

## Web 服务选择中信誉度评估方法\*

王尚广<sup>+</sup>, 孙其博, 杨放春

(北京邮电大学 网络与交换技术国家重点实验室, 北京 100876)

### Reputation Evaluation Approach in Web Service Selection

WANG Shang-Guang<sup>+</sup>, SUN Qi-Bo, YANG Fang-Chun

(State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China)

+ Corresponding author: E-mail: sguang.wang@gmail.com

**Wang SG, Sun QB, Yang FC. Reputation evaluation approach in Web service selection. Journal of Software, 2012, 23(6): 1350-1367. <http://www.jos.org.cn/1000-9825/4051.htm>**

**Abstract:** In an open Web service environment, it is difficult to guarantee the authenticity of the feedback ratings reported by users after they have already consumed some services, which leads to large deviation between evaluation results and actual values of service reputation. Therefore, there are much more failures in service selection. To address the above problem, this paper proposes a reputation evaluation approach for Web service selection. The main idea of the proposed approach is evaluating the reputation by three reputation evaluation model including feedback checking, feedback adjustment, and feedback detection. Simulation results demonstrate the proposed approach not only improves the objectivity of reputation evaluation effectively, but also reduces the deviation of service selection significantly.

**Key words:** Web service; service selection; reputation; QoS; HFS; CUSUM

**摘要:** 在开放的 Web 服务环境中,由于无法保证用户使用服务后给出的反馈等级是真实可靠的,导致服务的信誉度评估结果与实际值存在较大偏差,进而服务选择失败率较高.为了克服上述问题,提出了一种用于 Web 服务选择的信誉度评估方法.该方法的主要思想是,通过反馈核查、校正和检测这 3 个信誉度评估模块对服务的信誉度进行评估.仿真结果表明,所提出的方法不仅有效提高了信誉度评估的客观性,还显著减小了服务选择的偏离度.

**关键词:** Web 服务;服务选择;信誉度;QoS;层级模糊系统;累积和

**中图法分类号:** TP311      **文献标识码:** A

Web 服务作为一种新型的分布式计算模型和协作关系,为动态、跨组织的资源协同共享创造了前所未有的机会,得到了工业界和学术界的极大关注.随着 Web 服务标准和技术的逐步成熟,开发和部署 Web 服务已经变得更加容易.Web 服务的最终目标是,在面向服务的体系架构(service oriented architecture,简称 SOA)中,根据用户提出的需求,将分散、独立的 Web 服务进行无缝地自动组合.

\* 基金项目: 国家自然科学基金(60872042, 60821001); 国家重点基础研究发展计划(973)(2009CB320406); 国家高技术研究发展计划(863)(2008AA01A317); 核高基国家科技重大专项(2009ZX01039-001-002-01)

收稿时间: 2010-05-11; 修改时间: 2011-01-12; 定稿时间: 2011-04-28

随着 Web 服务在 e-Economy,e-Science,e-Government 等多个领域的蔓延,满足相同功能要求的服务越来越多,使得在为提供服务组合时不仅要考虑功能方面的需求,而且还要考虑到非功能属性服务 QoS(quality of service)<sup>[1,2]</sup>的需求,如信誉度、可用性、响应时间、费用等.因此在服务组合过程中,担任重要角色的基于 QoS 感知的 Web 服务选择技术引起了许多研究者的注意,并获得了较好的研究成果<sup>[1-9]</sup>.

然而在开放的 SOA 环境中,由于受多种网络的涌现(如 Ad-hoc,P2P,WLAN 等)、利益化的市场竞争以及 Web 服务的逻辑流程和资源管理对外的封闭性<sup>[10]</sup>等因素的影响,不能保证服务提供者发布的服务是客观、真实、可信的<sup>[6]</sup>.因此,通过对用户使用服务后给出的反馈等级进行评估(计算)得到的服务信誉度属性值,则成为了用户选择服务时考虑的一个重要参考指标.即被选择的服务除了满足常规的 QoS 要求外(如费用、响应时间),还要满足用户对服务信誉度的额外要求<sup>[11-13]</sup>.所以在开放的 Web 服务环境下,研究具有准确性和客观性的服务信誉度评估方法具有重要的现实意义.

目前,如何根据用户的反馈等级对服务信誉度进行准确、客观的评估以及获得高质量的组合服务,已得到了许多研究者的着重关注.尽管研究者<sup>[11-25]</sup>提出的多种解决方案有效地提高了信誉度评估的准确性和客观性,加快了 Web 服务组合的应用进程,但仍然面临以下挑战:

- 首先,由于 Web 用户的急剧增加(截止到 2009 年 12 月 31 日,Web 用户的数量已增加到全球人口总数的 26.6%),Web 服务用户分布广泛,知识、素质、能力各不相同,而无法保证每个用户使用服务后都具有提供合格反馈等级的能力.对于某个被使用过的服务,那些拥有丰富知识或对 Web 服务较为熟悉的用户,反馈的数据将比那些缺乏相应知识或能力用户的反馈更为可信<sup>[26]</sup>.由于缺乏上述能力的用户给出的反馈等级往往呈现出一定的杂乱或混乱性,本文将该类型的反馈称为混乱反馈.因此,为确保信誉度的准确性,在信誉度评估过程中,对于缺乏相应能力用户的混乱反馈并不能直接用于服务的信誉度评估;
- 其次,由于服务网络环境和运行平台样(如 Phones,Laptops,PDAs 等)的多样化,服务用户在使用某个服务后,给出的反馈等级常常受到用户偏好、上下文信息的影响<sup>[27]</sup>.对于该类型的反馈,本文称为偏好反馈.在偏好反馈类型中,对于相同的 Web 服务,一些宽容的用户给出的反馈据相对比较宽松,一些苛刻用户给出的反馈则相对保守,而其他的用户则居于中间<sup>[12]</sup>.因此,这就需要信誉度评估方法应该能够屏蔽或减弱不同用户在对其使用过的服务进行评估时附着的主观因素,同时还要考虑用户上下文信息对用户反馈所产生的影响;
- 最后,一些用户由于受市场竞争或商业利益的驱动,他们或许通过正常身份或女巫攻击(sybil attacks)<sup>[12,28,29]</sup>方式给予某些服务极高或极低的恶意反馈,导致反馈数据充斥虚假泡沫或遭到恶意打压.然后,由于常用的平均值方法缺少能够区分或剔除这些恶意反馈数据的方法,使得服务的信誉度与实际值偏离较大.对于该类型的反馈,本文称为恶意反馈.存在的恶意反馈表明,现有的方法中通常假定服务用户给出的反馈数据都是真实可信的这一假设在实际中往往很难保证<sup>[17]</sup>.

显然,如果无法克服上述挑战,将到无法获得准确、可靠的信誉度评估结果,从而使选择的服务偏离用户请求,导致用户满意度下降,甚至服务组合失败.因此,为了解决上述挑战,本文提出了一种开放分布式 Web 环境下服务选择的信誉度评估方法,即 REAWSS(reputation evaluation approach in Web service selection).REAWSS 的核心是一个具有反馈核查、反馈校正和反馈检测等 3 个功能模块的信誉度评估模型.其中:反馈核查的工作原理是,通过对进入 Web 环境中的服务用户建立档案,以对每个用户的反馈能力或准确性进行核查,然后根据核查结果有针对性地选择能力较强用户的反馈数据用于评估,从而达到减弱能力较差用户的反馈数据在信誉评估中所占权重的目的;反馈校正的工作原理是,通过对用户上下文推理,选择与该用户存在使用过相同 Web 服务或服务集的用户并根据用户之间的反馈等级和用户上下文信息计算反馈和谐度,然后将获得的反馈和谐度用于计算当前用户的虚拟反馈,并与实际反馈进行比较,对反馈冲突度较大的反馈进行相应处理,从而达到弱化用户偏好(主观因素)和考虑用户上下文(环境因素)的目的;反馈检测的工作原理是,采用累积和算法检测抽样间隔内反馈水平与正常情况下反馈水平的平均阶跃来发现恶意反馈,然后通过丢弃抽样间隔内的反馈等级来防止恶意反馈对信誉度的污染,从而达到提高信誉度评估准确性的目的.在搭建的仿真环境中的实验结果表明, REAWSS

具有较好的核查、校正和检测能力,不但能够准确地评估单个服务的信誉度,而且显著地降低了服务选择结果的偏离度.

本文第 1 节介绍相关工作.第 2 节介绍提出的 REAWSS 方法,包括反馈核查、反馈校正和反馈检测.第 3 节给出仿真实验,包括仿真环境搭建、实验结果对比、模块分析及结果讨论.最后一节总结全文并指出下一步研究内容.

## 1 相关工作

针对 Web 服务的信誉度评估问题,许多研究者提出了相关的框架和解决思路,并取得了较好的研究成果.

文献[1]提出了一种基于混合整数规划的自适应服务组合方法.该方法充分考虑了包括信誉度在内的 5 种 QoS 属性对服务选择结果的影响,并且当用户的 QoS 请求无法得到满足时(无解),能够通过用户与服务提供者对 QoS 的协商(negotiation)进行二次寻优(reoptimization),以发现具有 QoS 端到端保障的组合服务.虽然该方法具有较好的服务选择能力及自适应性,但由于作者提出的信誉度聚合函数缺少考虑用户反馈等级的真实性,导致服务组合结果无法满足用户对信誉度端到端的约束而偏离了用户需求(用户提交的信誉度请求与组合服务的实际信誉度不一致).即使该方法中的服务协商能够起到弥补作用,但如果服务提供者本身就是恶意或不诚实的(通过提供虚假或恶意服务来获得非法利益)<sup>[17]</sup>以及存在的恶意反馈<sup>[11]</sup>,将导致二次寻优的组合服务仍然偏离用户的信誉度要求.

文献[2]着重分析了准确高效完成基于语义 QoS 的 Web 服务描述和选择的必要条件.作者根据 QoS 属性在服务管理中扮演的角色,将信誉度划为领域无关 QoS 属性(domain-independent QoS attributes)中配置管理(configuration management)类型,同时给出了信誉度的计算方法.尽管该研究显著提高了基于语义 QoS 感知的 Web 服务发现进程.但与文献[1]相似,由于缺少从数据根源层面上考虑用户反馈等级客观性、真实性,导致了计算出的服务信誉度与实际值偏离较大.

文献[11]提出了一个面向开放分布式服务环境的基于信誉度的服务可信管理框架,核心是可信服务管理(trust management service,简称 TMS).TMS 不仅支持多种不同实体的信任关系,同时还支持每个实体采用不同的信誉度评分函数(scoring function)对同样的反馈数据进行评估.该框架的一个显著优点是支持多种信誉度评估方法,具有较高的实用价值.但也存在以下两个缺点:

- 虽然 TMS 将包含客户端、服务、反馈值(等级)最优属性集等 4 种元素项在内的服务调用历史记录(service invocation history record)用于信誉度评估,而比文献[1,2]中对反馈等级的直接计算获得的信誉度更准确、合理.但由于该记录,无法对用户反馈中包含的主观感觉、用户偏好等对用户反馈影响较大的因素信息进行识别、记录或处理,导致被信誉度评估中充斥大量的主观因素而导致信誉度准确性下降;
- 该框架缺少对环境因素(用户上下文)的支持.不同上下文环境下,用户对同一服务给出的反馈不尽相同,因此在信誉度评估中不考虑用户上下文对反馈等级的影响,以及对其反馈等级的区分或区别对待,将导致信誉度评估的客观性丧失<sup>[13,24]</sup>.

文献[12]提出了一个基于信誉度感知的服务评价和选择框架.框架中的服务评价采用了市场营销学(market science)中期望失验理论(expectancy-disconfirmation theory),用于研究用户满意度与服务质量之间的关系来得到用户反馈(该反馈来自服务质量感知函数(perception function)和失验函数(disconfirmation function)的综合,而不是用户的直接反馈).为了克服用户对所用服务的质量期望(expectation)和感知(perception)掺杂的主观感觉,使其达到客观、公平和可度量,作者使用服务水平协商(service level agreement,简称 SLA)来量化用户的服务质量期望,采用质量监测(quality monitoring)量化用户的服务质量感知.同时,该评价方法中信誉度计算支持对信誉度的预测.另外,在服务评价方法的基础上,作者提出了一种服务选择算法,该算法的亮点是将服务选择中的多属性决策问题转化为仅考虑 3 个属性的服务选择,即信誉度、质量、花费,并进一步将上述 3 种属性其聚合为单属性服务选择,提高了服务选择的性能.由于该框架中的信誉度系统是基于服务(per-service)而非用

户(per-customer),从而在设计之初就考虑了用户主观感觉、恶意反馈(sybil attacks)<sup>[28]</sup>对信誉度准确性和客观性的影响.虽然上述框架对服务的信誉度具有较高的评价优势,但也存在以下3点不足:

- 服务协商中过多地考虑用户的服务质量期望,忽略了用户的上下文信息,将造成失验函数(disconfirmation function)值较小而导致信誉度的客观、公正性下降.例如,用户在智能手机上运行多媒体服务,在服务协商中,他显然想得到与固定运行环境一样的服务质量.此时,服务协商中虽然满足了用户的要求为其提供了固定运行环境下信誉度较好的服务,但由于没有考虑其上下文环境(该服务不能得到相应的软硬件支持),导致用户质量期望与服务感知差别较大,服务反馈明显降低,造成在固定环境下运行信誉度较高的服务的信誉度下降;
- 虽然 SLA 在一定程度上能够遏制用户直接的恶意反馈,但仍然无法拒绝间接的 Sybil 攻击.例如,虽然用户在该系统中无法直接对服务进行反馈,但仍然能够通过 Sybil 攻击盗用或伪造多个用户.由于用户并不真正在意服务的协商结果,同样人为导致质量期望与服务感知差别较大,使得服务评价方法误认为该服务质量低下,而被误导地降低其反馈等级,从而能够间接达到 Sybil 攻击的目的.另外,在开放的 Web 环境中,对于伪造的服务用户,将无法在 SLA 中为其在服务提供者 and 用户之间建立统一的 QoS 度量标准<sup>[13]</sup>,同样导致 SLA 失效;
- 虽然信誉度计算中采用的简单指数平滑(simple exponential smoothing)方法对服务的信誉度具有预测能力,但由于其用户反馈本身的不可预知性,导致其预测准确度较低.

北京大学的梅宏团队<sup>[17]</sup>对基于 QoS 感知的 Web 服务选择进行了探讨,针对目前 QoS 计算中使用的质量属性过于局限以及缺少确保其可信的计算机制这两个问题,提出了一种考虑 QoS 数据可信性的服务选择方法.该方法的核心思想是,在确保质量属性可信的基础上,使 QoS 的计算相对准确,最终使服务选择结果更加可信,取得了较好的研究成果.

华中科技大学的徐兰芳等人<sup>[30]</sup>针对传统分析方法中存在恶意客户的虚假推荐,提出了以灰色系统理论为基础、以灰色聚类评估算法为主要内容的信誉报告机制方案.该方案克服了传统分析中每个用户的评级权视为相等的做法,能够使数据更为客观,具有评估可靠、可操作性强等优点.

其他的研究还有基于信誉度感知的服务组合<sup>[14]</sup>、增强信誉度的服务发现<sup>[15]</sup>、针对新加入服务的使用信息较少而无法准确地计算信誉度的情况提出的 Web 服务的信誉度评估方法,即通过用户之间的互相评估(P2P 方式)来评估信誉度<sup>[16]</sup>、REMAN<sup>[25]</sup>等.

## 2 信誉度评估方法(REAWSS)

**定义 1.** 反馈等级是用户为了表达对所使用服务的较为全面的感知而给出的一个值或向量,它常常受所服务的性能、可靠性等因素影响.

**定义 2.** 信誉度是众多 QoS 属性中的一种属性,它是所有服务用户在使用某个服务后给出的用户感知的概述,而这种概述一般是通过用户的反馈等级来量化表达.例如,服务  $s_j$  在某个抽样时间间隔内的信誉度  $q(s_j)$  可以通过下式(1)获得:

$$q(s_j) = \frac{\sum_{i=1}^n w_{j,i} \times r_{j,i}}{n} \quad (1)$$

$n$  是该抽样间隔内反馈等级的数量; $r_{j,i}$  表示第  $i$  个用户对服务  $s_j$  的反馈等级,本文中的取值为  $r_{j,i} \in [1, 10]$ ;  $w_{j,i} \in [0, 1]$  表示该反馈等级的权重.

由于在式(1)中无法保证用户反馈  $r_{j,i}$  的真实可靠,导致信誉度评估结果与其真实值偏离较大,经常导致组合服务执行失效或失败,用户满意度下降.为了解决上述问题,我们提出一种用于 Web 服务选择中的信誉度评估方法,即 REAWSS.REAWSS 中的信誉度评估模型是对我们早期提出的分布式的 Web 服务 QoS 注册系统 Q-peer<sup>[31-33]</sup>的扩展,即在 Data collector 和 Repository 之间增加了信誉度评估模型.它主要用来纠正或消除用户

反馈中的混乱反馈、偏好反馈及恶意反馈对 Web 服务信誉度评估的负面影响,进而增强服务选择结果的信誉度,提高用户满意度.基本原理是,利用 REAWSS 对来自数据收集器(data collector)中的用户反馈数据进行信誉度评估,并将评估结果(信誉度)存储到 QoS 存储器(repository)中,以便和其他类型的 QoS 属性一起用于 QoS 感知的 Web 服务选择.

如图 1 所示,我们提出的 REAWSS 主要包含 3 个模块.模块 1 即反馈核查(feedback checking)的目的是解决 Web 服务选择面临的混乱反馈,其原理是通过进入 Web 环境中的服务用户建立档案(survey form),通过该档案可以对每个用户的反馈能力或准确性进行核查,然后根据核查结果有针对性地选择能力较强用户的反馈数据用于评估或减弱能力较差用户的反馈数据在信誉评估中的权重.模块 2 即反馈校正(feedback adjustment)的目的是解决偏好反馈,其原理是通过对用户上下文推理(context inference),选择与该用户存在使用过相同 Web 服务或服务集的用户并根据用户之间的反馈等级和用户上下文信息计算反馈和谐度(feedback coordination),然后将获得的反馈和谐度用于计算当前用户的虚拟反馈,并与实际反馈进行比较,对反馈冲突度较大的反馈进行相应处理,从而达到弱化用户主观因素(偏好)和考虑用户环境因素(上下文)的目的,增强反馈数据的客观性和公正性;同时,该模块对部分用户的恶意反馈也具有一定的校正作用.模块 3 即反馈检测(feedback detection)的目的是应对恶意反馈,其原理是,引入统计学中的改变点检测理论,将其中的 CUSUM(cumulative sum)算法用于检测用户的恶意反馈.它通过检测抽样间隔内反馈水平与正常情况下反馈水平的平均阶跃来发现恶意反馈,然后通过丢弃抽样间隔内的反馈等级来防止恶意反馈对信誉度的污染,从而达到提高信誉度评估准确性的目的.

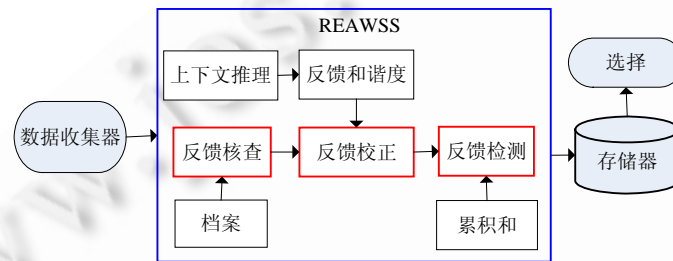


Fig.1 Framework of the REAWSS

图 1 REAWSS 框架

与文献[13,15,24,34]相似,在本文的研究中,Web 服务被众多用户使用,服务代理能够收集到大量的反馈数据.对于反馈数据较少的情况,超出本文的研究范围,而北京大学的梅宏团队<sup>[26]</sup>在该领域已经取得了较为显著的研究成果.另外,有关 Q-peer 及其 Data collector 和 Repository 等其他功能实体的更多介绍可见文献[31,33].

## 2.1 反馈核查

反馈核查意味对来自用户的反馈数据在用于信誉度评估前对其进行核对或检查,其目的是解决当前 Web 服务选择存在的用户混乱反馈,即减弱缺乏相应知识或能力的用户反馈所在信誉度评估中权重来提高信誉度的可信性,其功能实现依赖用户反馈档案.

**定义 3.** 反馈档案是一个包含四元组的用户反馈历史记录,即  $SF = \langle ID_i, FB_{i,j}, CL_{i,j,k}, Ratio_i^k \rangle$ . 其中,  $ID_i$  表示第  $i(i=1,2,\dots,n)$  个用户(如  $ID=2$  表示第 2 个用户),  $FB_{i,j}$  表示  $ID_i$  对第  $j(j=1,2,\dots,n)$  个 Web 服务的反馈等级(如  $FB=6$ ),  $CL_{i,j,k}$  表示  $ID_i$  给出的  $FB_{i,j}$  的可信水平为  $k$  级(本文中  $k=1,2,\dots,5$ , 最高级为 5),  $Ratio_i^k$  表示  $ID_i$  在可信水平  $CL_{i,j,k}$  上的反馈正确率.

根据定义 3,反馈等级的可信水平  $CL_{i,j,k}$  被量化为共  $k$  个等级,反馈正确率  $Ratio_i^k$  可由下式获得:

$$Ratio_i^k = \frac{RNC_j^k}{\sum_{j=1}^n NC_j^k} \quad (2)$$

其中,  $RNC_j^k$  表示  $ID_i$  在可信水平  $CL_{i,j,k}$  上给出的正确反馈的数量(正确反馈是指  $|FB_{i,j}-q(s_j)|/q(s_j) \leq cf, q(s_j)$  是评估后的信誉度; $cf$  是反馈等级正确的门限,一般取值不大于 0.01),  $\sum_{j=1}^n NC_j^k$  表示历史记录中  $ID_i$  在可信水平  $CL_{i,j,k}$  上给出的所有反馈的数量。

令  $Ratio_k'$  为具有相应能力的用户在第  $k$  个可信水平上给出的反馈等级正确率的门限值,如果用户  $ID_i$  的  $Ratio_i^k \geq Ratio_k'$ , 则意味着该用户给出的反馈等级是可靠的(反馈能力得到认可),可以直接进入模块 2(第 1.2 节);否则,如果该用户的反馈正确率小于设定门限值,则该用户的  $FB_{i,j}$  在公式(1)中的权重  $w_i$  将被赋予一个较小的值(用于减少了其对信誉度评估的可信性影响)并进入模块 2。 $w_i$  的赋值在本文中分为 0.1,0.2,0.4,0.4,0.5 这 5 个等级值(可以根据具体情况调整设置),每个等级值分别与 5 个可信水平等级相对应。例如在 1 级可信水平上,  $w_i=0.1$  表示该用户的反馈权重被极大降低,意味着  $ID_i$  的反馈能力极差,如果不对其赋予较小的权重,将会造成信誉度评估严重偏离其实际值。

本文中的模块 1 主要是通过反馈档案中的用户反馈正确率和可信水平来综合判断用户是否具有相应的反馈能力,对不符合要求的用户反馈,弱化其在信誉度评估中的权重。另外,本文建立的反馈档案明显地比简单地使用可信水平或直接地使用反馈等级更为合理,从而为信誉度评估可信、准确的数据来源提供了保障。然而,尽管该模块能够通过调节  $Ratio_k'$  的取值将混乱反馈予以减弱,但这并不能保证通过反馈核查的用户反馈一定是客观准确的。由于受到用户偏好或上下文的影响导致反馈核查无论是从粒度还是力度方面都显得较弱,因此需额外模块对来自模块 1 的数据进行校正,所以模块 2(反馈校正)被设计出来。

## 2.2 反馈校正

反馈校正意味对受用户偏好和上下文影响较大的反馈等级进行校对或调整,其目的是解决当前 Web 服务选择中存在偏好反馈,即调整用户反馈数据使其更准确,其功能实现依赖于用户上下文推理及反馈和谐度计算。具体实现如图 2 所示,通过层级模糊系统获得用户上下文信息后,计算与该用户上下文相近的使用过相同服务集的用户之间的反馈和谐度,然后将其作为该用户反馈数据的权重计算其虚拟反馈,最后通过与用户实际反馈进行对比得到反馈冲突度,并根据其所在区间范围做相应处理。

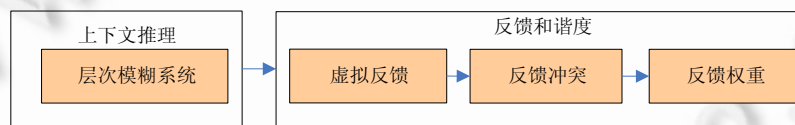


Fig.2 Feedback adjustment

图 2 反馈校正

### 2.2.1 用户上下文推理

该模块中的用户上下文推理意味着从用户运行服务期间的多个不同类型、不同意义的上下文中推理出一个能代表该用户上下文信息的一个综合量化值。由于用户上下文在领域或应用方面不尽相同,迄今还没有统一的定义。根据本文的研究及其要解决的科学问题,首先给出用户上下文定义。

**定义 4.** 用户上下文是指用户运行 Web 服务所处的网络和终端环境,可以用一个五元组表示  $UC=(Band_i, Power_i, CPU_i, Delay_i, Smooth_i)$ 。其中,  $i$  表示第  $i$  个用户在使用该服务,  $Band_i$  表示网络的下行带宽,  $Power_i$  表示服务运行终端的电量水平,  $CPU_i$  表示终端的 CPU 的利用率,  $Delay_i$  表示网络延迟,  $Smooth_i$  表示服务运行的平滑度(如对于多媒体服务中的平滑度=1-丢包率)。

从定义 4 可以看出,  $UC$  中不同元素取值的单位、范围及其所代表的意义不尽相同。如果将每个元素遍历化地用于表示用户上下文,显然无论从系统的性能负载还是应用开发方面都是无法接受的。因此,本文通过综合信息值  $UCV$  来表示用户上下文,即该值是对用户上下文信息的综合量化。本文中,  $UCV$  将通过层次模糊系统(hierarchical fuzzy system,简称 HFS)对用户上下文推理来获得,如图 3 所示。有关 HFS 的更详细介绍见文献[35]。



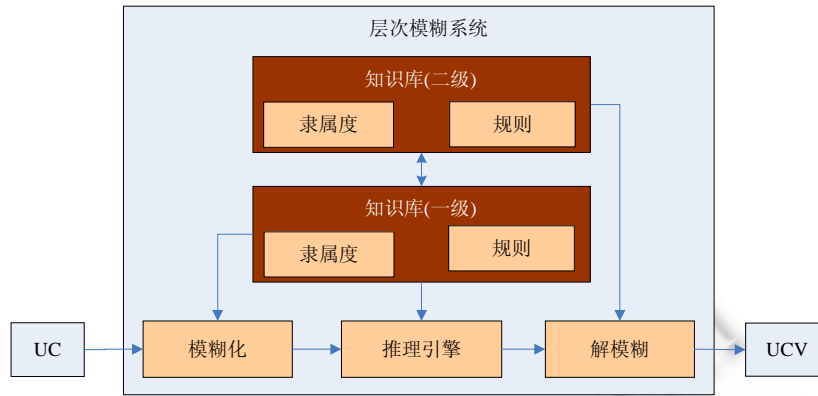


Fig.3 Framework of context inference

图3 用户上下文推理框架

本文中,HFS 通过以下 3 个步骤对  $UC$  进行推理得到用户上下文信息的综合直  $UCV$ :

步骤 1. 模糊化(fuzzification),即对 5 个输入变量( $UC$  中的 5 个元素)和输出变量  $UCV$  进行模糊化.设其模糊子集为  $\{L,M,B\}$ ,其相应的语言变量为: $L$ (little)=小, $M$ (middle)=中, $B$ (big)=大.这样将输入和输出变量量化为不同等级,然后使用知识库(knowledge base)中的隶属度函数根据其对应的隶属度(membership)进行模糊化.本文采用三角隶属度函数对所用变量进行模糊化,例如,图 4(a)和图 4(b)分别给出了  $Smooth$  和  $UCV$  的隶属度函数.同理,其他变量的隶属度函数与其类似,在此略去.

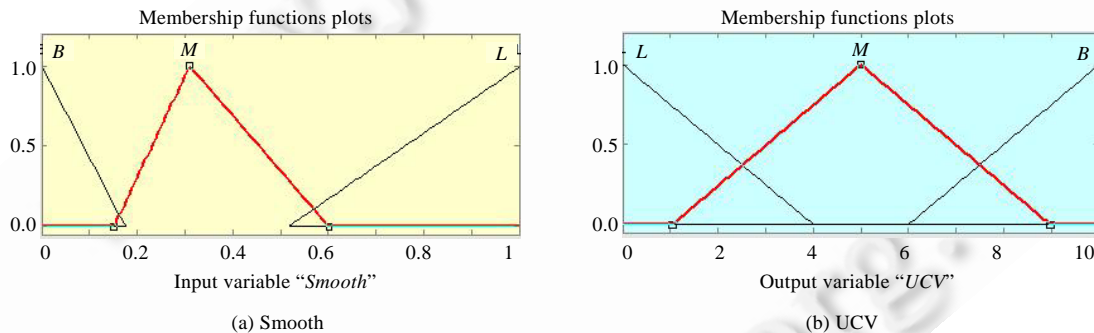


Fig.4 Membership functions

图4 隶属度函数

步骤 2. 推理引擎(inference engine),即对  $UC$  推理.图 5 给出了 HFS 中用户上下文推理的层级逻辑关系.本文选取 IF-THEN 的形式,此时规则库(rule base)中共制定了  $3^2+3^3+3^2=45$  条模糊规则.根据这些规则,可以通过图 3 中的推理引擎得到输出的模糊变量  $UCV$ .例如:If (CPU is low) and (power is middle) then (device status is low); If (smooth is middle) and (band is big) and (delay is low) then (network status is middle); If (device status is low) and (network status is middle) then (UCV is middle).

步骤 3. 解模糊(defuzzification).由于输出变量  $UCV$  是模糊变量,因此需要对其进行解模糊,以得到精确量化值.本文中使用重心法<sup>[9]</sup>对  $UCV$  进行解模糊,从而得到代表用户上下文信息的综合量化值(实数值).这样,通过上述 3 个步骤得到一个能代表用户上下文信息的实数值  $UCV \in [0,10]$ ,从而为查找或判断用户上下文相似或相同的用户反馈创造了条件.

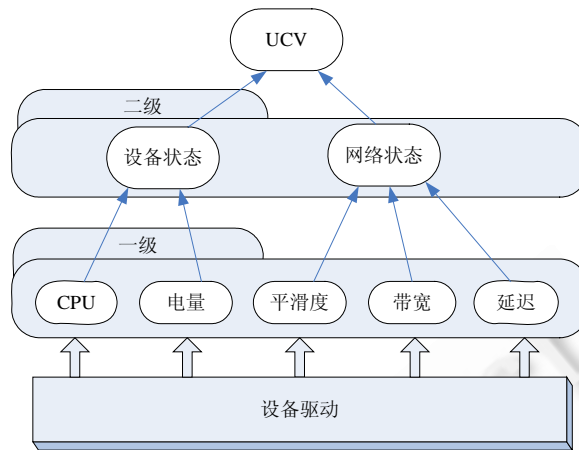


Fig.5 Hierarchy logic of user context inference

图 5 用户上下文推理的层级逻辑关系

本文中采用的 HFS 与常规模糊系统相比,明显地减少了模糊规则数量(如果使用常规模糊系统,则规则数高达  $3^5$ ,远远大于 HFS 的 45 条规则),从而有利于对规则库的管理,减少系统维护开销.另外,尽管给出的定义 4 是用户上下文的狭义定义,但其可以根据实际情况进行扩展(例如增加图 5 中的层级或每层中的元素),因此具有可扩展性.

2.2.2 反馈和谐度计算

本文将用户共同使用过的服务集合视为一个多维空间,用户对这些服务的反馈数据视为该空间中的一个点,则反馈和谐度(feedback coordination,简称 FC)可以通过用户之间的反馈等级和用户上下文信息几何距离来度量,即距离越近的两个用户和谐度越高,否则越低.用户反馈和谐度可由公式(3)获得:

$$FC_{AB}^i = \begin{cases} 1 - \sqrt{\frac{\sum_{j=1}^{|S^i|} (q_A^j - q_B^j)^2 \cdot (UCV_A^j - UCV_B^j)^2}{10^4 \cdot |S^i|}}, & |S^i| \neq 0 \\ 0, & |S^i| = 0 \end{cases} \quad (3)$$

上式中, $i$  代表  $A$  用户与其他用户存在的第  $i$  个服务集合, $FC_{AB}^i$  表示  $A$  用户和  $B$  用户的反馈和谐度, $S^i = \{s_1^i, \dots, s_m^i\}$  为  $A$  和  $B$  两用户以前共同使用过的一个包含  $m$  个服务的集合, $q_A^j (j=1, \dots, m)$  和  $q_B^j$  分别是  $A$  和  $B$  两用户对服务集合中  $s_j^i$  的反馈等级, $UCV_A^j$  和  $UCV_B^j$  分别是  $A$  和  $B$  两用户上下文信息.

**定义 5.** 虚拟反馈不是用户给出的实际反馈,而是指与该用户(假设为  $A$ )存在共同服务集的用户们的反馈等级与该服务集计算出的反馈和谐度,在所用服务集下的反馈和谐度中所占权重的乘积和由下式获得:

$$vq_A = \sum_{i=1}^K \frac{FC_i}{\sum_{i=1}^K FC_i} \hat{q}_i \quad (4)$$

其中, $vq_A$  表示  $A$  用户的虚拟反馈, $\hat{q}_i$  为与  $A$  用户共同使用过的第  $i$  个服务集中的用户给出的反馈数据.

**定义 6.** 反馈冲突度是指虚拟反馈与用户给出的实际反馈的不一致或矛盾程度.假设  $A$  用户给出的反馈数据为  $\bar{q}_A$ ,虚拟反馈为  $vq_A$ ,则用户  $A$  的反馈冲突度  $FI_A$  由下式获得:

$$FI_A = \begin{cases} 1, & \text{if } |\bar{q}_A - vq_A| / vq_A > 1 \\ |\bar{q}_A - vq_A| / vq_A, & \text{otherwise} \end{cases} \quad (5)$$

根据定义 5 和定义 6,本文将虚拟反馈的真实性划分为  $j$  个区间,第 1 个区间范围为  $[0, v_1]$ ,其余期间为  $[v_i, v_{i+1}] (i=1, 2, 3, \dots, j-1, v_j \text{ 可根据具体情况取值})$ .如果  $FI_A \leq v_1$ ,即  $FI_A$  落在第 1 个真实性区间,则表明  $A$  用户的反



反馈数据是客观的,予以保留;否则, $\bar{q}_i$ 被调整为 $vq_A$ ,并根据 $FI_A$ 落在的区间赋予其反馈相应的权重.例如:本文中 $j=5, FI_A$ 落在第2个区间,即 $v_1 < FI_A \leq v_2$ ,则公式(1)中的 $w_i=0.6$ ;若 $FI_A$ 落在第3个区间, $w_i=0.4$ ;第4个区间, $w_i=0.2$ ;最后一个区间, $w_i=0$ .因此,通过考虑用户上下文的反馈和谐度计算,能够起到调整或弱化用户反馈等级,以期达到信誉度评估的公正性和客观性.

另外,该模块对以下情况的恶意反馈具有一定的校正作用:

- 1) 如果与恶意用户A存在共同使用服务集的用户给出的反馈是正常的,此时将导致 $\bar{q}_A$ 与 $vq_A$ 偏离较大,因此可以通过 $w_i=0$ 或其他值进行清洗或校正;
- 2) 如果与恶意用户A存在共同使用服务集的用户给出的反馈也是恶意的,但性质不同,即一方为积极恶意,一方为消极恶意(恶意反馈分类的更多信息见第3.1节),同样将导致 $\bar{q}_A$ 与 $vq_A$ 偏离较大,因此,也可以通过 $w_i=0$ 进行清洗.

尽管反馈校正对偏好反馈以及部分恶意反馈具有校对和调整作用,但当存在共同使用的服务集用户,给出的反馈是性质相同的恶意反馈或大范围、有组织的恶意反馈(如 Sybil 攻击<sup>[28,29]</sup>)时,其校正作用将不能充分发挥.故仍然需要其他方法对恶意反馈进行检测或识别,因此模块3(反馈检测)被提出.

### 2.3 反馈检测

反馈检测意味着对恶意反馈进行检测,避免其恶化或影响信誉度评估的准确性,即检测恶意反馈并对其进行过滤或清洗.其功能实现依赖于CUSUM,具体实现是通过检测抽样间隔内反馈水平(即反馈等级的和)与正常情况下反馈水平的平均阶跃来发现恶意反馈.CUSUM 算法源于时序改变点检测理论,它被广泛地应用于异常检测,可以有效地检测到统计过程中均值的突变情况.有关CUSUM 更多的信息见文献[36].

本文将CUSUM 算法用于信誉度评估中对恶意反馈进行检测,其主要思想则是累积明显比正常运行情况下的平均反馈水平高的值(向上漂移),其优点在于能够以连续方式监测用户反馈,从而达到实时检测目的.由于该检测方法实现简单,无须建立一个较为复杂的用户反馈模型,更适合于开放分布式的Web 环境中.本文中以检测积极恶意反馈为例进行说明,即向上漂移;消极恶意反馈与其类似,即向下漂移,在此忽略.

基于文献[17]中有关用户反馈数据为正态分布的假设,令 $\{y_1, \dots, y_n\}$ 为 $n$ 个反馈数据抽样间隔 $T$ (反馈的数量)内反馈检测序列, $\{y_i\}$ 是一个均值为 $\mu_0$ ,方差为 $\sigma^2$ 的正态分布随机变量.该变量代表第 $i$ 个抽样间隔内反馈水平( $y_i = \sum_{t=1}^T q_t(s_j)$ ),当改变点发生时, $\{y_i\}$ 的均值变为 $\mu_1$ ,方差仍保持不变.经过计算<sup>[36]</sup>,可以得到用户反馈检测序列的累积和函数如下:

$$f_n = \left[ f_{n-1} + \frac{\mu_1 - \mu_0}{\sigma^2} \left( y_n - \frac{\mu_1 + \mu_0}{2} \right) \right]^+ \quad (6)$$

为了更好地适合恶意反馈检测,本文令 $x_n$ 为第 $n$ 个抽样间隔内的反馈水平, $\bar{\mu}_n$ 是统计 $x_n$ 的均值,它由式(7)的指数滑动加权平均得到:

$$\bar{\mu}_n = \lambda \bar{\mu}_{n-1} + (1 - \lambda)x_n \quad (7)$$

其中, $\lambda$ 为指数滑动加权平均系数, $\lambda \in [0, 1]$ .

然后令 $\tilde{x}_n = x_n - \bar{\mu}_{n-1}$ ,此时将CUSUM 用于检测 $\tilde{x}_n$ ,意味着在改变点发生前检测序列的均值将为0,即公式(6)中 $\mu_0=0$ .由于改变点发生后的均值无法获得,本文用 $\alpha \bar{\mu}_n$ ( $\alpha$ 为改变幅度)表示改变点变化后的均值,如图6所示.此时,公式(6)转变为公式(8):

$$f_n = \left[ f_{n-1} + \frac{\alpha \bar{\mu}_{n-1}}{\sigma^2} \left( x_n - \bar{\mu}_{n-1} \frac{\alpha \bar{\mu}_{n-1}}{2} \right) \right]^+ \quad (8)$$

至此, $f_n$ 可以通过下式的判决函数明显地反映出 $x_n$ 的变化情况:

$$d_h(f_n) = \begin{cases} 1, & f_n \geq h \\ 0, & f_n < h \end{cases} \quad (9)$$

其中, $h$ 是恶意反馈检测的门限; $d_h(f_n)$ 代表在第 $n$ 抽样间隔的判决:如果 $f_n \geq h$ ( $f_n$ 越大表明恶意反馈越严重),则该

值为 1,表示存在恶意反馈,丢弃该抽样间隔内的反馈,并发出报警;否则该值为 0,表示为正常用户反馈,可以直接用于信誉度评估,系统运行正常.

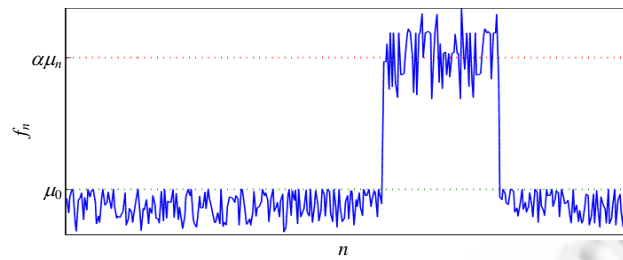


Fig.6 Change of user feedback mean

图 6 用户反馈均值的突变

### 3 仿真实验

为了验证提出的 REAWSS,本文进行了仿真实验,主要包括仿真环境的搭建、与已有研究的仿真对比、对 3 个模块的分析以及对仿真结果的讨论这 4 个小节.其中,仿真环境(第 3.1 节)主要通过 QualNet(<http://www.scalable-networks.com>)仿真开放分布式的 Web 服务环境,仿真结果(第 3.2 节)给出了与文献[1,11,12]中的相应方法在单个服务的信誉度准确度和组合服务信誉度的偏离度上的仿真对比,模块分析(第 3.3 节)对每个模块在信誉度评估中的贡献度以及其相互直接的互助关系进行了验证,结果讨论(第 3.4 节)重点分析了 REAWSS 比其他方法具有较好评估性能的深层原因.

#### 3.1 仿真环境

为了与已有研究进行比较,本文对开放分布式的 Web 环境进行仿真.考虑到该环境的复杂性及实现的困难性,出于简化实验的目的,我们通过 QualNet 仿真无线网络(wireless network)中的 Web 服务环境来替代,如图 7 所示.

图 7 中的 Web 服务器(Web server)运行在一台 PC 机上(Intel Core2 2.8GHz processor,2.0GB of RAM, Windows XP SP3,Java 1.4,MATLAB 7.6,lpsolve5.5),主要用于比较 REAWSS 与其他 3 种方法在信誉度评估方面的性能.仿真服务器(simulation server)运行在另外一台 PC 机上(VC++2005,QualNet4.5),主要通过仿真不同 QoS 属性的 Web 服务、用户及其反馈和上下文.PDAs 和 laptops 为仿真的 500 个用户/节点(node),其中,PDAs(250 个)与 laptops(250 个)相比具有较低的处理能力、内存和电量(延迟和平滑度是随机的),PDAs 和 laptops 共享 802.16 无线接口.

图 7 中的 Services 是 Web 服务器仿真的 500 个不同 QoS 属性的 Web 服务.为了使其更接近真实环境,Web 服务中的 QoS(信誉度除外)采用 QWS 真实数据集,该数据集包含了 2 500 个真实 Web 服务(每个服务包含 9 个 QoS 属性),该数据集更详细的介绍和描述见文献[37].用户上下文(UC)来自于 PDAs 和 laptops 中被模拟的 5 个变量.用户的反馈来自模拟的 500 个用户(一个用户可以使用多个服务并给出反馈)的 3 000 个反馈(包含混乱反馈、偏好反馈和恶意反馈),其中,300 个用户之间及与其他 200 用户之间存在有共同使用服务集(服务集中包含的服务数量在 1~10 之间),模拟的服务用户构造如图 8 所示.

图 8 中 300 个用户中的正常反馈等级服从正态分布<sup>[17]</sup>,取值范围为[1,10].另外,300 个用户中包含 50 个恶意用户,其反馈等级为 1 或 10(反馈等级为 1 表示消极恶意反馈,10 为积极恶意反馈),其余的为混乱反馈和偏好反馈.上述 50 个恶意用户存在 45 个共同使用服务集,其中,两两虚假用户的共同使用服务集(比如恶意用户 A 和 B 同时对多个服务提供了虚假信誉度)数量为 10,其他的 35 个为恶意和正常用户的共同使用服务集.在上述 10 个服务集(恶意+恶意)中,积极的恶意服务集(反馈为 10)和消极的恶意服务集(反馈为 1)数量均为 5.在上述 35 个服务集(恶意+正常)中,积极的恶意服务集和消极的恶意服务集数量分别为 17 和 18.另外,50 个恶意用户中剩余

的 5 个用户不存在共同使用服务集,其中,积极的恶意用户数量为 2,消极的恶意用户数量为 3.

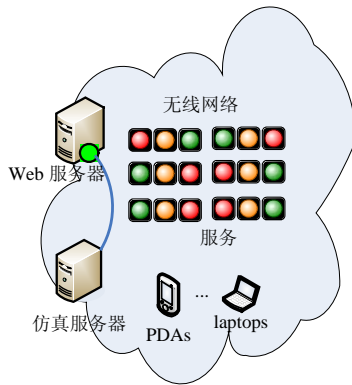


Fig.7 Simulation environment setup

图 7 仿真环境搭建

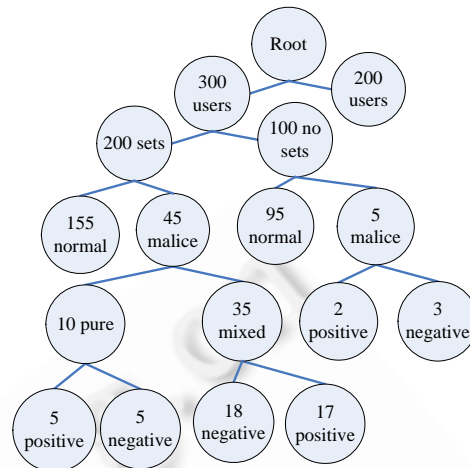


Fig.8 Service customers

图 8 服务用户

本文的仿真建立在一个  $1000 \times 1000m^2$ 、500 节点(随机分布)的无线网络环境中.每次仿真运行持续 1 000 秒,每次数据采集平均超过 15 次仿真运行.另外,REAWSS 中用户反馈档案中的部分数据见表 1.REAWSS 中的其他参数设置为  $\alpha=0.43, \lambda=0.48, h=0.75$ .仿真实验中,所有方法的测试用例和实验环境都是相同的,所有仿真实验对比结果均是运行 20 次后的平均统计.

Table 1 Survey form of feedbacks

表 1 反馈档案

ID	FB	CL	Ratio (%)
1	4.3	1	86.2
2	5.2	3	95.3
...	...	...	...
300	7.5	5	57.6

### 3.2 仿真对比

为了便于说明,ASC,TMS,ASSQ 分别表示文献[1,11,12]中的评估方法.实验中不仅对单个服务,而且还对服务组合下的信誉度评估进行了对比,同时对 REAWSS 中 3 个模块的贡献度进行了分析,并对实验结果进行了讨论.

#### 3.2.1 单个服务

实验中,我们将通过 4 种方法对单个服务的信誉度评估的准确度方面进行比较,即比较信誉度与理想值的差距.信誉度的理想值(ideality)是指对同一个服务,所有用户使用后给出的反馈等级都是相同无差异的情况下计算出的信誉度值.当然,尽管 Ideality 在实际 Web 环境中是不可能的,但可以通过比较与 Ideality 的差距来验证评估方法的优劣,即差距越小,方法准确度越好;反之越差.

实验中,4 种方法分别对 50 个独立的 Web 服务进行评估,其中 3 种反馈类型在所有反馈中的比例以 10% 递增,图 9 给出了它们的评估结果.由于该结果是 50 个独立服务评估结果的平均值,因而评估结果更具公正性.

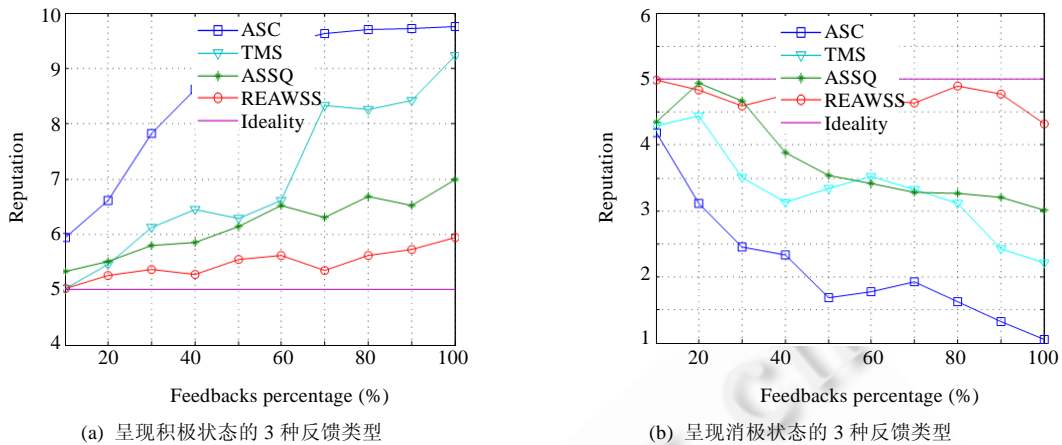


Fig.9 Simulation comparison on reputation evaluation of single service

图 9 针对单个服务的信誉度评估仿真对比

从图 9 可以看出,无论 3 种反馈类型呈现何种状态(积极状态意味着用户的反馈都比理想值高,消极状态则相反),REAWSS 的评估结果与其他方法相比更接近信誉度的理想值,即意味着 REAWSS 在所有对方法中具有最高的准确度.

在图 9(a)中,REAWSS 评估的信誉度平均为 5.471,与理想值很接近(差距仅为 0.471).而 ASC,TMS 和 ASSQ 方法的平均值分别为 8.622,7.015 和 6.165,显然准确度都低于 REAWSS,另外与理想值的差距也较大(最大差距和最小差距分别为 3.622 和 1.165).尤其当 3 种反馈类型数量的比例大于 60%时,上述 3 种方法与理想值差距急剧增大,导致准确度明显降低.而 REAWSS 与理想值的差距虽然也在增大,但幅度较小且较为平稳,因此准确度上的表现更为优秀.在图 9(b)中,REAWSS 与 ASC,TMS 和 ASSQ 的信誉度评估平均值分别为 4.701,2.148,3.331 和 3.757.可以看出,REAWSS 与理想值的差距更小,仅为 0.299,而其他方法的差距均大于 1.这意味着 REAWSS 的准确度在所有方法中仍是最好的.尤其是当 3 种反馈类型数量的比例大于 70%时,上述 3 种方法与理想值的差距也呈现急剧增大趋势,导致准确度下降.而 REAWSS 与理想值的差距变化较小且平稳,显然在信誉度评估的准确性上是最好的.

综合图 9 中的评估结果,REAWSS 与理想值的平均差距仅为 0.385.考虑到实际网络环境,由于该值较小,几乎可以认为该差距是由网络因素造成的而被忽略.因此,针对单个 Web 服务,REAWSS 的评估结果几乎和理想值相同.其原因在于 REAWSS 对上述 3 种反馈类型分别给出了解决模块,从而能够有效克服用户能力缺乏、用户偏好及上下文和恶意反馈等因素对信誉度评估准确度的影响.尤其在针对恶意反馈的检测中,REAWSS 中除模块 3 中的 CUSUM 算法在对恶意反馈的清洗方面发挥了决定性作用外,模块 2 中的反馈和谐度计算减少了恶意反馈对信誉度评估准确度的影响,提高了评估结果的准确度.

### 3.2.2 服务组合

尽管 REAWSS 对单个服务的信誉度评估效果较好,但在 Web 服务选择的实际应用中,更多是考虑用户的 QoS 请求是否得到满足.而本文中 QoS 请求主要指信誉度,即信誉度约束是否与服务组合的信誉度在结果上一致.因此,为了进一步验证 REAWSS 的性能,本文将与其他 3 种方法比较服务组合结果与用户请求在信誉度上的偏离度(deviation).

为便于比较,OUTI 代表了用户的 QoS 请求(信誉度是唯一的约束条件),RUW 代表采用 REAWSS 方法时组合服务的效用函数<sup>[9]</sup>(除信誉度属性外,其余两种属性来自 QWS).类似地,RUX,RUY,和 RUZ 分别表示采用 ASC, TMS 和 ASSQ 方法时组合服务的效用函数.此时,DW,DX,DY 和 DZ 分别表示 RUW,RUX,RUY 和 RUZ 与 OUTI 的偏离度,由式(10)获得:

$$D_i = \frac{|RU_i - OUT_i|}{OUT_i} \times 100\%, i = W, X, Y, Z \quad (10)$$

$D_i$  越小,越接近于 0,表示偏离度越小,即该方法的评估性能越好;否则,表示越差。

图 10 给出了 REAWSS 与其他 3 种方法在不同候选服务和 service 类数量下的仿真对比结果。图 10(a)中 service 类的数量在 3~5 之间随机选择,候选服务数量的取值范围为 10~100。图 10(b)中候选服务的数量在 4~10 之间随机选择,service 类数量的取值范围为 5~50。这 3 种反馈类型所占比例在 10%~80% 之间随机选择。

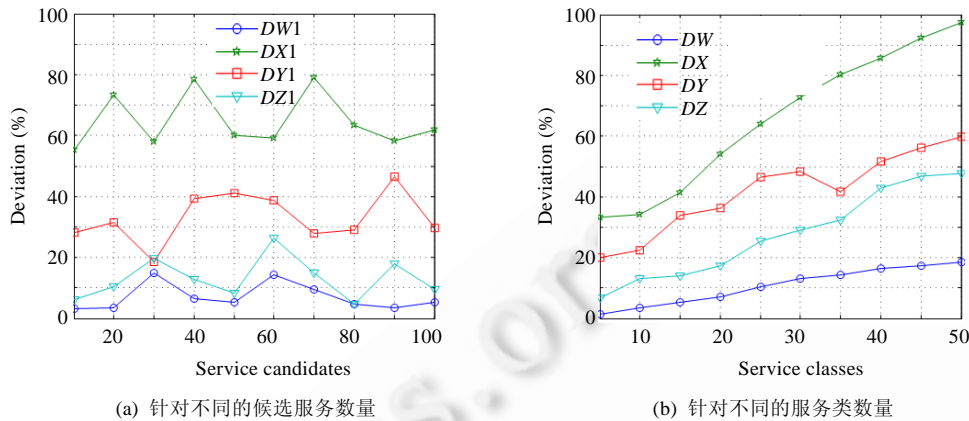


Fig.10 Simulation comparisons on deviation of service composition

图 10 针对服务组合的偏离度的仿真对比

由图 10 可以看出,与其他 3 种方法相比,REAWSS 的偏离度最小,评估性能最好,即服务组合通过使用 REAWSS 方法能够更好地满足用户需求。

在图 10(a)中,REAWSS 的偏离度平均为 10.7%,而 ASC,TMS 和 ASSQ 这 3 种方法的偏离度分别为 65.5%, 41.7%, 27.6%。显然,其他方法的偏离度均大于 REAWSS,意味着服务组合利用上述 3 种方法获得的组合服务与用户的要求偏离较大。在图 10(b)中,REAWSS 的偏离度更小,平均仅为 7.1%(几乎接近于 0),而 ASC,TMS 和 ASSQ 这 3 种方法的偏离度均大于 10%(分别为 64.7%, 33.1% 和 13.1%)。显然,在针对服务组合的信誉度评估,REAWSS 的表现同样优秀。尤其在 service 类数量较少的情况下,组合服务在信誉度上的偏离度几乎都小于 10%。同样,如果考虑到实际的应用环境,REAWSS 几乎不影响用户的服务体验。当然,在 service 类数量较多的情况下,尽管 REAWSS 的偏离度略大,但平均仍然好于其他方法。

在服务组合的对比中,REAWSS 偏离度较低的主要原因是,3 个模块之间具有较好的衔接和互补性,从而最大程度地降低了 3 种反馈类型对信誉度评估的影响,能够准确地评估服务的信誉度,进而为选择出满足用户请求的组合服务提供了可靠保障。例如,虽然模块 1 主要面对相应能力缺乏用户的混乱反馈,但其同样能够为恶意用户反馈提供初级处理,甚至对偏好反馈也起到积极作用。

### 3.3 模块分析

在这个实验中,为了进一步验证 REAWSS 的性能,并为实际部署提供参考,主要分析 3 个模块对 REAWSS 性能的贡献度。贡献度是指 REAWSS 在服务组合的偏离度仿真实验中,在缺少相应模块下的偏离度与 1 的绝对差值,即  $|1 - DW|$ 。表 2 给出了针对不同反馈类型的各个模块的贡献度。

由表 2 可以看出,针对反馈类型中包含的混杂反馈(chaos feedbacks),3 个模块都发挥了作用,其中模块 1 的作用最大,即贡献度最高,达到 62.1%,意味着 100 个混杂反馈中,约 62 个反馈能够被核查到。另外,模块 2 和模块 3 的总体贡献度高达 37.9%。尤其是模块 2,其贡献度达到了 30%,意味着在无法被模块 1 核查的 38 个反馈中,模块 2 能够核查出约 30 个。因此,其余两个模块的使用进一步降低了混杂反馈对信誉度评估可信性的影响,有效地弥补了模块 1 的不足。



Table 2 Contribution rate of modules

表 2 模块贡献度

反馈类型	反馈核查 (%)	反馈校正 (%)	反馈检测 (%)
混乱反馈	62.1	30.6	7.3
偏好反馈	12.4	82.9	4.7
恶意反馈	0.4	15.2	84.4

对于偏好反馈(preference feedbacks),模块 2 的贡献度最高,为 82.9%。因此,在应对 Web 服务选择存在的偏好反馈,模块 2 起到了决定性的作用。另外,模块 1 的贡献度为 12.4%,对模块 2 起到了较好的补充作用。虽然模块 3 的贡献度较低,但仍然能够降低偏好反馈对信誉度评估公正性和客观性的影响。

对于恶意反馈(malicious feedbacks),模块 3 的贡献度最高,达到 84.4%。另外,模块 2 的贡献度超过了 15%。这表明,虽然模块 3 在应对恶意反馈中发挥了决定性作用,但模块 2 的作用不可忽视。当然,尽管模块 1 的贡献度较低,评估作用不明显,但该模块在评估另外两种反馈类型时均具有一定的贡献度,因此,它仍然能够从总体上保障 REAWSS 具有较高评估性能而不可缺失的重要模块。

通过上述分析,REAWSS 中的 3 个模块针对不同反馈类型均发挥了各自的作用,效果明显。同时,3 个模块相互配合,具有较好的互补性,这也是 REAWSS 仿真结果比其他方法好的一个重要原因。当然,尽管上述分析结果来自仿真实验而非实际系统,但其仍然能为该系统的实际部署和应用提供参考。例如,在仅仅面临 Sybil 攻击威胁严重的 Web 服务环境,模块 3 的部署是必不可少的。另外,本文后续的工作是将其应用到实际系统中进一步验证 REAWSS 的各项性能。

### 3.4 结果讨论

讨论的目的是根据仿真结果阐述 REAWSS 的性能比其他较好的深层原因,以便从更客观的角度分析 REAWSS 的优秀性能。

正如第 3.2 节的简单分析,在总体角度上看,其主要原因是 REAWSS 中的 3 个模块被分别设计应对 Web 服务选择中影响信誉度评估准确性的 3 种反馈类型(混乱反馈、偏好反馈和恶意反馈),从而在调用反馈等级时,3 个模块不但针对不同反馈类型扮演了各自角色,而且相互之间还起到了互助作用,因而具有较好的评估性能。而从各个模块的角度来看,更详细的原因如下。

对于 REAWSS,为了减弱能力较差用户的反馈数据在对信誉评估可信性的影响,我们在 REAWSS 中设计了反馈核查模块并对反馈用户建立了反馈档案。以 Web 服务用户 Jack 的反馈档案为例,尽管 Jack 使用了服务,但由于其对 Web 服务缺乏相应知识或认知,其极可能在使用服务后仅仅给出一个随机的反馈等级  $q_{Jack}$ 。由于  $q_{Jack}$  既不代表 Jack 对该服务的全面感知也不包含他的主观因素,甚至与其上下文没有任何关系,显然,  $q_{Jack}$  不能直接用于信誉度评估并为其他用户使用该服务时提供信誉度参考。然而,对于像 Jack 这样的众多反馈,ASC, TMS 和 ASSQ 均没有提出清晰有效的应对方法或策略,而仅仅是笼统地将其以简单、粗粒度的方式划分为普通用户的反馈类型中。结果,由于缺乏针对该反馈类型细粒度的分析以及相应的解决方法,导致其评估的可信性降低。而我们的 REAWSS 中,对 Jack 建立的反馈档案可以根据 Jack 的历史记录以及系统要求的可信水平  $k$  通过公式(5)计算出正确率  $ratio_{Jack}^k$ ,并将其与门限值  $ratio'_k$  进行比较,然后根据比较结果赋予  $w_i$  不同的权重值,最后将根据  $w_i$  计算最终的信誉度。仿真结果显示,超过 60% 的混乱反馈被核查,显然,这比其他 3 种方法更合理、粒度更细。

随着 Jack 使用 Web 服务次数的增加,考虑到人类的学习能力,他的反馈能力也在相应增加。一旦其达到一定水平( $ratio_{Jack}^k > ratio'_k$ )后,其反馈不再属于混乱反馈类型。然而,这时考虑到人类行为特征及环境因素,Jack 的反馈将容易受其个人偏好及使用服务时的上下文信息影响。针对该现象,我们在 REAWSS 设计了反馈校正模块,主要用于解决用户反馈受其自身偏好和上下文影响,造成信誉度评估过多掺杂主观因素和环境因素而导致信誉度公正性和客观性降低。例如,Jack 在使用文献[38]中的多媒体分发服务时,他对编码服务的价格要求不高,而过分看重其响应时间(其他用户或许与其相反)。因此,如果服务的响应时间超出他的期望,Jack 将对该服务给予一个较低的反馈等级。另外,他在 PDA 上使用该服务时给出的反馈等级或许与 laptop 或智能手机不同。所以,在



反馈校正模块中,我们通过用户上下文推理和反馈和谐度计算两个方面综合校正 Jack 反馈等级对信誉度评估的影响.

众所周知,由于 PDA 的各项软硬件运行水平均低于 laptop,导致对于同一服务,使用 laptop 的用户给出的反馈等级普遍高于使用 PDA 的用户.另外,在上述服务运行样式上的上下文信息也不尽相同,对于这种定性上的不同,REAWSS 通过层级模糊推理将用户上下文转化为定量不同.例如,对于 PDA,通过输入 UC 中各个参数值及使用的模糊规则如“If (CPU is low) and (power is low) then (device status is low); If (smooth is middle) and (band is low) and (delay is middle) then (network status is middle); If (device status is low) and (network status is middle) then (UCV is middle)”,经过解模糊可以得到代表用户上下文信息的精确值  $UCV=0.47$ .而对于 laptop,根据 UC 及使用的模糊规则,如“If (CPU is middle) and (power is middle) then (device status is middle); If (smooth is big) and (band is middle) and (delay is low) then (network status is big); If (device status is middle) and (network status is big) then (UCV is big)”,得到的  $UCV=0.85$ .显然,两者上下文信息的推理结果表明它们明显不同(差距较大),而这种不同必然会影响到用户的反馈.因此,如果缺少对上下文因素的考虑,而直接将 Jack 的反馈用于信誉度评估是不合理的.这时,我们提出了反馈和谐度计算.它通过计算与 Jack 使用过共同服务集的用户(如 Rose)的反馈及其上下文信息与 Jack 的距离得到反馈和谐度( $FC_{Jack,Rose}^i$ ).根据获得的  $FC_{Jack,Rose}^i$  计算其虚拟反馈以及后续的反馈冲突度( $FI_{Jack}$ ),然后根据  $FI_{Jack}$  落在的相应区间对 Jack 的反馈分配相应权重.仿真结果显示,超过 82% 的反馈能够被校正.

当然,如果 Jack 是恶意用户,该模块也能根据对其反馈和谐度而得到其反馈冲突度,同样能够根据所落区间对其施加影响(在针对恶意反馈的检测中,该模块贡献度超过 15%).显然,这种互补性是 ASC, TMS 和 ASSQ 方法所不具备的.另外,在 ASC 中缺少考虑用户偏好和上下文对信誉度的影响,而是直接将其用于信誉度评估.在 TMS 中,虽然它可以兼容多种信誉度评估模型,但其反馈采集依赖于  $H=(C,S,Fdbk,Attrs)$ <sup>[11]</sup>,显然缺少对用户上下的考虑,以至于在其服务组合场景(composite scenarios), $H$  直接用于各个评分函数.因此,即使不考虑存在的恶意反馈,仅仅存在的偏好反馈也能使 TMS 信誉度的准确性偏离实际值,从而丧失其公正性和客观性.虽然在 ASSQ 中,它通过服务协商能够有效地提高信誉度评估的公平、客观性,但由于服务协商中缺少上下文信息的支持,导致用户质量期望与服务感知差别较大,导致其对部分偏好反馈束手无策.所以,REAWSS 中对用户偏好和上下文的考虑,显著地提高了信誉度评估的公平性和客观性,明显地比其他方法更具优势.

综上所述,如果存在众多像 Jack 这样的恶意反馈,并且如果公式(6)中 Rose 也是性质相同的恶意反馈,那么 REAWSS 中模块 2 的效能将大大降低,甚至失去作用.针对该问题,我们设计了反馈检测模块,它主要用于检测 Web 服务环境中充斥的恶意反馈或 Sybil 攻击.例如,Jack 通过伪造或盗用等手段同时获得多个服务用户身份,而对某个服务发起 Sybil 攻击以期抬高或打压该服务的信誉度,从而获得非法利益或破坏公平竞争.在攻击中,以 Jack 抬高该服务的信誉度为例,他将模拟众多用户给予该服务大量积极恶意反馈,而这些反馈在一定抽样间隔内的反馈水平将明显高于正常水平.此时,CUSUM 通过累积抽样间隔内反馈水平与正常水平的偏离( $f_n$ ),并将其与设定的阈值  $h$  进行比较.一旦  $f_n \geq h$ ,则意味着发生恶意反馈从而将其丢弃;否则,给予保留并用于信誉度评估.仿真结果显示,其贡献度达到 84.4%,意味着超过 80% 的恶意反馈能够被检测到.而在 ASC, TMS 和 ASSQ 等 3 种方法中,ASC 缺少对恶意反馈的检测.TMS 虽然提到了支持恶意反馈,但由于缺少应对偏好反馈和混乱反馈的解决方法,仍然不能从全局上提高信誉度评估性能.ASSQ 中的服务协商虽然能够对部分恶意反馈起到作用,但由于用户本身的恶意性,导致其无法回避 Sybil 的间接攻击,因此与 ASC 一样,无法对充斥恶意反馈的 Web 服务环境中已注册服务的信誉度给予准确评估.

#### 4 结论及未来工作展望

为了克服开放分布式 Web 服务环境存在的混乱、偏好和恶意这 3 种反馈类型对信誉度评估准确性的影响,本文提出了一种用于 QoS 感知 Web 服务选择中信誉度评估的方法,即 REAWSS.该方法通过使用反馈核查,反馈校正和反馈检测 3 个模块应对来自用户的 3 种反馈对信誉度评估客观性和准确性的影响.REAWSS 具有以下

显著特征:1) 建立的用户档案有效地降低了混乱反馈类型对信誉度评估可信性的影响;2) 采用的层级模糊系统使其能够通过模糊推理对用户上下文进行定性到定量的信息转换;3) 根据上下文信息,对反馈和谐度和反馈冲突度的计算提高了信誉度评估的准确性,极大地降低了用户偏好和上下文信息对信誉度评估的公正性和客观性的影响;4) CUSUM 被用于检测恶意反馈,不仅能够有效地检测积极恶意反馈,同样也支持对消极恶意反馈的检测,从而保证了信誉度评估的准确性,促进了公平竞争;5) 3个模块在履行自身义务的同时,模块之间还具有较好的链接和互补性,从而在总体上进一步提高了 REAWSS 的评估性能.REAWSS 的有效性已经在搭建的仿真环境中得到了验证.仿真结果显示,REAWSS 不仅能够提高单个服务信誉度评估的准确度,而且对服务组合同样具有优秀的表现.另外,组合服务结果与用户请求在信誉度偏离度上的降低显著地提高了基于 QoS 感知的 Web 服务选择进程.

到目前为止,我们还没有能力在实际环境中对提出的 REAWSS 方法进行部署和应用,以确认其在仿真环境之外的评估性能.尤其是 Web 服务基于 Internet,由于网络的不稳定因素,文中用户上下文中 5 种因素的取值势必也会呈现出不可预期的变化,而对于这些因素再进行模糊化处理并通过推理进行反馈校正,其设计的模糊规则可能导致推理出的用户上下文环境与实际环境存在偏差.另外,提出的方法中参数设置过多,缺少一种自适应的机制赋予每个参数合适的值,导致其自适应能力较差.因此,我们希望在今后,以当前本文研究为基础构建一个实际环境下的实验平台,以增加实验结果的准确性.其目标是在不久的将来为实际环境中的部署和商业应用提供参考和借鉴,以便通过提高 Web 服务选择中的信誉度评估的准确性来帮助服务用户寻找合适、可靠的服务.另外,通过分析文中所提仿真环境的非现实性,对比其与实际环境的差异,增加考虑网络不稳定等不确定性因素,研究如何在实际环境中更加准确地评估 REAWSS 方法,从而实现 REAWSS 的进一步实际应用.

**致谢** 在此,对提供 QWS 数据集的加拿大圭尔夫大学的 Eyhab Al-Masri 和 Qusay H. Mahmoud 博士表示感谢.

#### References:

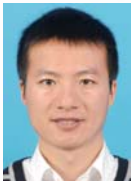
- [1] Ardagna D, Pernici B. Adaptive service composition in flexible processes. *IEEE Trans. on Software Engineering*, 2007,33(6): 369–384. [doi: 10.1109/TSE.2007.10111]
- [2] Kritikos K, Plexousakis D. Requirements for QoS-based Web service description and discovery. *IEEE Trans. on Service Computing*, 2009,2(4):320–337. [doi: 10.1109/TSC.2009.26]
- [3] Wang SG, Sun QB, Yang FC. Towards Web service selection based on QoS estimation. *Int'l Journal of Web and Grid Services*, 2010,6(4):424–443. [doi: 10.1504/IJWGS.2010.036406]
- [4] Zhang CW, Su S, Chen JL. Genetic algorithm on Web services selection supporting QoS. *Chinese Journal of Computers*, 2006, 29(7):1029–1037 (in Chinese with English abstract).
- [5] Hu CH, Wu M, Liu GP, Xu DZ. Approach to constructing Web service workflow based on business spanning graph. *Journal of Software*, 2007,18(8):1870–1882 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/18/1870.htm> [doi: 10.1360/jos181870]
- [6] Jiang ZY, Han JH, Wang Z. Optimization model for dynamic QoS-aware Web services selection and composition. *Chinese Journal of Computers*, 2009,32(5):1014–1025 (in Chinese with English abstract). [doi: 10.3724/SP.J.1016.2009.01014]
- [7] Liu X, Huang G, Mei H. Discovering homogeneous Web service community in the user-centric Web environment. *IEEE Trans. on Services Computing*, 2009,2(2):167–181. [doi: 10.1109/TSC.2009.11]
- [8] Yu T, Zhang Y, Lin KJ. Efficient algorithms for Web services selection with end-to-end QoS constraints. *ACM Trans. on the Web*, 2007,1(1):1–26. [doi: 10.1145/1232722.1232728]
- [9] Li Z, Yang FC, Su S. Fuzzy multi-attribute decision making-based algorithm for semantic web service composition. *Journal of Software*, 2009,20(3):583–596 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/20/583.htm> [doi: 10.3724/SP.J.1001.2009.03341]
- [10] Alrifai M, Dolog P, Balke WT, Nejdl W. Distributed management of concurrent Web service transactions. *IEEE Trans. on Services Computing*, 2009,2(4):289–302. [doi: 10.1109/TSC.2009.29]

- [11] Conner W, Iyengar A, Mikalsen T, Rouvellou I, Nahrstedt K. A trust management framework for service-oriented environments. In: Proc. of the 18th Int'l Conf. on World Wide Web (WWW 2009). 2009. 289–302. [doi: 10.1145/1526709.1526829]
- [12] Limam N, Boutaba R. Assessing software service quality and trustworthiness at selection time. IEEE Trans. on Software Engineering, 2010,36(4):559–574. [doi: 10.1109/TSE.2010.2]
- [13] Wang Y, Vassileva J. A review on trust and reputation for Web service selection. In: Proc. of the 27th Int'l Conf. on Distributed Computing Systems Workshops (ICDCSW 2007). 2007. 25–32. [doi: 10.1109/ICDCSW.2007.16]
- [14] Bianculli D, Binder W, Drago L, Ghezzi C. Transparent reputation management for composite Web services. In: Proc. of the 2008 IEEE Int'l Conf. on Web Services (ICWS 2008). 2008. 621–628. [doi: 10.1109/ICWS.2008.39]
- [15] Xu ZQ, Martin P, Powley W, Zulkernine F. Reputation-Enhanced QoS-based Web services discovery. In: Proc. of the 2007 IEEE Int'l Conf. on Web Services (ICWS 2007). 2007. 249–256. [doi: 10.1109/ICWS.2007.152]
- [16] Malik Z, Bouguettaya A. Reputation bootstrapping for trust establishment among Web services. IEEE Internet Computing, 2009, 13(1):40–47. [doi: 10.1109/MIC.2009.17]
- [17] Li Y, Zhou MH, Li RC, Cao DG, Mei H. Service selection approach considering the trustworthiness of QoS data. Journal of Software, 2008,19(10):2620–2627 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/19/2620.htm> [doi: 10.3724/SP.J.1001.2008.02620]
- [18] Malik Z, Bouguettaya A. RATEWeb: Reputation assessment for trust establishment among Web services. Journal of VLDB, 2009, 18(4):885–911. [doi: 10.1007/s00778-009-0138-1]
- [19] Li HH, Du XY, Tian X. A review-based reputation evaluation approach for Web services. Journal of Computer Science and Technology, 2009,24(5):893–900. [doi: 10.1007/s11390-009-9280-x]
- [20] Yahyaoui H, Mamar Z, Bentahar J, Sahli N, Elnaffar S, Thiran P. On the reputation of communities of Web services. In: Proc. of the 8th Int'l Conf. on New Technologies in Distributed Systems (NOTERE 2008). 2008. 1–8. [doi: 10.1145/1416729.1416735]
- [21] Lages AG, Pirmez L, Pires PF, Delicato VC. Satya: A reputation-based approach for service discovery and selection in service oriented architectures. In: Proc. of the 9th Annual ACM Int'l Workshop on Web Information and Data Management (Widm 2007). 2007. 153–60. [doi: 10.1145/1316902.1316927]
- [22] Jung K, Lee Y. Autonomic trust extraction for trustworthy service discovery in urban computing. In: Proc. of the 8th IEEE Int'l Conf. on Dependable, Autonomic and Secure Computing (DASC 2009). 2009. 502–507. [doi: 10.1109/DASC.2009.39]
- [23] Cai SB, Zou YZ, Xie B, Shao WZ. Mining the Web of trust for Web services selection. In: Proc. of the IEEE Int'l Conf. on Web Services (ICWS 2008). 2008. 809–810. [doi: 10.1109/ICWS.2008.114]
- [24] Sherchan W, Krishnaswamy S, Loke SW. Explaining reputation for informed Web services selection. In: Proc. of the IEEE Int'l Conf. on Web Services (ICWS 2008). 2008. 805–806. [doi: 10.1109/ICWS.2008.94]
- [25] Bianculli D, Binder W, Drago ML, Ghezzi C. ReMan: A pro-active reputation management infrastructure for composite Web services. In: Proc. of the 31th IEEE Int'l Conf. on Software Engineering (ICSE 2009). 2009. 623–626. [doi: 10.1109/ICSE.2009.5070571]
- [26] Zou YZ, Gu L, Li G, Xie B, Mei H. Rectifying prejudicial feedback ratings in reputation based trust management. In: Proc. of the IEEE Int'l Conf. on Services Computing (SCC 2007). 2007. 530–535. [doi: 10.1109/SCC.2007.91]
- [27] Li X, Ling L. PeerTrust: Supporting reputation-based trust for peer-to-peer electronic communities. IEEE Trans. on Knowledge and Data Engineering, 2004,16(7):843–857. [doi: 10.1109/TKDE.2004.1318566]
- [28] Ries S, Aitenbichler E. Limiting sybil attacks on Bayesian trust models in open SOA environments. In: Proc. of the 2009 Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing (UIC 2009 and ATC 2009). 2009. 178–183. [doi: 10.1109/UIC-ATC.2009.82]
- [29] Zhang YC, Fang YG. A fine-grained reputation system for reliable service selection in peer-to-peer networks. IEEE Trans. on Parallel and Distributed Systems, 2007,18(8):1134–1145. [doi: 10.1109/tpds.2007.1043]
- [30] Xu LF, Hu HF, Sang ZX, Xu FM, Zou DQ. A prestige reporting mechanism based on gray system theory. Journal of Software, 2007,18(7):1730–1737 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/18/1730.htm> [doi: 10.1360/jos181730]
- [31] Li F, Yang F, Shuang K, Su S. A policy-driven distributed framework for monitoring quality of Web services. In: Proc. of the IEEE Int'l Conf. on Web Services (ICWS 2008). 2008. 708–715. [doi: 10.1109/ICWS.2008.123]

- [32] Li F, Yang F, Shuang K, Su S. Q-Peer: A decentralized QoS registry architecture for Web services. In: Proc. of the 5th Int'l Conf. on Service-Oriented Computing (ICSOC 2007). 2007. 145–156. [doi: 10.1007/978-3-540-74974-5\_12]
- [33] Su S, Li F, Yang FC. Iterative selection algorithm for service composition in distributed environments. Science in China (Series F—Information Sciences), 2008,51(11):1841–1856 (in Chinese with English abstract). [doi: 10.1007/s11432-008-0147-5]
- [34] Khosravifar B, Bentahar J, Thiran P, Moazin A, Guiot A. An approach to incentive-based reputation for communities of Web services. In: Proc. of the IEEE Int'l Conf. on Web Services (ICWS 2009). 2009. 303–310. [doi: 10.1109/ICWS.2009.101]
- [35] Joo MG, Lee JS. A class of hierarchical fuzzy systems with constraints on the fuzzy rules. IEEE Trans. on Fuzzy Systems, 2005, 13(2):194–203. [doi: 10.1109/tfuzz.2004.840096]
- [36] Basseville M, Nikiforov IV. Detection of Abrupt Changes: Theory and Application. Englewood Cliffs: Prentice-Hall, 1993.
- [37] Al-Masri E, Mahmoud QH. Investigating Web services on the World Wide Web. In: Proc. of the 17th Int'l Conf. on World Wide Web (WWW 2008). 2008. 795–804. [doi: 10.1145/1367497.1367605]
- [38] Wagner M, Kellerer W. Web services selection for distributed composition of multimedia content. In: Proc. of the 12th ACM Int'l Conf. on Multimedia (ACM Multimedia 2004). 2004. 104–107. [doi: 10.1145/1027527.1027546]

#### 附中文参考文献:

- [4] 张成文,苏森,陈俊亮.基于遗传算法的 QoS 感知的 Web 服务选择.计算机学报,2006,29(7):1029–1037.
- [5] 胡春华,吴敏,刘国平,徐德智.一种基于业务生成图的 Web 服务工作流构造方法.软件学报,2007,18(8):1870–1882. <http://www.jos.org.cn/1000-9825/18/1870.htm> [doi: 10.1360/jos181870]
- [6] 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.计算机学报,2009,32(5):1014–1025. [doi: 10.3724/SP.J.1016.2009.01014]
- [9] 李祯,杨放春,苏森.基于模糊多属性决策理论的语义 Web 服务组合算法.软件学报,2009,20(3):583–596. <http://www.jos.org.cn/1000-9825/20/583.htm> [doi: 10.3724/SP.J.1001.2009.03341]
- [17] 李研,周明辉,李瑞超,曹东刚,梅宏.一种考虑 QoS 数据可信性的服务选择方法.软件学报,2008,19(10):2620–2627. <http://www.jos.org.cn/1000-9825/19/2620.htm> [doi: 10.3724/SP.J.1001.2008.02620]
- [30] 徐兰芳,胡怀飞,桑子夏,徐凤鸣,邹德清.基于灰色系统理论的信誉报告机制.软件学报,2007,18(7):1730–1737. <http://www.jos.org.cn/1000-9825/18/1730.htm> [doi: 10.1360/jos181730]
- [33] 苏森,李飞,杨放春.分布式环境中服务组合的迭代选择算法.中国科学(F 辑:信息科学),2008,38(10):1717–1732. [doi: 10.1007/s11432-008-0147-5]



王尚广(1982—),男,河南周口人,博士生,CCF 会员,主要研究领域为服务计算,网络安全.



杨放春(1957—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为服务计算,网络智能化,网络安全,通信软件.



孙其博(1975—),男,博士,副教授,CCF 会员,主要研究领域为网络智能化,下一代网络安全.