

面向用户均衡需求的 Web 服务资源智能推荐方法*

刘平峰, 张 慧, 章佩璐

(武汉理工大学经济学院 电子商务与智能服务研究中心, 武汉 430070)

摘要: 研究了满足用户在服务功能、服务属性和服务质量上的多层次均衡需求的 Web 服务资源智能推荐方法, 解决了服务信息过载问题。现有服务推荐技术存在未考虑用户对服务多层次均衡需求的不足, 结合 Web 服务资源的特点, 综合考虑用户服务需求与候选服务群之间的多层次相似度, 设计了综合 Web 服务资源的功能、属性和质量三个层次的适合度匹配算法, 以及用户主观兴趣度匹配算法。通过这两个算法建立了面向用户均衡需求的个性化 Web 服务资源智能推荐方法, 能将最大匹配度的 Web 服务资源推荐给目标用户。

关键词: Web 服务资源; 智能推荐; 功能属性; 非功能属性

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2012)09-3224-05

doi:10.3969/j.issn.1001-3695.2012.09.006

Method of intelligent recommendation of Web service resources towards users' balanced requirements

LIU Ping-feng, ZHANG Hui, ZHANG Pei-lu

(Research Center for E-business & Intelligence Service, School of Economics, Wuhan University of Technology, Wuhan 430070, China)

Abstract: This paper proposed an approach of intelligent recommendation of Web service resources to users. This approach can satisfy users' balanced multi-level requirements in service function, service attribute and service quality, as well as can solve the problem of service information overload. Present service recommendation techniques do not explore users' balanced multi-level requirements in service. Taking into consideration the characteristics of Web service resources and exploring the multi-level similarity between user's requirements in services and service resource candidates, this paper designed the algorithm of integrated fitness matching in function, attributes and quality of Web service resources, as well as algorithm of matching users' subjective interests. Through the above work, this paper proposed a method of personalized intelligent recommendation of Web service resources oriented towards users' balanced requirements. This method can ensure that the service resource with the maximum matching degree be recommended to the targeted user.

Key words: Web service resource; intelligence recommendation; FQoS; QoS

0 引言

随着管理实践的深化发展以及经营模式的变革影响, Web 服务市场已逐渐演变为一个以用户为导向、感性和理性并存的消费市场, 用户对于 Web 服务的需求也逐渐从单一的服务功能属性扩展到服务质量层面上来。同时, 由于 Web 服务信息的不断膨胀和服务资源的分散化, 使得服务用户逐渐陷入了“信息过载”的困境中。因此, 如何快速准确地掌握用户的多层次服务需求, 将服务提供商从被动接收用户的请求转换为主动感知用户的需求, 并在此基础上进行准确有效的 Web 服务资源推荐已然成为 Web 服务行业面临的挑战性问题。

在 Web 服务推荐研究方面, Thio 等人^[1]在认识到现阶段 Web 服务推荐主要侧重于从服务提供商绩效角度进行考虑的弊端的基础上, 提出了基于用户概要的 Web 服务推荐系统, 实现了以用户绩效为中心的最优 Web 服务供应商的选择。文献[2]结合领域本体和代理技术, 通过构建个性化的用户模型来

实现个性化 Web 服务推荐。朱红康等人^[3]构建了一个能够用来记录 QoS 执行信息的框架, 同时设计了一种基于时间加权的匹配算法, 从而实现了基于 QoS 执行信息的动态 Web 服务推荐方法。文献[4]提出了一个 WSRec 服务推荐系统, 该系统通过一种用户贡献度机制来收集 Web 服务的 QoS 信息, 并利用一个有效的协同过滤算法来实现 Web 服务 QoS 值的预测。蔡华利等人^[5]为了解决 Web 服务的优化选择问题, 利用数学方法来阐述基于 QoS 的 Web 服务组合, 提出了一种基于离散二进制粒子群算法的 Web 服务推荐策略。还有很多学者也都是通过 QoS 匹配优化选择方法、协同过滤推荐方法、用户偏好模型等相关手段来实施 Web 服务的推荐的。但是这些方法不仅存在冷启动、数据稀疏等显著问题, 而且没有结合 Web 服务资源的自身特点以及用户对于 Web 服务资源在服务功能、服务属性以及服务质量上的多层次需求目标, 从而无法真正实现用户需求 and 偏好的统一, 缺乏推荐系统的层次性、动态性和可扩展性。

收稿日期: 2012-02-14; **修回日期:** 2012-04-01 **基金项目:** 国家自然科学基金资助项目(71072077, 71172043); 国家科技支撑计划资助项目(2011BAH16B02); 国家教育部回国留学人员科研启动基金资助项目(2010-1561); 武汉理工大学中央高校基本科研业务专项资金资助项目(2012-IB-060)

作者简介: 刘平峰(1972-), 男, 湖北荆门人, 副教授, 博士, 主要研究方向为商务智能、服务管理、社会计算(lpf@whut.edu.cn); 张慧(1986-), 女, 湖北武汉人, 硕士研究生, 主要研究方向为语义网、商务智能、服务管理; 章佩璐(1986-), 女, 浙江宁波人, 硕士, 主要研究方向为语义网、商务智能、服务管理。

本文提出的 Web 服务资源推荐算法重点围绕适合度匹配和兴趣度挖掘两方面展开,利用适合度匹配得到满足用户功能属性 FQoS 和非功能属性 QoS 需求的 Web 服务资源,进一步利用兴趣度挖掘与匹配得到最能满足用户主观兴趣的 Web 服务资源,这样通过计算用户需求服务与候选服务群之间的多层次相似度,将最大匹配度的 Web 服务资源推荐给目标用户。该推荐模型主要由三大模块组成,自下而上分别是 Web 服务资源功能属性匹配模块、非功能属性匹配模块和用户兴趣度匹配模块。其具体实现路径如图 1 所示。

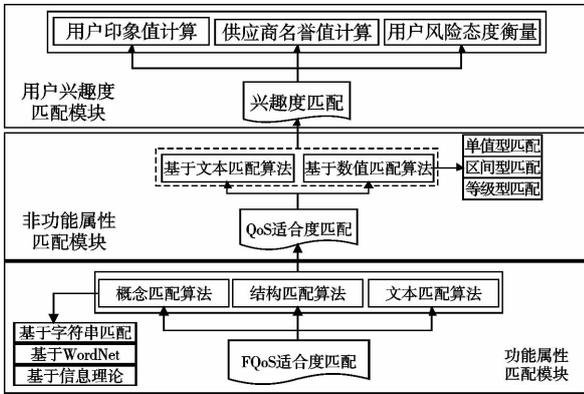


图1 Web服务资源智能推荐模型

1 基于 FQoS 适合度的 Web 服务资源功能属性匹配模块

适合度主要是用于衡量用户需求的 Web 服务资源在功能属性和非功能属性上与 Web 服务资源提供商提供的服务资源群之间的匹配程度,用于从候选 Web 服务资源群中选择出在功能、属性以及质量上与用户需求最为相近的 Web 服务资源。本文提出的基于适合度计算的 Web 服务资源智能推荐方法的工作流程如图 2 所示。

FQoS 适合度匹配是根据 FQoS 不同类型计算得到的,主要从概念匹配、结构匹配和文本匹配三方面考虑,并将三种方法得到的匹配结果求和取平均,即

$$\text{sim}(FQoS_1, FQoS_2) = \frac{\text{sim}(c_1, c_2) + \text{sim}(X_1, X_2) + \text{sim}(D_1, D_2)}{3} \quad (1)$$

其中: $\text{sim}(FQoS_1, FQoS_2)$ 是用户需求的 Web 服务资源与供应商提供的 Web 服务资源之间的 FQoS 适合度, $\text{sim}(FQoS_1, FQoS_2) \in [0, 1]$ 。 c_1, X_1 和 D_1 分别是用户需求 Web 服务资源对应的概念类型、结构类型和文本类型 FQoS; c_2, X_2 和 D_2 则是供应商提供的 Web 服务资源所对应的概念类型、结构类型和文本类型 FQoS; $\text{sim}(c_1, c_2)$ 表示概念 c_1 和 c_2 的相似度匹配值; $\text{sim}(X_1, X_2)$ 表示 XML 文档 X_1 和 X_2 的相似度匹配值; $\text{sim}(D_1, D_2)$ 表示非结构化文档 D_1 和 D_2 的相似度匹配值,其含义及算法将在下面进行详细讨论。

1.1 FQoS 的概念匹配算法

概念匹配主要用来处理诸如“服务分类”等原子标签类型的 FQoS。本文在现有概念语义相似度算法的基础上,提出一种基于字符串匹配算法、WordNet 以及信息理论的综合语义相似度算法,旨在利用语义相似度算法解决字符串匹配算法中存在的语义区分含糊问题,同时利用字符串匹配算法来减少复合词匹配带来的误差。 $\text{Sim}(c_1, c_2)$ 表示概念 c_1 和 c_2 的综合语义相似度,计算式为

$$\text{sim}(c_1, c_2) = w_1 \text{sim}_1(c_1, c_2) + w_2 \text{sim}_2(c_1, c_2) + w_3 \text{sim}_3(c_1, c_2) \quad (2)$$

其中: $\text{sim}_1(c_1, c_2)$ 、 $\text{sim}_2(c_1, c_2)$ 和 $\text{sim}_3(c_1, c_2)$ 分别表示三种不同概念匹配算法得到的相似度结果; w_1, w_2 和 w_3 为权重系数, $w_1 + w_2 + w_3 = 1$ 。

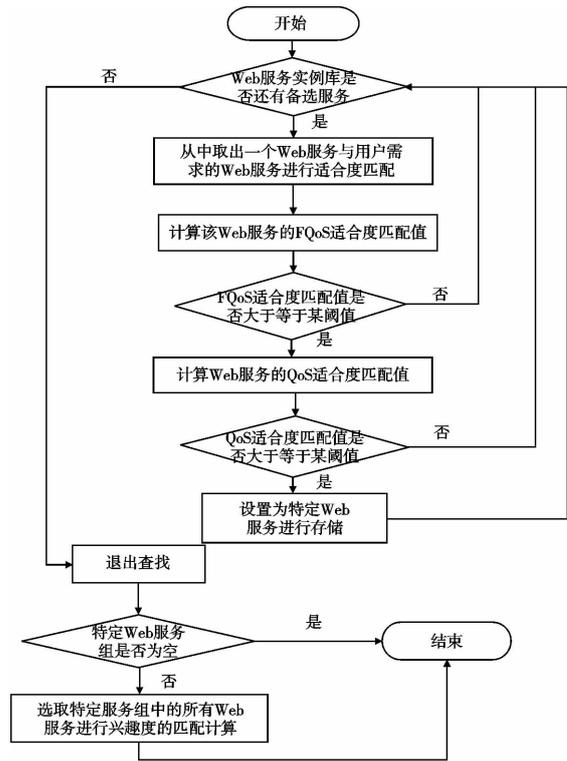


图2 基于适合度的Web服务资源智能推荐流程

1.1.1 基于字符串匹配的概念相似度算法

基于字符串匹配算法通过计算字符串的编辑距离实现元素间的相似度匹配。本文采用文献[6]中引用的 Maedche & Staab 所提出的概念相似度算法:

$$\text{sim}_1(c_1, c_2) = \max\left(0, \frac{\min(|c_1|, |c_2|) - D(c_1, c_2)}{\min(|c_1|, |c_2|)}\right) \quad (3)$$

其中: c_1 和 c_2 分别是两个不同的概念; $|c|$ 表示概念 c 所占的字符长度; $D(c_1, c_2)$ 则表示两个概念之间的编辑距离。这种方法将字符串的编辑距离转换成了一个 $[0, 1]$ 间的数值。

1.1.2 基于 WordNet 的概念语义相似度算法

WordNet 是一部基于认知语言学的英文语义词典,在语义匹配中有广泛应用。综合考虑 WordNet 的语义特点以及现阶段各学者对 WordNet 的研究成果,本文采用文献[7]中提到的 Wu-Palmer 语义相似度算法:

$$\text{sim}_2(c_1, c_2) = \frac{2 \times \text{depth}(\text{lso}(c_1, c_2))}{\text{depth}(c_1) + \text{depth}(c_2)} \quad (4)$$

其中: c_1 和 c_2 分别是两个不同的概念; $\text{depth}(c_1)$ 和 $\text{depth}(c_2)$ 是 c_1 和 c_2 在 WordNet 语义树中的深度,而 $\text{lso}(c_1, c_2)$ 描述的是 c_1 和 c_2 的最近共同祖先概念。 $\text{sim}_2(c_1, c_2) \in [0, 1]$ 。

1.1.3 基于信息理论的概念语义相似度算法

根据 Ross 的信息理论,本体中的一个概念 c 的频率 $\text{freq}(c)$ 可以定义为概念 c 与其所有子节点在本体中出现的次数,即 $\text{freq}(c) = \sum \{ \text{occur}(c_i) \mid c \in \text{ancestors}(c_i) \}$ 。而概念 c 的信息内容可定义为 $\text{IC}(c) = -\log(\text{prob}(c))$,其中 $\text{prob}(c) = \text{freq}(c)/N$, N 为本体语料库中所有概念总和。

语义相似度计算中两个概念的相似度与它们之间共享信息的程度有关。因此给定概念 c_1 和 c_2 ,则它们的共享信息可

以表示为两个概念之间信息相似程度最高的祖先节点的信息内容,即 $share(c_1, c_2) = \max \{IC(a) \mid a \in sub(c_1, c_2)\}$, 其中 $sub(c_1, c_2)$ 是同时包含 c_1 和 c_2 的概念。

为了实现标准化,使最终得到的计算结果取值为 $[0, 1]$, 本文定义基于信息理论的概念相似度算法式如下:

$$sim_3(c_1, c_2) = \frac{2 \times share(c_1, c_2)}{IC(c_1) + IC(c_2)} \quad (5)$$

值得注意的是,在整个 Web 服务的 FQoS 匹配过程中,出现的服务名称等概念很多都是诸如“ReservationBooking”这种类型的复合词。考虑到复合词的特殊性,本文采用文献[8]中的方法先将复合词进行标准化:首先分解复合词,得到若干原子词汇;分析这些原子词汇的形态,发现其基本形式;去掉一些没有表达意义的如停顿词的词汇。在对复合词标准化后,采取以下步骤来解决 FQoS 中的复合词匹配问题(本文以“ReservationBooking”和“HotelAccommodation”为例进行说明):

a) 将两个需要匹配的复合词进行标准化,并将得到的相关原子词汇作为节点构建一个完全连通的二部图。标准化后得到的四个原子词汇分别是 reservation、booking、hotel 和 accommodation,而它们构成的二部图如图 3 所示。

b) 利用上述概念语义相似度计算方法分别计算某一复合词的原子词汇与另一个复合词的原子词汇之间的相似度,并将所得值作为两原子词汇连接边上的权重。如图 4 所示,计算得到的结果为

$$\begin{aligned} sim(reservation, booking) &= 0.954 \\ sim(reservation, accommodation) &= 0.748 \\ sim(hotel, booking) &= 0.207 \\ sim(hotel, accommodation) &= 0.632 \end{aligned}$$

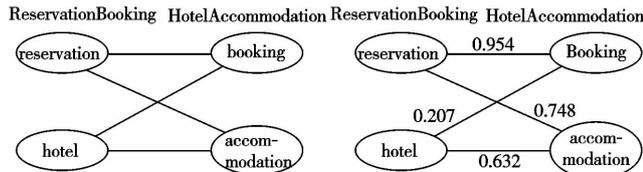


图3 构建二部图

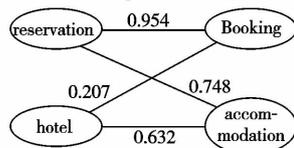


图4 计算二部图的边权重

c) 利用二部图的最大流匹配方法来解决节点之间的分配问题。由于 $(0.954 + 0.632) > (0.748 + 0.207)$, 则得到的分配结果如图 5 所示。

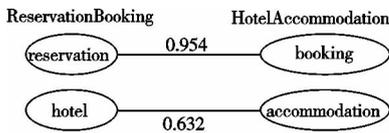


图5 二部图最大流分配

d) 以匹配后得到的权重总和为分子,以两个复合词分解得到的有效原子词汇个数中的较大值为分母,构建计算复合词相似度的分数,即

$$sim(ReservationBooking, HotelAccommodation) = (0.954 + 0.632) / \max(2, 2) = 0.793$$

1.2 FQoS 的结构匹配算法

结构匹配用来处理如“输入输出数据定义”等结构化 XML 类型的 FQoS。结构匹配比概念匹配过程更为复杂,因为需要在原子标签的基础上对更多的复杂树型结构(如 XML DOM 树型结构)进行分析。现阶段使用比较多的树型结构匹配主要包括节点、边界和路径匹配,树的编辑距离(TED)计算,标签相似度计算以及基于转换次数的匹配等,这些方法不仅存在着时间复杂度大的问题,而且没有充分考虑到 XML 文档结构中的

重复元素和嵌套元素。

根据 XML 文档的特点,本文采用文献[9]提出的方法对 XML 树型结构进行相似度计算。该方法首先利用 XML 文档树的层次结构关系抽取相应的文档特征信息;然后根据建模后的层次模型对不同层次的节点元素进行不同的权重分配,使高层次的节点元素权重大于低层次的节点元素权重;最后利用相似度计算式对两 XML 文档进行匹配,计算式如式(6)所示。

$$sim(X_1, X_2) = \frac{X_1 \text{ 和 } X_2 \text{ 中公共节点的权重}}{X_1 \text{ 和 } X_2 \text{ 所有节点的权重}} = \frac{\sum_{i=1}^{L_1} w_{1i}^{L_1-i} \times s_{1i} + \sum_{j=1}^{L_2} w_{2j}^{L_2-j} \times s_{2j}}{\sum_{i=1}^{L_1} w_{1i}^{L_1-i} \times s_{1m} + \sum_{j=1}^{L_2} w_{2j}^{L_2-j} \times s_{2n}} \quad (6)$$

其中, L_1 和 L_2 分别表示文档 X_1 和文档 X_2 的层数; w_{1i} 和 w_{2j} 分别指文档 X_1 中第 i 层节点的权重以及文档 X_2 中第 j 层节点的权重; s_{1i} 表示文档 X_1 中第 i 层节点与文档 X_2 中相同的节点个数; s_{2j} 表示文档 X_2 中第 j 层节点与文档 X_1 中相同的节点个数; 而 s_{1m} 和 s_{2n} 则分别表示文档 X_1 中第 m 层节点的总数以及文档 X_2 中第 n 层节点的总数。式(6)的计算结果取值也为 $[0, 1]$ 。

1.3 FQoS 的文本匹配算法

文本匹配用来处理诸如“注释”等文本数据类型的 FQoS, 可以通过静态文本挖掘技术得以实现,主要是通过计算两个或多个文本之间的相似度来获得其匹配程度的度量值。本文主要采用统计学方法中的向量空间模型(VSM)^[10]来对文本类型的 FQoS 进行相似度匹配。该方法的主要思想是:首先将文本 D 构成一个 n 维的文本向量空间;然后利用式(7)比较向量之间的余弦系数,余弦值越大则表明文本之间的相似度越高。

$$sim(D_1, D_2) = \cos\theta = \frac{\sum_{k=1}^n w_{1k} \times w_{2k}}{\sqrt{(\sum_{k=1}^n w_{1k}^2) \times (\sum_{k=1}^n w_{2k}^2)}} \quad (7)$$

其中: w_{1k} 是特征项 k 在文本 D_1 中的权重值; w_{2k} 是特征项 k 在文本 D_2 中的权重值。它们均通过 TF-IDF 公式计算得到:

$$w_{ik} = \frac{tf_{ik} \log(N/n_k + 0.5)}{\sqrt{\sum_{k=1}^n (tf_{ik})^2 \times [\log(N/n_k + 0.5)]^2}} \quad (8)$$

其中: w_{ik} 描述的是特征项 k 在文本 D_i 的权重值; tf_{ik} 是特征项 k 在文本 D_i 中出现的次数,即词频; $\log(N/n_k + 0.5)$ 表示特征词 k 的倒排文本频率, N 为匹配文本的个数(这里为 2), n_k 是含有特征项 k 的文本数。同样地,由于这里计算的相似度是通过余弦值得到的,故最终结果取值为 $[0, 1]$ 。

2 基于 QoS 适合度的 Web 服务资源非功能属性匹配模块

用户对 Web 服务的非功能性需求 QoS 参数所包含的内容同时具有语义信息和数值信息,传统的概念相似度算法无法满足 QoS 的这种双重特性。因此 QoS 参数适合度的匹配过程是“对象”之间的匹配,即不仅是 QoS 参数名称之间的概念匹配,还包括整个 QoS 值和其所对应的单位之间的匹配。例如,“响应时间”和“价格”描述的是两个完全不相容的 QoS 属性,因此将两者进行比较是毫无意义的。同样地,由于存在着单位不一致,将 5 美元和 6 元人民币进行适合度计算也是没有必要的。因此必须将基于文本的 QoS 参数和基于数值的 QoS 参数综合起来进行考虑。本文将 QoS 适合度定义为

$$\text{sim}(QoS_1, QoS_2) = \frac{\text{sim}(w_1, w_2) + \text{sim}(d_1, d_2)}{2} \quad (9)$$

其中: $\text{sim}(QoS_1, QoS_2)$ 是用户需求的 Web 服务资源与供应商提供的 Web 服务资源之间的 QoS 适合度, $\text{sim}(QoS_1, QoS_2) \in [0, 1]$; w_1 和 d_1 分别是用户需求 Web 服务资源对应的文本型和数值型 QoS, 而 w_2 和 d_2 则是供应商提供的 Web 服务资源对应的文本型和数值型 QoS; $\text{sim}(w_1, w_2)$ 是文本型 QoS 之间的匹配相似值, 而 $\text{sim}(d_1, d_2)$ 表示数值型 QoS 的匹配相似值。

2.1 基于文本的 QoS 适合度匹配算法

文本可以指 QoS 参数的属性名称, 如可靠性、安全性; 也可以指 QoS 属性的单位, 如美元、毫秒等。由于单纯从字面上对属性单位进行语义层面上的匹配没有意义, 需要引入单位之间的换算机制。本文主要从属性名称的角度来考虑文本型的 QoS 参数。

文本型 QoS 参数的概念匹配算法采用 1.1.1 节中介绍的语义相似度算法。将用户需求的 QoS 参数从文本概念层次上与 Web 服务提供商提供的服务群 QoS 参数进行语义相似度计算, 即

$$\text{sim}(w_1, w_2) = \text{sim}(C_1, C_2) \quad (10)$$

其中: C_1 和 C_2 分别表示用户需求 Web 服务资源 QoS 参数名称和提供商提供的 Web 服务资源 QoS 参数名称。

2.2 基于数值的 QoS 适合度匹配算法

数值型 QoS 参数包括单值型 (如服务价格为 \$90)、区间型 (如响应时间在 0 ~ 100 ms 之间) 以及等级型 (如安全性可以为很低、低、一般、高、很高)。基于不同类型 QoS 参数的不同数值特点, 其匹配方法如下:

1) 单值型 QoS 参数的匹配

假定 Q_1 和 Q_2 是两个具有正向趋势的单值型 QoS 参数; m 和 n 分别是 QoS 参数标准化后的属性值, 即 m 和 n 取值在 $[0, 1]$; $\text{sim}(m, n)$ 表示 m 和 n 之间的数值相似度, 且 $\text{sim}(m, n) \leq 1$ 。由相似度特性可知, 当 m 和 n 取值越接近时, $\text{sim}(m, n)$ 越接近于 1, 反之当两者差距越大时, $\text{sim}(m, n)$ 越接近 0。同时, 根据正向属性趋势的特点, 假设当 m 取值为两个与 n 距离相同的值 a 和 b ($a < n < b$ 且 $|an| = |nb|$) 时, m 等于 b 时的相似度比 m 等于 a 时的值要大。因此可以得到如图 6 所示的坐标轴关系。

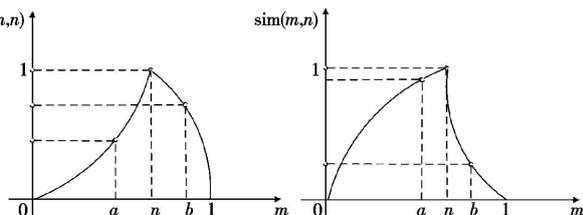


图6 坐标轴 (1)

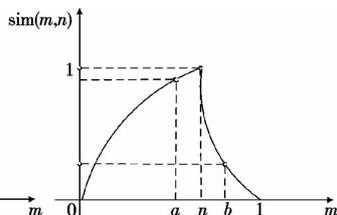


图7 坐标轴 (2)

根据图 6、7 的性质, 本文参考文献 [11] 中提到的仿真函数

数来进行定义。令 $\alpha = \min\left\{\frac{1}{n^2}, \frac{1}{(n-1)^2}\right\}$, 由于 QoS 参数的属性趋势只能为正向或反向中的一种, 则有:

正向 QoS 属性趋势为

$$\text{sim}_Z(m, n) = \begin{cases} \frac{m}{n} + \alpha n(m-n) & 0 \leq m \leq n \text{ 且 } n \neq 0 \\ \frac{m-1}{n-1} - \alpha(m-1)(m-n) & n \leq m \leq 1 \text{ 且 } n \neq 1 \end{cases} \quad (11)$$

反向 QoS 属性趋势为

$$\text{sim}_F(m, n) = \begin{cases} \frac{m}{n} - \alpha n(m-n) & 0 \leq m \leq n \text{ 且 } n \neq 0 \\ \frac{m-1}{n-1} + \alpha(m-1)(m-n) & n \leq m \leq 1 \text{ 且 } n \neq 1 \end{cases} \quad (12)$$

在 Web 服务推荐过程中, 用户 QoS 值与 Web 服务提供商提供的 Web 服务 QoS 值一一进行匹配时, 满足“正向趋势属性值越大越好, 负向趋势属性值越小越好”的基本原则, 因此采用式 (11) (12) 计算得到的数值相似度能够有效表明相似度越大的服务越符合用户的 QoS 需求。

2) 区间型 QoS 参数的匹配

由于区间型 QoS 参数的取值特点, 在对区间型 QoS 参数进行匹配时必须考虑区间的上界和下界, 即最大值和最小值。假设 Q_1 和 Q_2 表示两个区间型的 QoS 参数, 且 m 和 n 为 Q_1 的上界和下界, s 和 t 为 Q_2 的上界和下界, 即 Q_1 和 Q_2 可以分别用 $[m, n]$ 和 $[s, t]$ 度量表示。当且仅当 $m = s$ 且 $n = t$ 时, 两个区间才算是完全重合, Q_1 和 Q_2 的数值相似度为 1, 否则两个 QoS 参数的数值相似度将会随着两个区间差异的不断扩大而不断减小。假定 $\text{sim}([m, n], [s, t])$ 表示两个参数 Q_1 和 Q_2 的区间值 $[m, n]$ 和 $[s, t]$ 的数值相似度, $\text{sim}([m, s])$ 表示区间最小值之间的数值相似度, $\text{sim}([n, t])$ 表示区间最大值之间的数值相似度, 则 Q_1 和 Q_2 的数值相似度描述为

$$\text{sim}([m, n], [s, t]) = \sqrt{\frac{\text{sim}([m, s])^2 + \text{sim}([n, t])^2}{2}} \quad (13)$$

式 (13) 表明 Q_1 和 Q_2 的数值相似度由两区间的顶点值确定, 并且很好地阐释了两个区间之间的相近程度能够反映出对应 QoS 参数数值相似度的事实。

3) 等级型 QoS 参数的匹配

等级型 QoS 参数的值可以表示为等级范围 $[1, 2, 3, \dots, n]$ 中的其中一个, 其相当于是个模糊化的单值, 在经过标准化定量后取值范围也为 $[0, 1]$ 。两个等级型的 QoS 参数之间的数值匹配可以简单看成是模糊的单值匹配, 因此可以采用上述单值型 QoS 参数的匹配方法得以实现。

3 基于兴趣度的 Web 服务资源匹配模块

以往基于兴趣度的研究一般都是通过获取用户的消费偏好来分析的。而本文中的兴趣度指的是用户有可能对某 Web 服务资源产生消费冲动的程度, 是通过综合考虑用户对 Web 服务供应商的印象值、Web 服务供应商的名誉值以及用户接受风险的程度三方面来进行研究的。

在典型的商业环境中, 每一个用户会根据自身与服务供应商的互动而对其产生不同的印象。同样地, 在 Web 服务过程中, 只有当服务供应商之前提供过令用户满意的产品服务或是有过成功的服务推荐, 才能让用户对其产生较好的印象。本文参照 Chiu 等人 [12] 将用户 i 对于服务供应商 j 的印象值 IMP_{ij} 定义为

$$\text{IMP}_{ij} = f_i(\sum \text{gain}_{ij}, \sum \text{loss}_{ij}, s_{ij}, n_{ij}) \quad (14)$$

其中 $\sum \text{gain}_{ij}$ 是用户 i 过去在服务供应商 j 对其做的服务推荐中所获得的效用总和; $\sum \text{loss}_{ij}$ 指的则是用户 i 过去在服务供应商 j 对其做的服务推荐中所获得的损失总和; s_{ij} 表示服务供应商 j 在过去对用户 i 所做的服务推荐中成功次数的总和; 而 n_{ij} 表示用户 i 从服务供应商 j 那里获得服务推荐的总次数。根据实际消费经验和消费者的消费心态, 用户印象值 IMP_{ij} 需满足以下

条件:

a) $IMP_{ij} \in [-1, 1]$, -1 为最差印象值, 1 为最佳印象值; 用户的初始印象值设为 0 。

b) 当服务供应商推荐服务的成功率上升时, 用户印象值也随之增加。

c) 当 $\sum gain_{ij}$ 增加, 用户印象值也相应增加; 相反地, 当 $\sum loss_{ij}$ 增加, 则用户印象值相应减小。

d) 如果 $\sum gain_{ij} \geq \sum loss_{ij}$, 那么当 $s_{ij} = n_{ij} - s_{ij}$ 时, $IMP_{ij} = 0$; 当 $s_{ij} > n_{ij} - s_{ij}$ 时, $IMP_{ij} > 0$; 当 $s_{ij} < n_{ij} - s_{ij}$ 时, $IMP_{ij} < 0$ 。

e) 如果 $\sum gain_{ij} < \sum loss_{ij}$, 那么 $IMP_{ij} < 0$ 。

为满足上述条件, 本文引入以下仿真函数作为用户印象值的计算方法:

$$IMP_{ij} = \begin{cases} 0 & n=0 \\ \frac{s-(n-s)}{n} & \sum gain_{ij} = \sum loss_{ij} \\ -\left(\frac{s-(n-s)}{n}\right)\left(\frac{\sum gain_{ij}-\sum loss_{ij}}{\sum gain_{ij}+\sum loss_{ij}}\right) & \sum gain_{ij} < \sum loss_{ij} \text{ 并且 } s \leq n-s \\ \left(\frac{s-(n-s)}{n}\right)\left(\frac{\sum gain_{ij}-\sum loss_{ij}}{\sum gain_{ij}+\sum loss_{ij}}\right) & \text{其他情况} \end{cases} \quad (15)$$

Web 服务供应商的名誉值是其所有用户对其印象值的综合度量, 是服务供应商一个重要的社会象征。每一个用户的诚信以及是否对印象值准确度度量将会影响到名誉值的计算。本文引入了权重的概念, 即针对不同的用户印象值赋予不同的百分比信任度。则服务提供商 j 名誉值 REP_j 的计算式为

$$REP_j = \frac{\sum w_{ij} \times IMP_{ij}}{N} \quad (16)$$

其中: w_{ij} 表示用户 i 对服务供应商 j 印象值的可信度, 取值在 $[-1, 1]$ 之间; N 是所有用户总数。

不同的消费者在面对风险时的态度是不一样的。当某个 Web 服务供应商提供的服务失败时, 有些用户会坚决不再相信该服务供应商并终止重复消费, 而有些用户则可能再给一次机会, 这就意味着不同的用户之间存在着不同的风险接受程度。用户根据对风险的接受程度不同可以分成风险偏好型、风险逃避型以及风险中立型, 这种面对风险所采取的消费态度会大大影响用户的消费习惯。因此, 在对 Web 服务进行推荐的过程中, 考虑用户对于风险接受程度的高低也是十分必要的。本文将用户 i 的风险接受程度设为 $r_i, r_i \in [0, 1]$ 。当 $r_i = 0$ 时, 表明用户面对风险采取的是逃避的方式; 而当 $r_i = 1$ 时表明用户是风险寻求型的消费态度。

根据上述对于印象值、名誉值和风险接受度的阐述, 本文提出用户兴趣度函数 $interest_{ij} = f(IMP_{ij}, REP_j, r_i)$, 其中, IMP_{ij} 、 REP_j 和 r_i 分别为用户 i 对服务供应商 j 的印象值、服务供应商 j 的名誉值以及用户 i 的风险接受度。同样, 联系实际消费习惯和用户的消费心理, $interest_{ij}$ 也必须满足以下条件:

- a) 当印象值减少时, 用户兴趣度也相应减少, 否则反之;
- b) 当服务供应商的名誉值减少时, 用户兴趣度也相应减少, 否则反之;
- c) 当用户接受风险程度降低时, 用户兴趣度也相应减少, 否则反之。

因此, 用户兴趣度的计算式可以表示为

$$interest_{ij} = \frac{2r_i \times IMP_{ij} + REP_j}{3} \quad (17)$$

4 结束语

本文针对目前服务资源推荐技术存在的不足, 结合 Web

服务资源的特点, 提出了一种以用户需求为中心的 Web 服务资源智能推荐算法。该算法重点围绕适合度匹配和兴趣度挖掘两方面展开, 集中考虑了 Web 服务资源的功能属性 FQoS 和非功能属性 QoS, 并引入兴趣度概念来综合计算用户需求服务与候选服务群之间的多层次相似度, 旨在将最大匹配度的 Web 服务推荐给目标用户。

本文对于 Web 服务智能推荐研究还处于初级阶段, 在算法设计以及实现机制上还有待深入研究。本文在 QoS 适合度匹配计算时忽略了 QoS 的单位换算机制, 从而导致最终得到的结果可能不够精准。例如, Web 服务 a 的价格为 1 元, Web 服务 b 的价格为 10 角, 实际情况两者的价格是一样的, 但是在本文的相似度计算中却只考虑了 1 和 10 这种数值, 并没有覆盖单位的讨论, 因而致使最终得到的结果为服务 b 优于服务 a。另外, 不同的 QoS 属性对于不同用户在 Web 服务资源推荐的不同阶段中所起的影响程度是不一样的, 应该在计算过程中根据不同情况设置 QoS 影响权重。而本文还没有对这方面进行仔细的分析。在以后的工作中将着手研究这方面的问题, 以进一步优化算法, 提高 Web 服务推荐的精确度。

参考文献:

- [1] THIO N, KARUNASEKERA S. Web service recommendation based on client-side performance estimation [C]//Proc of Australian Software Engineering Conference. Washington DC:IEEE Computer Society, 2007:81-89.
- [2] LIAN Wen-juan, LIANG Yong-quan, ZENG Qing-tian, et al. An agent-based framework for personalized Web service recommendation [C]//Proc of the 2nd International Workshop on Computer Science and Engineering. Washington DC:IEEE Computer Society, 2009:229-233.
- [3] 朱红康, 余雪丽. 基于服务 QoS 执行信息的 Web 服务推荐研究 [J]. 计算机工程与应用, 2010, 46(18): 237-239.
- [4] ZHENG Zi-bin, MA Hao, LYU M R. WSRec: a collaborative filtering based Web service recommender system [C]//Proc of IEEE International Conference on Web Services. Washington DC:IEEE Computer Society, 2009: 437-444.
- [5] 蔡华利, 刘鲁, 樊坤, 等. 基于 BPSO 的 Web 服务推荐策略 [J]. 深圳大学学报:理工版, 2010, 27(1): 49-55.
- [6] ZADEH L A. The concept of a linguistic variable and its application to approximate reasoning [J]. Information Science, 1975, 8(3): 199-249.
- [7] 聂规划, 左秀然, 陈冬林. 本体映射中一种改进的概念相似度计算方法 [J]. 计算机应用, 2008, 28(6): 1563-1565.
- [8] JEONG B. Machine learning-based semantic similarity measures to assist discovery and reuse of data exchange XML schemas [D]. Pohang: Pohang University of Science and Technology, 2006.
- [9] 孙霞, 程宏斌. 基于加权层次结构的 XML 文档相似度算法 [J]. 武汉理工大学学报, 2009, 31(18): 76-79.
- [10] YATES R B. Modern information retrieval [M]. Harlow: Addison Wesley Longman Limited, 1999: 27-29.
- [11] 刘斌. 基于 QoS 本体的语义 Web 服务选择研究 [D]. 北京: 北京邮电大学, 2008.
- [12] CHIU D K W, LEUHG Ho-fung, LAM Ka-man. On the making of service recommendations: an action theory based on utility, reputation and risk attitude [J]. Expert System with Applications, 2009, 36(2): 3293-3301.