

基于 QoS 综合评价的服务选择方法

何干志,刘茜萍

(南京邮电大学 计算机学院,江苏 南京 210003)

摘要:从海量服务中准确选择出符合用户真实需求的服务变得日益重要,而服务质量的评定在选择服务的过程中至关重要。基于 QoS 的已有服务选择方法,通常假定服务提供者 and 使用者给出的 QoS 数据都是真实可信的,但这一假设实际上很难得到保证。为此,提出了基于 QoS 综合评价的服务选择方法。该方法针对来自服务提供者的质量属性,基于历史数据的统计对当前 QoS 数据进行修正;而针对来自服务使用者的质量属性,基于具有时间系数的历史评价计算推荐用户和目标用户之间的相似程度,以权衡其历史评价的可靠程度,进而得出针对该目标用户的最终 QoS 参考评价。同时,采用熵值法计算用户的主客观合成权重,并基于提供者修正 QoS 发布信息和用户可靠的 QoS 参考评价信息完成服务选择。实例验证结果表明,所提出的方法有效可行。

关键词:服务选择;用户相似度;熵值法;QoS 综合评价

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2017)08-0164-07

doi:10.3969/j.issn.1673-629X.2017.08.035

A Service Selection Method with QoS Synthetic Evaluation

HE Gan-zhi, LIU Xi-ping

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

Abstract: It is increasingly important for users to exactly select a service meeting their requirements from the mass services. Assessment on the quality of service is crucial in the process of service selection. Existed QoS-based service selection methods generally assume that QoS data from service providers and users are trustable. However, it is not realistic. Therefore, a service selection method based on comprehensive evaluation of QoS is proposed, by which the values of quality attributes provided have been adjusted based on historical statistical data and that from user reviews have been integrated into a comprehensive QoS evaluation for a target user based on the user similarity computed through reviews on time factors. Moreover, entropy method is adopted to calculate the user's subjective and objective synthesis weights with further operation of service selection based on the adjusted publishing QoS from providers and the reliable reviewing QoS from users. The experimental results show that it is effective and feasible.

Key words: service selection; user similarity; entropy method; QoS synthetic evaluation

1 概述

网络环境中服务技术的出现为组织间建立一种灵活多样的协作关系创造了前所未有的机会^[1]。目前,随着服务的发展和广泛应用,服务的大量出现,提供相同功能的服务越来越多,如何准确从庞大的服务集中选择最符合要求的服务是服务技术面临的巨大挑战。因此,基于服务质量对服务进行选择已成为热点问题之一^[2-3]。

目前提出的不少基于 QoS 的服务选择方法大多假定服务提供者描述的 QoS 数据和用户反馈的 QoS

评价都是真实可靠的,因而在对服务进行排序和选择时,常常直接使用服务提供者发布的 QoS 数据,同时采用平均值的方法处理服务使用者反馈的 QoS 数据^[4-5]。然而在实际应用中,对于 QoS 数据可靠性的假设往往很难保证:一方面,服务提供者可能出于利益考虑,发布高于服务实际水平的 QoS 数据以吸引更多的用户^[6-7];另一方面,用户反馈的 QoS 评价,常常受到自身因素的影响,甚至是一些恶意的虚假数据,而平均值的方法缺少能够度量这些质量属性数据有效性的标准。这些不可靠的 QoS 数据将直接影响 QoS 计算

收稿日期:2016-09-30

修回日期:2017-01-04

网络出版时间:2017-07-05

基金项目:国家自然科学基金资助项目(61402241)

作者简介:何干志(1992-),男,硕士研究生,研究方向为服务计算;刘茜萍,副教授,硕士生导师,研究方向为服务计算、工作流。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20170705.1652.074.html>

的准确性和可靠性,以及最终的服务选择结果^[8-10]。

随着数据可靠性问题的日益突出,一些学者开始渐渐关注这类问题的研究。文献[11]根据用户反馈的 QoS 数据,从而得到可靠的 QoS 评价信息,当服务提供者描述不符合事实的 QoS 数据,该提供者这种不诚实的行为将受到处罚,但是该方法却没有提出服务使用者 QoS 可靠性问题的计算方法,忽略了其在服务选择过程中的作用;文献[12]同时考虑了来自服务使用者和服务提供者两方面的 QoS 属性,同时通过不同的计算方法来修正 QoS 属性值,从而得到可靠的 QoS 质量属性值参考信息,但是该方法却忽略了 QoS 质量属性值的动态变化,却将其视为了常量;文献[13]定性分析了直接经验和间接经验在服务可信性判断中的作用,考虑了服务的时效性,并提出了 QoS 可靠性的评估函数,但片面强调了直接经验的主导作用,忽略了当交往次数很少时间接经验的重要性;文献[14]提出了一种基于历史经验和可信概率的贝叶斯模型,给出了奖励函数,但偏重于对用户的未来行为进行约束。

然而,目前基于可靠 QoS 的服务选择研究大多仅根据服务提供者描述或用户反馈信息开展,难以为目标用户提供全面可靠的 QoS 选择依据。为此,提出了一种基于 QoS 综合评价的服务选择方法。对于来自服务提供者描述的 QoS,基于历史数据的统计对服务提供者当前描述的 QoS 数据加以修正,对于来自用户反馈的 QoS 评价,基于用户相似度计算得到各候选服务的可靠综合 QoS 评价。此外,还采用了熵值法以计算得到用户的主客观合成权重。基于提供者的修正 QoS 发布信息 and 用户的可靠综合 QoS 评价信息完成符合用户权重的服务选择。

2 基于历史数据的服务提供者 QoS 信息修正

对于来自服务提供者描述的质量属性数据,如服务的执行时间、价格等,这类属性通常在服务注册时会随着功能属性发布到注册中心,但一些服务提供者为了吸引服务使用者会发布高于实际性能的数值,所以,在进行服务选择时如果直接使用服务提供者发布的数据是不合适的。对于这方面数据的修正,主要通过在中加入采集机制的方式收集服务提供者描述的历史数据。该机制基于历史数据的统计定期对当前描述的 QoS 数据进行修正,此类质量属性的修正是由提供者当前描述的数据和统计的历史数据加权决定,并引入时间因子,从而保证 QoS 数据的可靠性。下面给出该过程的若干相关定义。

定义 1 ($S, \mathbf{R}, \mathbf{PT}$): S 表示所有用户使用过的服务集合,有 $S = \{s_1, s_2, \dots, s_m\}$; \mathbf{R} 表示服务描述的矩阵,

有 $R = \{R_{i,j} \mid R_{i,j}$ 表示服务 s_i 第 j 次发布的 QoS 数据, $R_{i,j} = \langle r_{i,j,1}, r_{i,j,2}, \dots, r_{i,j,k}, \dots, r_{i,j,n_i} \rangle$, 其中 $i = 1, 2, \dots, m$, $k = 1, 2, \dots, n_i$, $j = 0, 1, \dots, v$, $r_{i,j,k}$ 表示服务 s_i 第 j 次发布的 QoS 数据中属性 k 的描述; \mathbf{PT} 表示服务发布 QoS 数据时间的矩阵,有 $\mathbf{PT} = \{pt_{i,j} \mid pt_{i,j}$ 表示服务 s_i 第 j 次发布 QoS 的数据时间 ($i = 1, 2, \dots, m$, $j = 0, 1, \dots, v$) $\}$ 。

针对服务提供者当前发布的数据,使用式(1)对其进行修正。

$$q = wR_{i,v} + (1 - w)f(R_{i,0}, R_{i,1}, \dots, R_{i,v-1}) \quad (1)$$

其中, $R_{i,0}, R_{i,1}, \dots, R_{i,v}$ 是从服务提供者发布了多次数据后采集到的 $v+1$ 次相关质量属性向量值; w ($0 < w < 1$) 用于调整服务提供者发布的质量属性值和采样获取值在修正结果中的比值; $f(R_{i,0}, R_{i,1}, \dots, R_{i,v-1})$ 是个函数,可基于采集到的 $R_{i,0}, R_{i,1}, \dots, R_{i,v-1}$ 计算综合历史的 QoS 向量值; q 为服务 s_i 修正的 QoS 值向量; $R_{i,v}$ 为当前正发布的质量属性向量。

另外需要说明地是,当服务无法收集历史服务质量属性数据时,则只有 $R_{i,v}$ 这一当前发布信息,此时可设 $w = 1$,即服务提供者当前发布的数据为最终该质量属性值,具体计算方法为:

$$q = R_{i,v} \quad (2)$$

假设某服务 s_i 发布了多次数据后,从中采集 $v+1$ 次相关质量属性向量值,分别为 $R_{i,0}, R_{i,1}, \dots, R_{i,j}, \dots, R_{i,v}$ (按时间先后),其对应的时间分别是 $pt_{i,0}, pt_{i,1}, \dots, pt_{i,j}, \dots, pt_{i,v}$, 其中 $pt_{i,0}$ 为服务 s_i 最初发布该服务的时间,故 $pt_{i,j} - pt_{i,0}$ 为第 j 次服务与最初发布服务时间间隔。一般而言,时间越近值越可靠,所以发布服务质量属性的时间距离最初发布服务的时间越长,所占的权重就越大。具体计算方法为:

$$f(R_{i,0}, R_{i,1}, \dots, R_{i,v-1}) = \sum_{j=0}^{v-1} \frac{pt_{i,j} - pt_{i,0}}{\sum_{j=0}^{v-1} (pt_{i,j} - pt_{i,0})} R_{i,j} \quad (3)$$

3 可靠的用户评价获取

对于来自服务使用者反馈的服务质量属性,如满意度,由于受到服务使用者所处环境、主观想法等各种因素的影响,对于同一种服务而言,不同的使用者在使用过该服务后都可能会有截然不同的评价结果,而且不排除恶意诋毁的情形,将所有评价平等对待是不合适的。因此在 QoS 计算中,对所有的用户评价取平均是不合适的。为此,采用了“用户相似度”作为衡量不同用户评价所占的权重大小,以它们的加权平均求得最终评价。所谓“用户相似度”是指两个用户对于使用过相同服务所反馈评价的相似程度,即相似程度越

大,其用户评价在计算中所占的权重就越大。下面给出该过程的若干相关定义。

定义 2(U, \mathbf{EV}): U 表示一组使用过相关服务的用户构成的集合,有 $U = \{u_1, u_2, \dots, u_l\}$; \mathbf{EV} 表示用户对服务的评价矩阵,有 $\mathbf{EV} = \{EV_i \mid EV_i \text{ 是用户评价集合, 为一个四元组, 有 } EV_i = (E_i, T_i, C_i, ST_i)\}$ 。其中, E_i 是用户 u_i 所访问服务的有限集合,有 $E_i = \{s_j \mid s_j \text{ 为 } u_i \text{ 用过的服务, } s_j \in S\}$; T_i 表示用户 u_i 使用的服务次数矩阵,有 $T_i = \{t_{i,j} \mid t_{i,j} \text{ 表示用户 } u_i \text{ 使用服务 } s_j \text{ 的次数, } i = 1, 2, \dots, l, j = 1, 2, \dots, m\}$; C_i 是用户 u_i 对所访问服务评价的有限集合,有 $C_i = \{C_{i,j,p} \mid C_{i,j,p} \text{ 为用户 } u_i \text{ 对服务 } s_j \text{ 第 } p \text{ 次的评价向量, 而 } C_{i,j,p} = \langle c_{i,j,p,1}, c_{i,j,p,2}, \dots, c_{i,j,p,k}, \dots, c_{i,j,p,n_2} \rangle, \text{ 其中 } i = 1, 2, \dots, l, j = 1, 2, \dots, m, p = 1, 2, \dots, t_{i,j}, k = 1, 2, \dots, n_2, c_{i,j,p,k} \text{ 表示用户 } u_i \text{ 对服务 } s_j \text{ 中属性 } k \text{ 第 } p \text{ 次的评价值}\}$; ST_i 表示用户 u_i 使用服务的时长矩阵,有 $ST_i = \{st_{i,j,p} \mid st_{i,j,p} \text{ 表示服务 } s_j \text{ 第 } p \text{ 次使用服务 } s_j \text{ 时长, 其中 } i = 1, 2, \dots, l, j = 1, 2, \dots, m, p = 1, 2, \dots, t_{i,j}\}$ 。

定义 3(推荐用户): 用户 u_i 和 u_j 的评价分别是 $V_j = (E_j, T_j, C_j, ST_j)$ 和 $V_i = (E_i, T_i, C_i, ST_i)$, 也就是如果 $E_i \cap E_j \neq \emptyset$, 且存在 $S_j \in E_i, S_j \notin E_i$, 同时用户 u_i 希望获取服务 S_j 的评价信息, 则称用户 u_i 为 u_j 关于服务 S_j 的推荐用户。

尽管目前有很多对用户相似度计算的相关研究, 但是均未考虑到多次使用服务和使用时间长短对评价的影响。比如用户住酒店的次数越多或者住的时间越长, 则对酒店了解程度就越高, 所以该评价的可靠性就越高。因此, 在计算用户相似度时, 引入了用户领域相关度和用户时间系数进行修正。所谓的用户领域相关度是指推荐用户对服务领域的了解程度高低, 一般情况下, 推荐用户使用某服务的次数占该服务总使用次数的比重越大, 说明其对服务了解程度越高, 那么推荐用户评价的权重系数越高。为此, 使用式(4)对其进行计算:

$$\alpha = \frac{t_{i,j}}{\sum_{i=1}^l t_{i,j}} \quad (4)$$

其中, α 为推荐用户 u_i 在服务 s_i 的领域相关度; l 为用户总数。

所谓的用户时间系数是指用户对服务性能的熟知程度。一般情况下, 用户使用该服务的平均时长占有用户使用该服务平均时长之和的比重越大, 说明其对服务了解程度越高, 那么推荐用户评价的权重系数越高; 同理相反。为此, 使用式(5)对其进行计算:

$$\eta = \frac{ct_{i,j}}{\sum_{i,j>0} ct_{i,j}} \quad (5)$$

其中, η 为用户 u_i 在服务 s_j 的时间系数; $ct_{i,j}$ 为用户 u_i 使用服务 s_j 的平均历史时长, 即 $ct_{i,j} = \frac{\sum_{p=1}^{t_{i,j}} st_{i,j,p}}{t_{i,j}}$ 。

目前为止, 计算服务相似度的公式有很多, 例如皮尔逊相关系数、余弦相似度等。为了使服务的推荐方法更加准确, 从而需要改进目前已有的相似度计算方法。因此引入了服务的个性化特征这一概念。例如, 假设用户 a 和用户 b 以前都使用过服务 1 和服务 2, 给定的服务 1 和服务 2 分别为用户不敏感服务和用户敏感服务, 不同用户在使用过服务 1 后, 给出的 QoS 评价大体相同, 因此所对应的方差比较小, 通过目前已有的计算公式计算出用户 a 和用户 b 的相似程度偏高, 然而这并不能代表真实的相似程度, 因为服务 1 对于所有的用户而言是没有差别的, 这样的服务 1 用于用户 a 和用户 b 的相似度计算时应该贡献程度越低。相反, 不同用户在使用过服务 2 后, 给出的 QoS 评价所对应方差比较大, 表明服务 2 对不同的用户来说所产生的评价差别很大, 此时, 如果两个用户在使用过服务 2 后给出的 QoS 评价相似, 这两个用户才能算是真正意义上的相似。这样的服务 2 用于用户 a 和用户 b 的相似度计算应该贡献程度越高。

总而言之, 在计算两个用户的相似度时, 不能忽略这两个用户共同调用的服务个性化特征。为此, 使用式(6)进行计算:

$$\delta_{j,k} = \sqrt{\frac{1}{\varphi} \sum_{i=1}^{\varphi} (\overline{c_{i,j,k}} - \bar{c})^2} \quad (6)$$

其中, $\overline{c_{i,j,k}}$ 为用户 u_i 对服务 s_j 中某一属性 k 的平

均评价, 即 $\overline{c_{i,j,k}} = \frac{\sum_{p=1}^{t_{i,j}} c_{i,j,p,k}}{t_{i,j}}$; \bar{c} 为所有使用过服务 s_j 的

用户对属性 k 的平均评价, 即 $\bar{c} = \frac{\sum_{i=1}^{\varphi} \overline{c_{i,j,k}}}{\varphi}$; $\delta_{j,k}$ 为服务 s_j 在属性 k 上评分的标准差; φ 为使用过服务 s_j 的用户个数。

因此, 使用结合服务个性化特征的距离向量法来计算目标用户服务评价和参考用户服务评价的偏离大小, 服务的 QoS 属性可以使用向量形式表达, 将用户共同使用过的服务集合视为一个多维空间, 用户对这些服务反馈的质量属性数据视为该空间的一个点。则两用户之间的反馈相似度可以通过它们之间的几何距离来衡量, 距离越小, 说明目标用户和参考用户反馈的服务评价相似度越大, 则该用户的评价信息所占的权重值越大。距离越大, 说明目标用户和参考用户反馈的服务评价相似度越小, 则该用户的评价信息所占的权

重值越小。

结合服务个性化特征的距离向量法为:

$$\text{dist}(u_i, u_{i'})_{j,k} = \sqrt{\sum_{s_j \in E_i \cap E_{i'}} \frac{|E_i \cap E_{i'}| \delta_{j,k}^\lambda}{\sum_{s_j \in E_i \cap E_{i'}} \delta_{j,k}^\lambda} (\overline{c_{i,j,k}} - \overline{c_{i',j,k}})^2} \quad (7)$$

其中, $u_i, u_{i'}$ 为任意两个偏好的相似用户; λ 为一个可调节参数, 表示服务的个性化程度, 如果 $\lambda = 0$, 则为没有考虑服务的个性化特征; $\text{dist}(u_i, u_{i'})_{j,k}$ 为 $u_i, u_{i'}$ 为在服务 s_j 在属性 k 上的几何距离; $|E_i \cap E_{i'}|$ 表示 $u_i, u_{i'}$ 共同使用过服务的集合; $\overline{c_{i,j,k}}$ 为用户 u_i 对服务 s_j 中某一属性 k 的平均评价; $\overline{c_{i',j,k}}$ 为用户 $u_{i'}$ 对服务 s_j 中某一属性 k 的平均评价。

通过式(8)可以将用户 u_i 和 $u_{i'}$ 之间的距离进行归一化, 将其限制到 $[0, 1]$ 这个区间内, 即可得到相似度。可以采用不同的归一化标准, 这里采用的是分数归一化, 具体的形式为:

$$r_{i-i',j,k} = \frac{1}{(1 + \text{dist}(u_i, u_{i'})_{j,k})} \quad (8)$$

其中, $r_{i-i',j,k}$ 为用户 u_i 和 $u_{i'}$ 在服务 s_j 属性 k 上的评价相似度; $\text{dist}(u_i, u_{i'})$ 为 $u_i, u_{i'}$ 在服务 s_j 属性 k 上的几何距离。

如果两个用户之间共同评分项目少, 比如只有一个共同评分, 而且两个用户的评分相同, 用距离向量法算出的值为 0, 但仅仅因为两个用户对某个服务的看法相同, 就认为两用户的兴趣高度一致, 显然不符合现实, 所以需要在用户推荐度上加上加权因子 β 。

则得到:

$$\beta = \begin{cases} 1 & |E_i \cap E_{i'}| \geq \mu \\ \frac{|E_i \cap E_{i'}|}{\mu} & |E_i \cap E_{i'}| < \mu \end{cases} \quad (9)$$

其中, $|E_i \cap E_{i'}|$ 为用户 u_i 和 $u_{i'}$ 共同使用过服务的数量; μ 为设定的共同使用过服务的限定。

采用这种方法, 两个用户之间共同评分项目少会降低用户推荐度, 而两个用户之间共同评分项目多会增大用户推荐度。给定用户 $u_{i'}$ 为 u_i 关于服务 s_j 的推荐用户。则用户 $u_{i'}$ 针对用户 u_i 关于服务 s_j 在属性 k 上的用户推荐度为:

$$R_{i-i',j,k} = \beta \times r_{i-i',j,k} \times (1 + \alpha + \eta) \quad (10)$$

为此, 引入“用户推荐度”作为不同反馈数据的权重, 以它们的加权平均作为最终的质量属性值。则:

$$e_{i,j,k} = \frac{\sum_{E_i \cap E_{i'} \neq \emptyset, S_j \in E_i, S_j \in E_{i'}} R_{i-i',j,k}}{\sum_{E_i \cap E_{i'} \neq \emptyset, S_j \in E_i, S_j \in E_{i'}} R_{i-i',j,k}} \times \overline{c_{i',j,k}} \quad (11)$$

其中, $e_{i,j,k}$ 为所有推荐用户向 u_i 推荐服务 s_j 时第 k 个属性的参考数据。

4 基于综合 QoS 评价的加权服务选择

在应用上述方法得出可靠的服务提供者 QoS 值和服务使用者评价的 QoS 值后, 可基于综合 QoS 评价进行服务选择。由于在进行服务选择时, 需要综合全面考虑服务的 QoS 属性值。

通常采用加权平均的方法, 权重的设置可以采用如下几种方式: 一是由用户完全给出每个属性的权重, 即主观权重; 二是根据服务的属性值, 通过一定的算法求出权重, 即客观权重; 三是既考虑服务的主观权重又考虑服务的客观权重, 通过主观权重和客观权重再加权求得。

以上三种方式均有各自的优点。对于第一种加权方式来说, 用户可以根据自己真实的需求对不同 QoS 属性做出正确判断, 从而对相应的属性赋予满足真实需求的权重; 对于第二种加权方式来说, 由于服务对于用户而言有可能是一种全新的服务, 在使用服务时对于不同 QoS 属性并不知道如何关注。如果用户在不了解服务的情况下随意赋予权重值, 由此选择出的服务很可能并不是最优的, 针对此问题, 此时就可以采用客观权重; 对于第三种加权方式来说, 用户对某些服务并不熟悉, 虽然对某些 QoS 属性有要求, 但是并不能确定用自己的方式设置的权重选择出来的服务是最优的, 所以需要参考下属性的客观权重, 使主观权重和客观权重各占一定的比例, 用户自己完全可以根据个人需求设置该比例大小。如果希望用户自己设置的权重, 即将主观权重的比例调大一点, 如果希望客观权重, 即将客观权重的比例调大。综合上述可知, 采用第三种方式更具优势。

服务质量矩阵可以表示为:

$$X = (x_{ij})_{m \times (n_1 + n_2)} \quad (12)$$

其中, m 为可选服务的数量; $n_1 + n_2$ 为服务的属性个数。

对于客观权重, 即 $\text{cw} = (\text{cw}_1, \text{cw}_2, \dots, \text{cw}_{n_1+n_2})^T$, 采用熵值法进行计算。由熵值法可知, 某个属性信息的价值取决于 1 和该属性熵的差, 该差影响属性权重大小, 信息的价值越大, 对评价结果的影响就越大, 属性权重值也得到相应增大, 因此, 熵值法确定某个属性的权重是利用该指标属性的价值系数来计算, 该值越高对评价的影响就越大, 则权重越大。熵的计算方法如下:

$$e_j = -\tau \sum_{i=1}^m \left(\frac{x_{ij}}{\sum_{i=1}^m x_{ij}} \times \ln \frac{x_{ij}}{\sum_{i=1}^m x_{ij}} \right) \quad (13)$$

其中, τ 为常数 $\tau = 1/\ln m$; e_j 为 QoS 矩阵 Y 的属性 j 的熵值。

因此, 计算属性 j 客观权重 cw 的方法为:

$$(\mathbf{cw})_j = \frac{1 - e_j}{\sum_{j=1}^{n_1+n_2} (1 - e_j)} (0 < \mathbf{cw}_j < 1, \sum_{j=1}^{n_1+n_2} \mathbf{cw}_j = 1)$$

(14)

由于所采用的权重是主客观合成的权重,假定用户给出的主观权重为 $\mathbf{zw} = (\mathbf{zw}_1, \mathbf{zw}_2, \cdots, \mathbf{zw}_{n_1+n_2})^T$, 通过以上方法求得客观权重为 $\mathbf{cw} = (\mathbf{cw}_1, \mathbf{cw}_2, \cdots, \mathbf{cw}_{n_1+n_2})^T$, 则合成的权重, 即 $\mathbf{sw} = (\mathbf{sw}_1, \mathbf{sw}_2, \cdots, \mathbf{sw}_{n_1+n_2})^T$, \mathbf{sw}_j 的计算方法为:

$$\mathbf{sw}_j = a \times \mathbf{zw}_j + (1 - a) \mathbf{cw}_j$$

(15)

其中, a 的值可以由用户自己设置。 a 设置得大, 说明用户认为自己设置的主观权重比较重要, a 设置得小, 说明用户认为 QoS 的属性的客观权重比较重要。当 a 为 1 时, 表示合成权重完成参考于用户主观权重。一般情况下设置为 0.5; \mathbf{sw}_j 表示第 j 个属性的权重值, 可以由服务消费者自己给出, 满足 $0 \leq \mathbf{sw}_j \leq 1$ 且 $\sum_{j=1}^{n_1+n_2} \mathbf{sw}_j = 1$ 。根据属性权重的大小, 可以看出服务消费者对各属性的重要性偏向程度。

最终将矩阵 $\mathbf{X} = (x_{ij})_{m \times (n_1+n_2)}$ 和合成权重 $\mathbf{sw} = (\mathbf{sw}_1, \mathbf{sw}_2, \cdots, \mathbf{sw}_{n_1+n_2})^T$ 相乘得到的矩阵表示为 $\mathbf{Z} = (z_{ij})_{m \times 1}$, 通过式(16)求得:

$$z_{ij} = \mathbf{sw}_j \times x_{ij}$$

(16)

同时根据最终矩阵 \mathbf{Z} 选择最优服务。

5 实例分析

下面将给出一个具体实例以演示上述方法的有效性。以旅游服务为例, 生成 10 个用户集合 $U = \{u_1, u_2, \cdots, u_{10}\}$ 和 10 个旅游商家集合 $S = \{s_1, s_2, \cdots, s_{10}\}$, 并选取旅游商家的服务费用、服务时间、服务满意度、服务可靠性 4 个指标作为服务的 QoS 属性。其中旅游商家的服务费用和服务时间这两个属性数据主要来源于服务提供者, 服务满意度和服务可靠性这两个属性数据主要来源于服务使用者。所采用的是 10 分制。

表 1 中 $R_{i,11}$ 表示服务提供者当前发布的 QoS 属性数据, $R_{i,0}, R_{i,1}, \cdots, R_{i,10}$ 表示利用采集机制收集的 11 次历史数据(其中 $R_{i,0}$ 是服务最初发布的数据, 采集的数据时间间隔为 1 天)。

表 2 为用户评价矩阵, 由于用户可以多次使用服务, 所以每个单元格对应每次使用服务的评价信息, 在单元格中前两个数字表示用户的评价值而后一个数字表示服务的使用时间(由于用户评价数据较多, 所以只抽取了 u_1, u_2, u_{10} 的数据)。

5.1 获取可靠的服务提供者发布的属性值

Step1: 基于采集到的 $R_{i,0}, R_{i,1}, \cdots, R_{i,10}$ 数据, 通过式(3)计算综合历史的 QoS 值。以 s_2 中服务费用属性

表 1 10 个服务对各属性 12 次历史描述信息表

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
$R_{i,0}$	4,5	4,8	3,6	5,6	6,5	4,6	7,4	8,6	4,7	7,8
$R_{i,1}$	4,6	5,7	8,6	4,6	7,5	6,7	5,4	7,6	5,5	7,6
$R_{i,2}$	3,6	4,4	4,6	5,5	4,7	4,7	6,6	7,6	5,6	4,6
$R_{i,3}$	4,5	3,5	5,6	3,3	3,6	5,6	3,5	5,4	4,5	5,5
$R_{i,4}$	3,2	5,6	4,5	4,6	4,5	3,7	4,5	5,5	4,7	7,6
$R_{i,5}$	2,6	6,6	4,7	5,6	6,5	4,6	5,6	7,6	6,6	3,7
$R_{i,6}$	4,7	4,3	3,3	7,5	7,5	5,6	7,6	5,6	5,6	6,6
$R_{i,7}$	5,6	7,4	4,5	4,5	5,6	5,7	3,5	3,7	4,5	3,7
$R_{i,8}$	3,3	5,6	6,6	4,4	4,6	6,7	5,5	4,6	4,6	4,6
$R_{i,9}$	3,5	3,4	3,5	5,6	5,7	4,6	4,7	6,5	3,7	5,6
$R_{i,10}$	4,6	5,3	5,6	6,5	4,7	5,7	3,6	3,4	5,6	6,8
$R_{i,11}$	5,6	3,6	4,6	4,7	5,6	6,5	6,6	3,5	5,7	7,5

表 2 3 个用户对 10 个服务各属性的历史评价信息表

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
	3,6,2		5,6,5	5,5,5	8,9,3		4,6,7	5,3,2		4,4,6
u_1					6,5,5			7,7,3		
								9,2,1		
		3,6,5	5,5,6		5,7,6			7,6,4		3,6,7
u_2					8,2,4					
					5,3,2					
									
	5,5,6			4,5,3			5,4,6	7,5,3	6,5,8	5,6,4
	7,2,7			7,8,1			7,5,4		4,9,6	4,5,2
u_{10}	4,1,6			7,8,2			3,4,7			9,7,3
	8,3,9						1,3,3			

为例, 有:

$$f(R_{i,0}, R_{i,1}, \cdots, R_{i,10}) =$$

$$(1 \times 5 + 2 \times 4 + 3 \times 3 + 4 \times 5 + 5 \times 5 + 6 \times 4 +$$
$$7 \times 7 + 8 \times 5 + 9 \times 3 + 10 \times 5) / (1 + 2 + 3 + 4 +$$
$$5 + 6 + 7 + 8 + 9 + 10) = \frac{257}{55}$$

同理服务时间计算的值为 $\frac{244}{55}$, 即 $< \frac{257}{55}, \frac{244}{55} >$ 。

同理求得 s_6, s_9 分别为 $< \frac{273}{55}, \frac{364}{55} >$, $< \frac{240}{55}, \frac{332}{55} >$ 。

Step2: 通过式(1)对服务 s_2 发布的数据进行修正, 设 $w = 0.5$ 。计算结果为 $< 3.836\ 36, 5.218\ 18 >$ 。同理求得 s_6, s_9 分别为 $< 5.481\ 81, 5.790\ 91 >$, $< 4.681\ 81, 6.518\ 18 >$ 。

5.2 获取可靠的服务使用者反馈值

Step1: 通过平均评价公式, 针对表 2 数据, 以用户 u_1 使用服务 s_8 为例, 则用户 u_1 关于服务 u_1 中第一个属性的平均评价为 $\frac{5+7+9}{3} = 7$, 关于第二个属性平均评价为 $\frac{3+7+2}{3} = 4$ 。同理可得其他用户关于各个服务的平均评价, 最终得到的用户平均评价矩阵如表

3 所示。

表 3 10 个用户对 10 个服务各属性的平均评价信息表

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
u_1	3,6		5,6	5,5	7,7		4,6	7,4		4,4
u_2		3,6	5,5		6,4			7,6		3,6
u_3	3,5			8,3		2,4		5,6	6,6	
u_4	6,6	2,9	6,7				5,6	3,5		
u_5	4,5		6,3		3,3		3,6			
u_6	4,7			2,6		4,6		6,4	5,6	5,7
u_7		5,7	5,5		7,6			6,5		4,3
u_8	7,6		5,6		5,6	7,5	6,7	3,4		
u_9	4,5	3,5				6,5	4,5		3,5	5,4
u_{10}	6,3			6,7			4,4	7,5	5,7	6,6

Step2:通过平均评价公式,针对表 2 数据,以用户 u_1 使用服务 s_8 为例,则用户 u_1 使用服务 s_8 的次数为 3,关于服务使用平均时长为 $\frac{2+3+1}{3}=2$ 。同理可得到其他用户关于各个服务的使用次数和服务使用平均时长,最终得到用户使用服务次数和服务使用平均时长两个属性组成的矩阵,具体结果见表 4。

表 4 用户使用服务次数和服务使用平均时长信息表

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
u_1	1,2		1,5	1,5	2,4		1,7	3,2		1,6
u_2		1,5	1,6		3,4			1,4		1,7
u_3	2,4			2,2		2,2		2,2	1,5	
u_4	1,3	4,6	1,4				1,2	4,2		
u_5	1,2		1,6		1,4		2,8		1,6	
u_6	2,2			1,4		1,5		1,5		1,8
u_7		3,2	3,2		1,2			3,2		1,2
u_8	1,2		2,2		1,5	1,2	3,2	1,2		
u_9	3,6	2,3				1,2	1,2		2,5	5,4
u_{10}	4,7			3,2			4,5	1,3	2,7	3,3

Step3:通过式(4)和式(5),针对表 4 数据,以服务 s_2 为例, $\alpha = \frac{1}{1+4+3+2} = \frac{1}{10}$, $\eta = \frac{1}{5+6+2+3} = \frac{5}{16}$ 。则 u_2 在该服务中的用户领域相关度和时间系数分别为 $\{\alpha = 1/10, \eta = 5/16\}$;同理得 u_4, u_7, u_9 用户领域相关度和时间系数分别为 $\{\alpha = 4/10, \eta = 6/16\}$, $\{\alpha = 3/10, \eta = 2/16\}$, $\{\alpha = 2/10, \eta = 3/16\}$ 。同理求得服务 s_6 中 u_3, u_6, u_8, u_9 的用户领域相关度和时间系数分别为 $\{2/5, 2/11\}, \{1/5, 5/11\}, \{1/5, 2/11\}, \{1/5, 2/11\}$;服务 s_9 中 u_3, u_5, u_9, u_{10} 的用户领域相关度和时间系数分别为 $\{1/6, 5/23\}, \{1/6, 6/23\}, \{2/6, 5/23\}, \{2/6, 7/23\}$ 。

Step4:通过式(6)和式(7),针对表 2 数据,以服务 s_2 为例,求得目标用户 u_1 和参考用户 u_2 关于第一个属性的几何距离数据 470 23。然后通过式(8)求得相似

度为 $\frac{1}{1+1.470\ 23} \approx 0.404\ 82$,同理求得目标用户 u_1 和参考用户 u_4, u_7, u_9 关于第一个属性的相似度分别为 0.143 31,0.463 81,0.812 68。同理求得目标用户 u_1 和参考用户 u_2, u_4, u_7, u_9 关于第二个属性的相似度为 0.186 85,0.414 99,0.366 03,0.428 65。同理求得服务 s_6 中目标用户 u_1 和参考用户 u_3, u_6, u_8, u_9 关于第一个属性的相似度分别为 0.205 10,0.203 02,0.125 22,0.406 34;关于第二个属性的相似度为 0.249 65,0.218 41,0.259 75,0.428 65。服务 s_9 中目标用户 u_1 和参考用户 u_3, u_6, u_9, u_{10} 关于第一个属性相似度分别为 0.205 10,0.203 02,0.406 34,0.213 15;关于第二个属性相似度分别为 0.249 65,0.218 41,0.428 65,0.180 94。

Step5:假设 $\beta = 1$;以服务 s_2 为例,通过式(10)求得 u_2, u_4, u_7, u_9 的关于第一个属性推荐度分别为 0.571 81,0.254 37,0.660 93,1.127 59;关于第二个属性推荐度为 0.263 92,0.736 61,0.521 59,0.594 75。同理求得服务 s_6 中 u_3, u_6, u_8, u_9 关于第一个属性推荐度分别为 0.324 43,0.335 90,0.173 03,0.561 48;关于第二个属性推荐度分别为 0.394 89,0.361 37,0.358 93,0.592 31。服务 s_9 中 u_3, u_6, u_9, u_{10} 关于第一个属性推荐度分别为 0.283 87,0.289 81,0.630 12,0.349 07;关于第二个属性推荐度分别为 0.345 53,0.311 78,0.664 72,0.296 32。

Step6:以服务 s_2 为例,据式(11)得第一个属性可靠反馈值为 $\frac{0.571\ 81}{2.614\ 70} \times 3 + \frac{0.254\ 37}{2.614\ 70} \times 2 + \frac{0.660\ 93}{2.614\ 70} \times 5 + \frac{1.127\ 59}{2.614\ 70} \times 3 \approx 3.408\ 26$,同理第二个属性反馈值为 7.009 35;即服务 s_2 可靠的属性值为 $\langle 3.408\ 26, 7.009\ 35 \rangle$ 。同理求得服务 s_6 可靠反馈的属性值为 $\langle 4.712\ 03, 4.980\ 35 \rangle$,服务 s_9 可靠反馈的属性值为 $\langle 4.371\ 24, 5.772\ 34 \rangle$ 。

5.3 基于 QoS 综合评价的服务选择

Step1:通过前两个过程得出的初始矩阵为: $\begin{Bmatrix} 3.836\ 36 & 5.218\ 18 & 3.408\ 26 & 7.009\ 35 \\ 5.481\ 81 & 5.790\ 91 & 4.712\ 03 & 4.980\ 35 \\ 4.681\ 81 & 6.518\ 18 & 4.371\ 24 & 5.772\ 34 \end{Bmatrix}$ Step2:通过式(13),可以求得 e_1, e_2, e_3, e_4 分别为 0.990 47,0.996 22,0.991 78,0.991 01。 Step3:通过式(14),有: $cw_1 = (1 - 0.990\ 47)/((1 - 0.990\ 47) + (1 - 0.996\ 22) + (1 - 0.991\ 78) + (1 - 0.981\ 01)) \approx 0.312\ 25$,同理 cw_2, cw_3, cw_4 分别为 0.123 85,0.269 33,0.294 56。即 $cw = (0.312\ 25, 0.123\ 85,$

$0.269\ 33, 0.294\ 56)^T$ 。

Step4: 假设目标用户的主观权重值为 $\mathbf{zw} = (zw_1, zw_2, \dots, zw_4)^T = (0.3, 0.2, 0.4, 0.1)^T$, 设 $a = 0.5$, 通过式 (15), 有 $sw_1 = 0.5 \times 0.3 + (1 - 0.5) \times 0.312\ 25 = 0.306\ 13$, 同理求得 $sw_2 = 0.161\ 92$, $sw_3 = 0.334\ 67$, $sw_4 = 0.197\ 28$; 即 $\mathbf{sw} = (0.306\ 13, 0.161\ 92, 0.334\ 67, 0.197\ 28)^T$ 。

Step5: 根据式 (16), 有:

$$\begin{Bmatrix} 3.836\ 36 & 5.218\ 18 & 3.408\ 26 & 7.009\ 35 \\ 5.481\ 81 & 5.790\ 91 & 4.712\ 03 & 4.980\ 35 \\ 4.681\ 81 & 6.518\ 18 & 4.371\ 24 & 5.772\ 34 \end{Bmatrix} \times (0.306\ 13, 0.161\ 92, 0.334\ 67, 0.197\ 28)^T = \begin{Bmatrix} 4.542\ 79 \\ 5.175\ 30 \\ 5.339\ 28 \end{Bmatrix}$$

Step6: 由于 $5.339\ 28 > 5.175\ 30 > 4.542\ 79$, 所以服务 s_9 最适合目标用户 u_1 。

6 结束语

为了向目标用户提供全面可靠的 QoS 选择依据, 提出了一种基于 QoS 综合评价的服务选择方法。对于来自服务提供者描述的 QoS, 基于历史数据的统计对服务提供者描述的 QoS 数据加以修正; 对于来自用户反馈的 QoS 评价, 通过计算用户之间以往反馈的相似程度, 同时引入用户领域相关度, 时间系数和服务个性化等概念对用户相似程度进行修正, 得到最终推荐度, 然后由用户推荐度所占的比重作为权重计算出最终评价; 在计算出所有的结果后, 针对目标用户对主观给出的各 QoS 属性权重值可能不太确定的现状, 采用熵值法计算客观权重并结合用户的主观权重加权进行整合, 从而进行服务选择。通过实例验证了该方法的有效性, 提高了服务选择的准确性。下一步工作中, 将结合数理统计理论对相似度计算公式进行深入研究。

参考文献:

[1] 王海艳, 杨文彬, 王随昌, 等. 基于可信联盟的服务推荐方

法[J]. 计算机学报, 2014, 37(2): 301-311.

- [2] 曹利培, 李爱玲, 刘 静. 基于 QoS 的两阶段 Web 服务选择方法[J]. 计算机工程与设计, 2009, 30(3): 747-751.
- [3] Zheng Z, Ma H, Lyu M R, et al. Wsrec: a collaborative filtering based web service recommender system[C]//IEEE international conference on web services. [s. l.]: IEEE, 2009: 437-444.
- [4] 邵凌霄, 李 田, 赵俊峰, 等. 一种可扩展的 Web Service QoS 管理框架[J]. 计算机学报, 2008, 31(8): 1458-1470.
- [5] 郭得科, 任 彦, 陈洪辉, 等. 一种基于 QoS 约束的 Web 服务选择和排序模型[J]. 上海交通大学学报, 2007, 41(6): 870-875.
- [6] 杨胜文, 史美林. 一种支持 QoS 约束的 Web 服务发现模型[J]. 计算机学报, 2005, 28(4): 589-594.
- [7] 巫 茜, 周 庆. 基于 QoS 与可信度融合的 Web 服务选择机制研究[J]. 计算机科学, 2012, 39(7): 108-111.
- [8] Mehdi M, Bouguila N, Bentahar J. A QoS-based trust approach for service selection and composition via Bayesian networks[C]//Proceedings of ICWS' 13. [s. l.]: [s. n.], 2013: 211-218.
- [9] 马 友, 王尚广, 孙其博, 等. 一种综合考虑主客观权重的 Web 服务 QoS 度量算法[J]. 软件学报, 2014, 25(11): 2473-2485.
- [10] 宋海霞, 严 馨, 余正涛, 等. 基于自适应聚类的虚假评论检测[J]. 南京大学学报: 自然科学版, 2013, 49(4): 433-438.
- [11] Bellare M, Namprempre C, Pointcheval D, et al. The power of RSA inversion oracles and the security of Chaum's RSA-based blind signature scheme[C]//Lecture notes in computer science. Berlin: Springer, 2002: 319-338.
- [12] Okamoto T. Efficient blind and partially blind signatures without random oracles[C]//Lecture notes in computer science. Berlin: Springer, 2006: 80-99.
- [13] Camenisch J, Koprowski M, Warinschi B. Efficient blind signatures without random oracles[C]//Lecture notes in computer science. Berlin: Springer, 2005: 134-148.
- [14] Boneh D, Boyen X. Efficient selective-ID secure identity based encryption without random oracles [C]//Lecture notes in computer science. Berlin: Springer, 2004: 223-238.