoS 属性 k 评价标度的不确定情况。若 A_k 为负属性判断条件 $k < r$, 则 $Ex_i^k \leq ex^k, En_i^k \leq en^k, He_i^k \leq he^k$, 若条件满足则返回步骤 2, 判断下一属性 A_{k+1} ; 若属性 A_k 为正属性, 判断条件 $k < r$, 则 $Ex_i^k \geq ex^k, En_i^k \leq en^k, He_i^k \leq he^k$, 若条件满足则返回步骤 2, 判断下一属性 A_{k+1} 。

步骤 6 若 $k = r$, 即所有属性都满足条件, 则将服务 WS_i 作为新的候选服务添加到集合 WSC' 中。

步骤 7 若存在尚未判定的候选服务, 则重置 QoS 判定属性 $k=0$, 返回步骤 1, 开始判断下一个服务 WS_{i+1} 。

步骤 8 返回过滤后的 Web 服务集合 WSC' 。

算法 1 严格按照用户定义的 QoS 约束条件, 通过 Ex 滤除不满足用户期望 QoS 属性值的服务来保证服务的性能, 通过 En 滤除 QoS 数据波动比较大的服务, He 滤除用户对 QoS 属性评价标度随机性较大的服务, 来保证稳定性, 从而解决 Web 服务 QoS 数据的不确定性问题。然而, 当 WSC' 集合规模较大时, 表示仍然存在较多的候选服务, 还需要对 WSC' 做进一步处理。

3.3 QoS 属性值聚合

QoS 属性值聚合是通过 QoS 反馈数据获得用户对服务的综合评价。在聚合用户 QoS 反馈数据之前, 需要对 QoS 数据进行归一化处理。本文采用文献[19]提出的方法, 区别 QoS 属性的正负, 正属性表示该属性值越大, 其服务质量越高, 如吞吐量、可靠性等; 负属性表示该属性值越小, 其服务质量越高, 如响应时间、费用等。因此, 正属性使用式(2), 负属性使用式(3)来进行具体变换, 其中 f_{\max} 和 f_{\min} 分别为该属性反馈数据中的最大值和最小值。

$$f = \begin{cases} \frac{f - f_{\min}}{f_{\max} - f_{\min}}, & f_{\max} - f_{\min} \neq 0; \\ 1, & f_{\max} - f_{\min} = 0; \end{cases} \quad (2)$$

$$f = \begin{cases} \frac{f_{\max} - f}{f_{\max} - f_{\min}}, & f_{\max} - f_{\min} \neq 0; \\ 1, & f_{\max} - f_{\min} = 0. \end{cases} \quad (3)$$

传统的 QoS 数据聚合方法对量化后的 QoS 数据进行简单的加权求和, 其中 QoS 属性的权重大都来自用户主观设置, 这种方法缺乏客观性。因此, 借鉴文本分类领域中向量距离分类的思想, 将用户共

同使用过的 Web 服务集合视为一个多维空间, 用户对这些服务反馈的 QoS 数据视为该空间中的一个点, 则用户对 QoS 属性的反馈相似程度可以通过他们之间的几何距离来度量, 距离越近的两个用户的 QoS 属性反馈相似度越大。

定义 6 QoS 属性反馈相似度。设服务请求者 u_r 和服务使用者 u_u 共同使用的服务集合为 $WSC^u = \{WS_1^u, WS_2^u, \dots, WS_m^u\}$, u_r 和 u_u 对属性 A_i 的评价向量空间为 $\{(f_{1r}^i, f_{1u}^i), (f_{2r}^i, f_{2u}^i), \dots, (f_{mr}^i, f_{mu}^i)\}$, 则 u_r 和 u_u 关于 QoS 属性 A_i 的评价相似度 fs^i 定义为

$$fs^i = \begin{cases} 1 - \frac{\sum_{j=1}^{|WSC^u|} |f_{jr}^i - f_{ju}^i|}{|WSC^u|}, & |WSC^u| \neq 0; \\ 0, & |WSC^u| = 0. \end{cases} \quad (4)$$

服务使用者 u_j 对服务 WS_i 的 QoS 数据聚合值

$$r_{ij} = \sum_{k=1}^l \frac{fs_j^k}{sum_{fs}} \times f_{ij}, sum_{fs} = \sum_{k=1}^l fs_j^k. \quad (5)$$

式(5)将服务使用者和请求者之间的 QoS 属性反馈的相似程度作为计算 QoS 聚合值的权重, 考虑了用户 QoS 数据反馈的相似程度, 减小了因为用户评价尺度不同对服务的 QoS 聚合值计算产生的影响, 使计算出的 QoS 聚合值更加接近服务请求者的实际评价, 保证了 QoS 聚合值计算的合理性和客观性。

3.4 服务排序推荐

结合上两节的计算, 得到 QoS 数据稳定可靠的服务集合 WSC' 和每一个服务使用者对服务的反馈 QoS 数据的聚合值。由于每个服务的使用者数量不同, 为了节省数据存储空间, 将用户对服务的 QoS 反馈数据以链表形式的结构存储, 由 WSC' , $User$ 和 QoS 数据聚合值得到的服务使用评价链表如图 3 所示, 用户评价 r_{1a}, r_{2b}, r_{mc} 中的 a, b, c 分别表示对应 Web 服务的使用者数量。

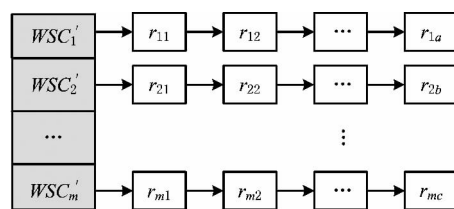


图3 服务使用评价链表

为了对服务进行排序选取, 本文综合考虑用户

的 QoS 聚合值和服务使用者数量对推荐度的影响,以式(6)计算的服务推荐度作为排序的标准。

$$\overline{R(WS_i)} = \sum_{j=1}^{|U_i|} \frac{r_{ij}}{|U_i|},$$

$$R(WS_i) = \overline{R(WS_i)} + (1 - \overline{R(WS_i)}) \times \beta(|U_i|). \quad (6)$$

式中: $|U_i|$ 表示服务 WS_i 使用者的数量; $\overline{R(WS_i)} \in [0, 1]$ 表示用户推荐度均值, $1 - \overline{R(WS_i)}$ 表示服务推荐度增长空间; $\beta(|U_i|) \in [0, 1]$ 表示推荐度增长系数,反映了服务使用者数量对推荐度的影响。服务使用者数量越多说明该服务越值得推荐,因此函数 $\beta(x)$ 应具有如下性质:

(1) 当 $1 \leq x_1 < x_2$ 时, $\beta(x_1) < \beta(x_2)$ 。

(2) $\beta(1) = 0$, 且 $\lim_{x \rightarrow \infty} \beta(x) = 0$ 。

依据上述 2 个性质,构造函数 $\beta(x)$ 如下

$$\beta(x) = 1 - \frac{1}{\delta \sqrt[e^{x-1}]{x}}, x \geq 1. \quad (7)$$

置信数 $\delta \geq 2$ 用于控制推荐度增长系数 $\beta(x)$ 趋于 1 的速度, δ 的取值越大, $\beta(x)$ 趋于 1 的速度就越慢,其中 δ 的取值可以根据系统中用户数量的多少动态调整。由性质(1)和(2)可知, $\beta(x)$ 是单调递增函数,当 $x = 1$ 时,其值最小为 0,表示服务只有一个推荐用户时推荐度的增长空间为 0;当 x 趋于无穷大时,其值最大趋于 1,表示推荐度的增长空间为 $1 - \overline{R(WS_i)}$ 。当有多个推荐者的数量达到一定数量时, $\beta(|U_i|)$ 都趋于 1,导致服务的推荐度也趋于 1,说明这些服务都能很好地满足用户的需求,推荐的服务都是可信且可靠的。式(6)综合考虑了使用者对服务的评价和服务使用者数量对推荐度的影响,服务的 QoS 反馈的 QoS 数据越多,通过式(6)所计算的服务推荐度越准确和可信。

4 原型系统及实验结果分析

4.1 实现原型

考虑到 QoS 数据的不确定性会影响服务选取结果的准确性和可靠性,利用本文提出的服务选取方法,采用 ASP.NET MVC3 和 SQL Server 数据库技术实现了一个 Web 服务选取原型系统。原型系统的主要功能包括:

(1) 服务提供者可以在平台注册、更新和删除服

务,提供服务的功能描述和 QoS 数据等信息。

(2) 服务推荐原型可以动态定义服务的 QoS 模型,提供服务 QoS 属性性能测量工具,收集、更新用户的 QoS 反馈数据,依据服务选取方法产生服务推荐。

(3) 服务请求者可以使用该原型系统查询服务自定义 QoS 需求,根据系统推荐调用服务并反馈 QoS 数据。

4.2 实验环境设置

原型系统的实验环境参数为: Pentium(R) Dual-Core 2.93 GHz CPU 和 2 GB 内存,使用 Windows XP 操作系统和 Visual Studio 2010 开发环境。实验方法基于用户 QoS 反馈数据,需要建立用户评分数据集。为了验证基于 QoS 数据不确定性服务选取方法的有效性和可靠性,原型系统采用 WS-DREAM 提供的数据集^[20-21],包括来自全球 24 个国家的 339 个用户对 5 825 个 Web 服务的 QoS 评估反馈,其中包括响应时间和吞吐量两个 QoS 属性真实数据。此外,还扩展了此数据集,添加了服务费用、可靠性、可用性、成功率和信誉度等五个 QoS 属性,其参数说明如表 1 所示。

表 1 Web 服务 QoS 属性说明

QoS 属性	描述	单位	性质	取值
响应时间	请求和执行服务所用的时间	s	负	$[0, +\infty]$
吞吐量	在给定的时间内服务可以被访问的数量	#/sec	正	$[0, +\infty]$
费用	调用一次服务的价格	#/call	负	$[0, +\infty]$
可靠性	错误信息占总信息的比率	%	正	$[0, 1]$
可用性	服务能够被访问或调用的程度	%	正	$[0, 1]$
成功率	调用服务并成功接收到服务响应的概率	%	正	$[0, 1]$
信誉度	用户对服务的信任程度,取决于用户对服务的使用情况	#	正	$[0, 1]$

随机选取 WS-DREAM 中的 1 000 个 Web 服务,参考 WebserviceList 站点 (<http://webservice-list.com/>) 上服务的分类情况,按照文献[22]的分类方法将这 1 000 个服务分为 10 类,分类情况如表 2 所示。用户根据个人需求发起一类服务选取请求。

表 2 Web 服务分类情况

类别	描述	实例	数量
访问控制/安全	对访问者的授权和安全控制	http://xmltools.oio.dk/AuthorityCode/AuthorityCodeConversion.asmx?wsdl	82
地址/定位	根据地址、IP 或坐标实现地理位置定位	http://www.webxml.com.cn/WebServices/IpAddressSearchWebService.asmx?wsdl	113
商业金融	商业数据及交易信息查询转换等服务	http://forums.genom-e.com/_vti_bin/BusinessDataCatalog.asmx?wsdl	182
Web 开发工具	可动态嵌入到系统中的 Web 开发工具	http://soft.72ec.com/WebService/ECToolsService.asmx?wsdl	102
目录数据库	开放数据库的浏览查询	http://projects.dnv.com/reference_data/RD4Browser/RD4WebServices.asmx	127
在线验证	验证字符及图片动态在线验证	http://www.webxml.com.cn/WebServices/ValidateCodeWebService.asmx?wsdl	84
证券	股票债券基金查询分析等服务	http://www.arikan.at/axis/services/BankService?wsdl	115
搜索/探测	与某一领域相关的服务搜索和发现	http://www.webservicex.com/UDIBusinessFinder.asmx?wsdl	50
自动销售	完成服务或商品在线销售的 Web 服务	http://theplasticbags.com/Services/BagsOnNet.asmx?wsdl	79
通讯和通信	提供电话通讯记录和主流通讯软件的检测	http://webservice.webxml.com.cn/webservices/qgOnlineWebService.asmx?wsdl	66

4.3 实验结果及分析

4.3.1 实验 1——服务选取实例分析

用户在原型系统中查询一组“搜索/探测类服务”，在获取到该类型的候选 Web 服务后，为了更好地区分这些服务的性能，用户需要定义服务选取的 QoS 约束条件，如图 4 所示。由于服务的 QoS 数据具有不确定性，用户很难知道当前服务 QoS 数据的分布情况，难以输入符合实际的 QoS 约束条件。因此，在用户定义 QoS 约束条件之前需要了解服务的 QoS 数据分布情况。

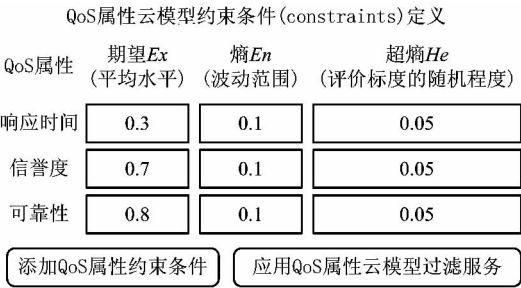


图4 QoS属性约束条件定义

图 5 所示为在用户定义 QoS 约束条件前根据其关注的 QoS 属性给出的服务响应时间、可靠性和信誉度三个 QoS 属性的分布情况，它根据候选服务三个 QoS 属性的反馈数据计算服务的相应属性的数字特征，将候选服务的使用者数 242 和数字特征作为正向云生成器的输入生成 242 个云滴，预测服务响应时间、可靠性和信誉度的分布情况。用户可以依据此分布定义合理的 QoS 约束条件，提高服务选取的准确性，否则可能因为输入不合理的 QoS 约束条件导致服务选取失败或者不准确。

在定义完 QoS 属性的约束条件后，根据算法 1

的 QoS 数据不确定决策方法， Ex 将不满足用户对服务 QoS 属性期望的服务过滤掉，同时 En 和 He 将波动较大的服务剔除。实验中候选服务 WSC 的数量为 50 个，过滤后新的候选服务 WSC' 的数量为 11，即有 39 个服务被过滤掉。表 3 列出了响应时间、信誉度和可靠性不满足约束条件的典型服务，其中服务被过滤掉的原因在表中加粗表示。从表 3 可以看出被过滤掉的服务大部分是因为不满足 En 和 He 这两个约束，说明 QoS 数据的不确定性是影响服务选取的主要原因，但也不能忽略 Ex 对服务选取的作用，例如 $SID=230$ 的服务就是因为 Ex 不满足条件而被过滤掉。

对 WSC' 应用式(5)和式(6)计算服务的推荐度。表 4 依据此方法给出了最终的 Web 服务推荐结果，其中置信数 $\delta=20$ 。在此 Web 服务推荐列表中，将选择推荐度最高的服务 $SID=232$ 推荐给服务请求者。

4.3.2 实验 2——验证 QoS 数据不确定决策的有效性

Web 服务选取的主要目的是给用户推荐准确可靠的 Web 服务，提高用户体验。为了验证 QoS 数据不确定决策方法的有效性，采用用户满意度作为评价指标，用户完成一次 Web 服务选取，根据本文提出的选取方法计算服务的推荐度 R_i ，服务的实际 QoS 性能记为 Q_i ，实验重复 n 次，则用户对服务的满意度按式(8)计算。其中服务推荐度与服务的实际性能差距越小，用户的满意度越高。

$$Sat(u) = 1 - \frac{\sum_{i=1}^n |R_i - Q_i|}{n} \quad (8)$$

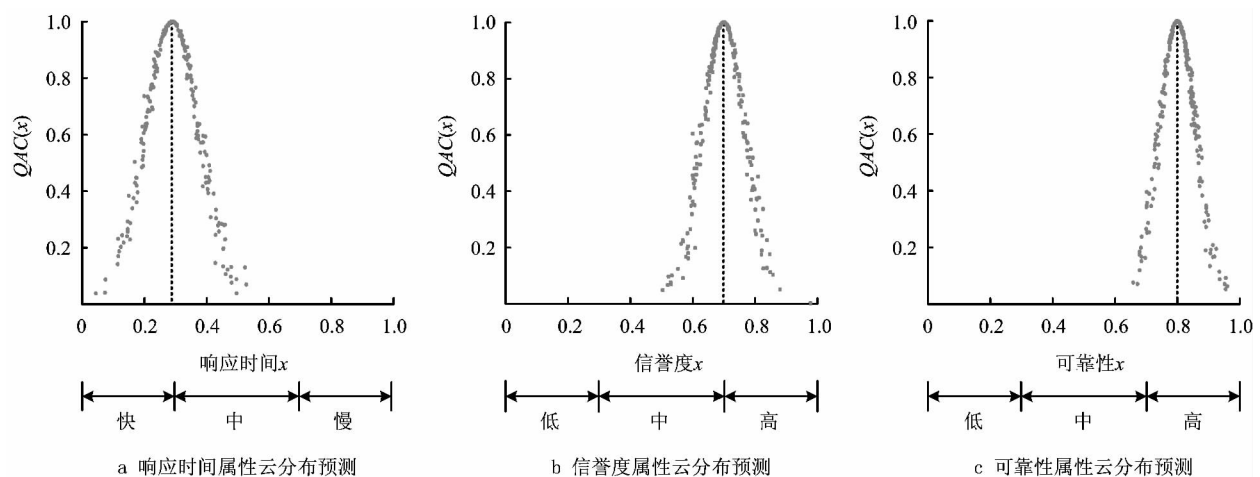


图5 QoS数据分布预测

表 3 不确定决策剔除的典型服务

SID	服务名称	期望	熵	超熵	服务反馈数据(响应时间、信誉度、可靠性)
97	FullTextSearch	0.199 7	0.085 6	0.023 3	0.375 5, 0.186 7, 0.355 7, 0.220 2, 0.231 2, 0.179 6, 0.209 6, 0.081 5, 0.318 9, 0.103 7, 0.153 6, 0.189 4, 0.188 8, 0.130 0, 0.072 3
		0.669 7	0.208 9	0.058 7	0.756 1, 0.652 3, 0.910 1, 0.869 1, 0.679 5, 0.688 7, 0.542 1, 0.371 7, 0.827 8, 0.378 0, 0.988 6, 0.699 8, 0.858 8, 0.337 3, 0.486 9
		0.902 3	0.083 9	0.039 7	0.928 3, 0.965 9, 0.920 0, 0.973 1, 0.886 6, 0.778 2, 0.972 5, 0.977 1, 0.954 9, 0.780 1, 0.824 7, 0.813 0, 0.829 3, 0.947 9, 0.984 0
389	SearchService	0.255 9	0.060 6	0.033 7	0.228 1, 0.316 0, 0.317 0, 0.220 8, 0.198 0
		0.808 2	0.099 5	0.034 0	0.819 8, 0.871 7, 0.932 0, 0.798 9, 0.619 0
		0.813 9	0.110 6	0.045 0	0.862 6, 0.695 7, 0.831 4, 0.968 6, 0.711 4
230	radsearch	0.176 7	0.093 9	0.032 8	0.009 6, 0.233 5, 0.136 0, 0.379 1, 0.224 6, 0.183 4, 0.200 4, 0.157 1, 0.067 2
		0.653 1	0.150 4	0.056 0	0.669 5, 0.444 8, 0.527 2, 0.633 4, 0.662 8, 0.657 2, 0.466 8, 0.848 2, 0.968 3
		0.896 6	0.071 5	0.034 7	0.990 4, 0.823 9, 0.979 1, 0.833 3, 0.841 6, 0.940 5, 0.933 6, 0.833 5, 0.894 3
377	SearchCard	0.251 1	0.125 1	0.049 0	0.361 9, 0.338 7, 0.281 1, 0.358 0, 0.154 2, 0.249 2, 0.115 7, 0.036 1, 0.365 0
		0.825 7	0.170 6	0.073 7	0.760 7, 0.556 4, 0.998 0, 0.983 6, 0.762 4, 0.894 4, 0.888 4, 0.976 9, 0.610 8
		0.837 5	0.089 0	0.002 5	0.639 7, 0.864 8, 0.903 0, 0.890 0, 0.827 6, 0.908 0, 0.863 4, 0.725 6, 0.915 9
465	BINDIdSearch	0.196 2	0.086 9	0.028 4	0.341 4, 0.031 0, 0.136 0, 0.175 0, 0.088 0, 0.188 0, 0.199 0, 0.272 0, 0.207 0, 0.343 1, 0.178 0
		0.621 4	0.124 5	0.020 0	0.393 0, 0.765 7, 0.701 3, 0.441 4, 0.495 5, 0.623 8, 0.647 3, 0.609 5, 0.628 1, 0.815 8, 0.714 9
		0.871 7	0.068 9	0.017 8	0.832 1, 0.866 1, 0.858 4, 0.849 4, 0.772 6, 0.967 2, 0.911 2, 0.749 3, 0.897 0, 0.931 8, 0.954 2

表 4 Web 服务选取结果($\delta=20$)

SID	服务名称	推荐度	数量	服务用户评价
232	IpAddressSearch	0.776 7	14	0.378 4, 0.600 4, 0.600 3, 0.556 9, 0.523 9, 0.654 1, 0.523 6, 0.613 5, 0.618 2, 0.880 0, 0.374 4, 0.615 4, 0.538 7, 0.533 9

续表 4

230	FindLocations	0.775 1	16	0.347 1,0.625 1,0.735 4,0.466 2,0.477 5,0.687 5,0.542 9,0.378 2,0.575 5,0.310 9, 0.539 6,0.748 4,0.576 3,0.400 1,0.497 5,0.475 2
262	SearchWebService	0.772 0	14	0.277 1,0.775 4,0.501 4,0.436 5,0.551 9,0.458 1,0.411 2,0.414 0,0.758 4,0.610 7, 0.507 9,0.674 2,0.822 4,0.687 3
377	clsFaqSearchWS	0.745 8	15	0.147 0,0.366 4,0.670 2,0.620 7,0.642 8,0.536 7,0.286 3,0.606 3,0.604 9,0.350 5, 0.565 2,0.314 7,0.728 3,0.289 9,0.590 9
85	SearchLeftBox	0.720 2	14	0.450 9,0.673 4,0.375 1,0.571 0,0.360 9,0.383 1,0.695 3,0.767 7,0.620 7,0.317 4, 0.478 7,0.151 4,0.296 8,0.354 2
27	addressfinder	0.715 1	11	0.193 4,0.347 9,0.490 2,0.563 1,0.628 5,0.559 4,0.603 8,0.473 8,0.677 1,0.624 4, 0.670 8
340	CurrencyCode	0.705 2	13	0.526 0,0.452 7,0.255 9,0.402 0,0.521 4,0.437 5,0.318 8,0.650 8,0.429 8,0.640 9, 0.499 9,0.347 4,0.533 3
380	ProductSearch	0.701 0	9	0.471 9,0.670 7,0.623 0,0.306 3,0.570 1,0.622 1,0.507 7,0.690 1,0.523 8
2	NewsSearch	0.695 9	12	0.344 3,0.353 9,0.578 0,0.429 1,0.564 0,0.616 6,0.650 7,0.306 3,0.395 4,0.611 2, 0.419 9,0.406 3
12	dzSearch	0.692 3	9	0.664 7,0.493 4,0.509 2,0.559 2,0.540 2,0.622 4,0.386 0,0.574 2,0.519 0
152	RoomFinder	0.691 9	9	0.655 7,0.821 9,0.377 0,0.318 7,0.492 1,0.417 1,0.822 2,0.604 1,0.354 5

图 6 比较了在候选服务中随机选取服务 (Random Service Selection, RSS)、不考虑 QoS 数据不确定性直接应用式(6)进行服务选取 (Without QoS Data Uncertainty Decision-making Service Selection, WQDUDSS)和考虑 QoS 数据不确定性的服务选取 (QoS Data Uncertainty Decision-making Service Selection, QDUDSS)三种方法。在服务选取时, K 个服务请求者采用不同的服务选取策略进行服务选取, 在每一轮选取中, 用户根据相应的策略调用服务并反馈使用服务后的 QoS 数据。设服务请求者数量 $K=50$, 实验重复 10 次, 记录每一轮所有用户对选取服务的 QoS 反馈数据。从实验结果来看, 迭代初期服务的 QoS 反馈数据很少, 能够获取的 QoS 数据大都由少数服务使用者和服务发布者提供, 由于服务提供者可能提供不符合实际 QoS 性能数据的 Web 服务, 从而导致不同用户对推荐服务的满意度有较大波动, 会出现采用 WQDUDSS 和 QDUDSS 的服务选取方法的用户满意度低于采用 RSS 方法的现象。然而, 随着迭代次数的增加, 已有的 QoS 数据趋近于服务的真实性能, WQDUDSS 和 QDUDSS 选取方法的用户满意度也相应增加。由于 WQDUDSS 不采用 QoS 属性云不确定算法过滤服务, 应用式(6)虽然能得到较高满意度的服务推荐, 但是用户反馈的 QoS 数据的不确定性会引起用户满意度也发生一定波动。然而, 基于 QoS 属性云的不确定算法会将 QoS 数据不满足用户需求和表

现不稳定的服务过滤掉, 以保证用户满意度一直保持在 0.9 以上, 说明 QDUDSS 在过滤 QoS 数据不确定方面是有效的。

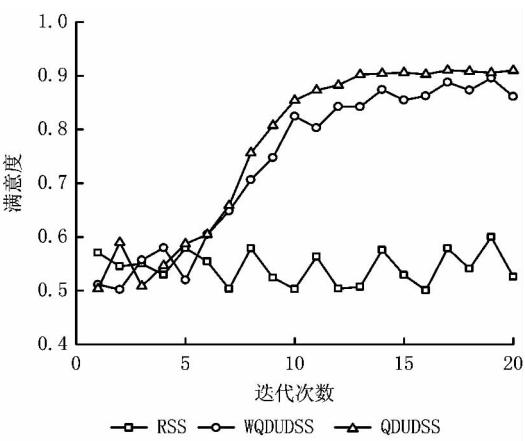


图6 不同选取方法用户满意度比较

4.3.3 实验 3——服务选取效率对比及讨论

随着 Internet 中相同功能的 Web 服务越来越多, 服务选取的计算时间也相应地增加。本节讨论候选服务数和 QoS 属性数量对服务选取效率的影响, 服务选取的效率采用服务选取所使用的时间来衡量, 主要包括 QoS 不确定算法的运算时间、QoS 数据规范化、QoS 数据聚合和用户推荐度的计算时间。因为 QoS 约束条件主要受用户主观因素的影响, 不同 QoS 约束条件下不确定算法过滤掉的服务数不同, 相应的计算时间也就有所不同, 所以在分析

候选服务数量和 QoS 属性数量对 QoS 属性不确定算法计算时间的影响时,设置约束条件为一个 QoS 属性,并且将属性的期望值作为 QoS 约束条件。当候选服务数增加时,服务计算时间的变化如图 7 所示。当候选服务小于 50 时,采用 QoS 不确定算法的服务计算时间略大于没有采用 QoS 不确定算法的时间,主要因为候选服务较少时过滤的服务较少,而 QoS 不确定决策时间的开销比较大;当 Web 服务数大于 50 时,没有采用 QoS 不确定算法进行服务选取的时间开销基本上随候选服务数量的增加呈正比例增加,而基于 QoS 不确定算法的运行时间增加较慢,主要因为 QoS 不确定算法过滤了大部分候选服务,缩小了服务检索空间,节省了服务计算时间。

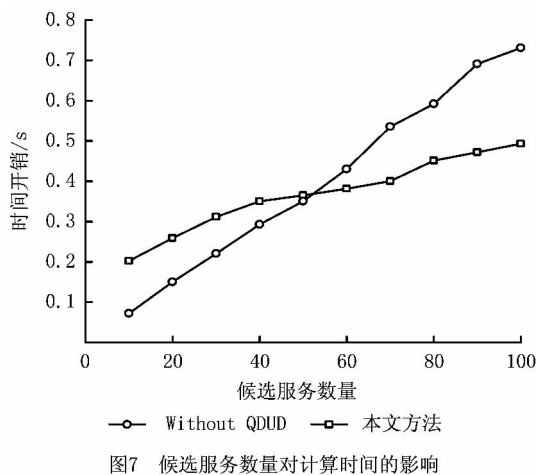


图7 候选服务数量对计算时间的影响

图 8 所示为候选服务为 50 的情况下,采用 QDUD 算法和不采用此算法时,服务选取时间随 QoS 属性约束条件的变化比较。不采用 QDUD 选取方法的计算时间基本保持在 0.35 s 之间,而当 QoS 属性约束条件增加时,QDUD 算法需要判断每一个 QoS 属性数据的不确定性,这将相应地增加不确定决策的时间。然而,这种增幅不是很大,而且 QoS 属性的增加在很大程度上提高了 QDUD 算法的适应性。此外 QoS 属性的增加,使用户对服务的描述更完全,减少了 QoS 模型的不完整性导致的 QoS 数据对推荐结果的影响。因此,QoS 属性的增加对提高算法扩展性和最终服务推荐结果准确性的优势大于服务选取方法时间开销的小幅增加。

5 结束语

本文探讨了 QoS 数据呈现出不确定性的原因,

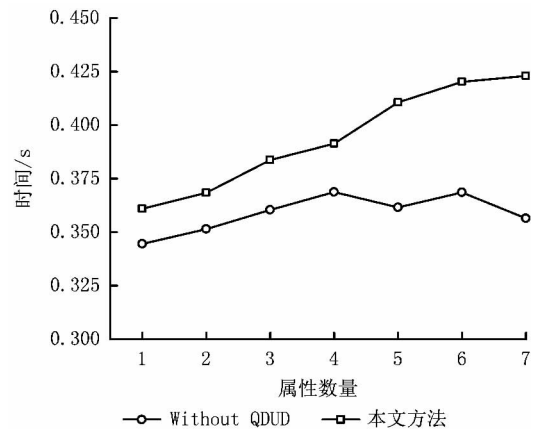


图8 QoS属性数量对计算时间的影响

建立了 QoS 属性云模型,利用 Ex , En , He 三个数字特征预测服务 QoS 数据的分布情况,并使用 Ex 过滤不满足用户期望的服务, En 和 He 过滤 QoS 数据波动比较大的服务,减少了 QoS 数据不确定性对服务选取的影响,保证了推荐结果的准确性和可靠性。在此基础上引入 QoS 属性相似度作为计算用户服务评价聚合值的权重,综合用户评价和服务使用者数量计算服务的推荐度,给出了优选服务的排序,并以原型系统、实际数据和多个实验分析说明了方法的有效性。未来将进一步研究 Web 服务 QoS 属性数据不确定性之间的关系优化不确定算法,完善原型系统,支持更多 Web 服务的分类和 QoS 数据采集。

参考文献:

- [1] PAPA ZOGLOU M P, TRAVERSO P, DUSTDAR S, et al. Service-oriented computing: state of the art and research challenges [J]. Computer, 2007, 40(11):38-45.
- [2] RAN S. A model for Web service discovery with QoS [J]. ACM SIGecom Exchanges, 2003, 4(1):1-10.
- [3] ZHANG Yinghui, FU Fengrui, RU Chao. Semantic Web service selection based on QoS ontology[C]//Proceedings of the 2010 International Conference on Information, Network and Automation. Washington, D. C., USA: IEEE Computer Society, 2010:2492-2496.
- [4] LIU Y T, NGU A H, ZENG L Z. QoS computation and policing in dynamic Web service selection[C]//Proceedings of the 13th International World Wide Web Conference. New York, N. Y., USA: Association for Computing Machinery, 2004: 798-805.
- [5] PAN Jing, XU Feng, LU Jian. Reputation-based recommender discovery approach for service selection[J]. Journal of Software, 2010, 21(2):388-400 (in Chinese). [潘静,徐锋,吕

- 建. 面向可信服务选取的基于声誉的推荐者发现方法[J]. 软件学报, 2010, 21(2):388-400.]
- [6] ZHU Rui, WANG Huaimin, FENG Dawei. Trustworthy services selection base on preference recommendation [J]. Journal of Software, 2011, 22(5):852-864 (in Chinese). [朱锐, 王怀民, 冯大为. 基于偏好推荐的可靠服务选择[J]. 软件学报, 2011, 22(5):852-864.]
- [7] BOUTILIER C, BACCHUS F, BRAFMAN R I. UCP-networks: a directed graphical representation of conditional utilities (UAI-2001) [C]//Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence. San Francisco, Cal., USA: Morgan Kaufmann, 2001:56-64.
- [8] BRAFMAN R I, DOMSHLAK C, SHIMONY S E. On graphical modeling of preference and importance [J]. Journal of Artificial Intelligence Research, 2006, 25(3):389-424.
- [9] STEPHAN R M, YU H Q, TILLY M. Service selection based on non-functional properties [C]//Proceedings of International Conference on Service-Oriented Computing (ICSOC 2007). Berlin, Gemany: Springer-Verlag, 2009:128-138.
- [10] YU Hongqing, STEPHAN R M. A method for automated Web service selection [C]//Proceedings of IEEE Congress on Services. Washington, D. C., USA: IEEE Computer Society, 2008:513-520.
- [11] LI Deyi, LIU Changyu, DU Yi, et al. Artificial intelligence with uncertainty [J]. Journal of Software, 2004, 15(11):1583-1594 (in Chinese). [李德毅, 刘常昱, 杜 鹄, 等. 不确定性人工智能[J]. 软件学报, 2004, 15(11):1583-1594.]
- [12] LI Feng, SHEN Limin, SI Yali, et al. A trust forecasting model based on entity context and time tamp [J]. Journal of Electronics & Information Technology, 2011, 33(5):1217-1223 (in Chinese). [李 峰, 申利民, 司亚利, 等. 一种基于实体上下文和时间戳的信任预测模型[J]. 电子与信息学报, 2011, 33(5):1217-1223.]
- [13] LI Deyi. Artificial intelligence with uncertainty [M]. Beijing: National Defense Industry Press, 2005:171-177 (in Chinese). [李德毅. 不确定性人工智能[M]. 北京:国防工业出版社, 2005:171-177.]
- [14] LI Deyi, MENG Haijun, SHI Xuemei. Membership clouds and membership cloud generators [J]. Journal of Computer Research and Development, 1995, 32(6):15-22 (in Chinese). [李德毅, 孟海军, 史雪梅. 隶属云和隶属云发生器[J]. 计算机研究与发展, 1995, 32(6):15-22.]
- [15] MEI Hongbiao, WANG Jian. Cloud model for data distribution management [J]. Computer Integrated Manufacturing Systems, 2011, 17(3):540-551 (in Chinese). [梅宏标, 王 坚. 基于云模型的数据分发管理[J]. 计算机集成制造系统, 2011, 17(3):540-551.]
- [16] LU Huiying, CAO Yuanda, LIU Yulong. Subjective trust model for collaborative commerce environment [J]. Computer Integrated Manufacturing Systems, 2007, 13(8):1545-1551 (in Chinese). [吕慧颖, 曹元大, 刘玉龙. 协同商务环境中的主观信任模型研究[J]. 计算机集成制造系统, 2007, 13(8):1545-1551.]
- [17] ZHANG Wei, PAN Xiaohong, LIU Zhi, et al. Manufacturing service scheduling strategy based on cloud model ant colony optimization [J]. Computer Integrated Manufacturing Systems, 2012, 18(1):201-207 (in Chinese). [张 卫, 潘晓弘, 刘 志, 等. 基于云模型蚁群优化的制造服务调度策略[J]. 计算机集成制造系统, 2012, 18(1):201-207.]
- [18] LI Deyi, CHEUNG D, SHI Xuemie, et al. Uncertainty reasoning based on cloud models in controllers [J]. Computer & Mathematics with Applications, 1998, 35(3):99-123.
- [19] KSALAN M, ZIONT S. Multiple criteria decision making in the new millennium [C]//Proceedings of the 15th International Conference on Multiple Criteria Decision Making (MCDM). Berlin, Germany: Spring-Verlag, 2001:1-14.
- [20] ZHENG Zibin, ZHANG Yilei, LYU M R. Distributed QoS evaluation for real-world Web services [C]// Proceedings of the 8th International Conference on Web Services (ICWS2010). Washington, D. C., USA: IEEE Computer Society, 2010:83-90.
- [21] ZHANG Yilei, ZHENG Zibin, LYU M R. Exploring latent features for memory-based QoS prediction in cloud computing [C]//Proceedings of the 30th IEEE Symposium on Reliable Distributed Systems (SRDS 2011). Washington, D. C., USA: IEEE Computer Society, 2011:1-10.
- [22] SAHA S, MURTHY C A, PAL S K. Classification of Web services using tensor space model and rough ensemble classifier [C]// Proceedings of the 17th International Symposium on Foundation of Intelligent Systems (ISMIS'08). Berlin, Germany: Springer-Verlag, 2008:508-513.

作者简介:

申利民(1962—),男,黑龙江佳木斯人,教授,博士生导师,研究方向:协同计算, E-mail: shenlmm@sina.com;
陈 真(1987—),男,陕西宝鸡人,硕士研究生,研究方向:Web 服务计算;
李 峰(1978—),男,山东德州人,博士研究生,研究方向:可信计算、动态信任管理。