



基于用户和服务区域信息的 个性化 Web 服务质量预测

鲁城华^{1,2}, 寇纪淞¹

1 天津大学 管理与经济学部, 天津 300072

2 天津财经大学 珠江学院, 天津 301811

摘要: 在基于Web服务质量的研究中,服务质量信息均被假设为已知值。然而,在实际中很多服务质量信息是未知的且具有个性化特征。已有研究在进行服务质量预测时只采用全局服务质量信息,未考虑用户和服务的局部区域信息对服务质量预测的影响。因此,综合全局和局部角度对服务质量进行个性化预测具有重要的理论价值和实践意义。

基于服务质量的个性化特征以及用户和服务两方面影响因素的视角,提出一种基于用户和服务区域信息的服务质量预测方法。将全局的服务质量信息和局部的区域信息相结合构建预测模型,采用随机梯度下降法对模型进行优化学习,最终能够得到满意的预测结果。实验采用通用的综合数据集WSDream-QoSDataSet2,运用Matlab R2010b中机器学习的相关软件包训练预测模型,分析模型中不同参数对服务质量预测结果的影响,并与NIMF方法和Colbar方法进行对比分析。

研究表明,考虑用户和服务区域因素有助于提高服务质量预测的准确性;服务所在的区域影响比用户区域影响稍大,主要是由于服务器端服务运行条件不同所致;已有服务质量信息的密度对未知服务质量的预测具有一定影响,密度越大,预测效果越好;与NIMF方法和Colbar方法对比,MAE和RMSE评价指标显示所提出的方法具有较高的服务质量预测准确性。同时,采用区域内平均值的方法计算新用户和新服务的服务质量信息,有效地解决了冷启动问题。

研究结果有助于提供较为准确的服务质量信息,对基于服务质量的服务选择、服务推荐和服务组合等研究工作具有重要的支持作用,也为面向服务计算的发展提供一些有益的现实启示。

关键词: Web服务;服务质量;区域信息;QoS预测;个性化预测

中图分类号: C931

文献标识码: A

doi: 10.3969/j.issn.1672-0334.2020.02.005

文章编号: 1672-0334(2020)02-0063-13

引言

随着面向服务计算的普及,涌现大量Web服务,基于服务质量(quality of service, QoS)的研究成为热点^[1-2],如服务选择^[3]、服务推荐^[4]和服务组合等^[5]。因此,提供准确的QoS信息是所有基于QoS研

究工作的前提。

QoS信息包括两种类型,第1种独立于用户,其值由服务提供者统一提供,如服务价格^[6]。第2种QoS信息与用户和服务的环境相关,如响应时间^[7],其值容易受到异步的用户端网络环境、拓扑结构以及服

收稿日期: 2017-12-31 **修返日期:** 2018-04-06

基金项目: 国家自然科学基金(71631003)

作者简介: 鲁城华,天津大学管理与经济学部博士研究生、天津财经大学珠江学院讲师,研究方向为管理信息系统和数据挖掘等, E-mail: iamlichenghua@sina.com

寇纪淞,管理学博士,天津大学管理与经济学部教授,研究方向为信息系统与信息管理、数据挖掘和知识管理等,代表性学术成果为“Dynamics of fitness sharing evolutionary algorithms for coevolution of multiple species”,发表在2010年第3期《Applied Soft Computing》, E-mail: jskou@tju.edu.cn

务器端设备水平和传输效率的影响,测量QoS信息需要用户对服务调用之后才能得到。

面对QoS信息的千变万化,获得准确并具有个性化的QoS信息非常重要^[8]。然而,要求每个用户调用所有服务并返回准确的QoS信息是不现实的。一方面,Web服务数量庞大,这项工作耗时巨大;另一方面,用户并非专家,难以保证用户评估QoS信息的准确性。因此,有必要找到一种有效的QoS预测方法而无需进行繁琐的服务调用^[9]。

已有研究采用协同过滤方法对历史QoS信息进行分析并实现预测^[10],其核心是使用相似用户或服务的QoS信息预测未知QoS信息,已经被证实具有良好的预测效果^[11]。传统的协同过滤方法仅使用已有的QoS信息,忽略了一些客观因素的影响,如用户和服务所在国家、省市地区等区域因素。同一地区的用户和服务具有相似的经济水平,使其具有相似的硬件环境。因此,在同一区域的用户调用相同的服务通常产生相似的QoS信息。为了提高QoS预测的准确性,本研究将全局QoS信息与用户和服务区域信息相结合,提出一种基于用户和服务区域信息的QoS预测方法。

1 相关研究评述

近年来,互联网上涌现出越来越多具有相同或者相似功能的Web服务^[12],因此,QoS属性成为大众关注的热点^[13]。同时,很多基于QoS的研究工作逐渐发展起来,如服务选择^[14-15]、服务推荐^[16]、服务组合^[17-18]和服务信任^[19]等。因此,为这些研究工作提供准确的QoS信息是极其重要的。

目前对QoS预测的研究还处于起步阶段^[20]。最简单的预测方法是平均值方法,该方法包括用户平均值方法和服务平均值方法,前者是以计算该用户使用的其他服务的QoS信息平均值作为该用户对当前未使用服务的QoS信息;后者是以计算所有使用该服务的用户的QoS信息平均值作为当前用户使用该服务的QoS信息。平均值方法较为简单,只是计算QoS信息的平均值,这种方法不能准确地反映出不同用户QoS之间的个性化差异^[21]。本研究将对未知QoS进行个性化的预测,以提高预测的准确性。

受用户协作思想的启发,协同过滤方法被应用到QoS的预测中^[22]。协同过滤方法对领域知识没有要求^[23],因此被广泛应用于不同的领域。通常将协同过滤方法分为基于记忆^[24]和基于模型两种类型^[25]。

基于记忆的协同过滤方法主要包括基于用户的计算、基于项目的计算以及两者的混合^[26]。基于记忆的QoS预测方法是根据用户使用Web服务之后所反馈的QoS信息计算两个用户(或服务)之间的相似程度,进而选择相似程度比较高的一组用户(或服务)的QoS信息预测未知的QoS值。SHAO et al.^[27]提出使用基于用户的协同过滤方法进行个性化的QoS预测,以用户的历史使用经验(即使用服务之后反馈的QoS信

息)为基础,采用皮尔逊相关系数的相似性度量方法计算两个用户之间的相似性,然后根据当前用户的邻居用户集合对其未使用服务的QoS信息进行预测;DING et al.^[28]提出一种基于服务相似性计算的QoS预测方法,然后将QoS属性值的预测与用户满意度相结合,评价服务的可信度。该研究改进了传统的皮尔逊相关系数的相似性度量方法,降低了消极QoS属性对预测结果的影响,提高了QoS预测的准确性。ZHENG et al.^[11]提出一种混合的基于用户和服务的协同过滤方法预测未知的QoS信息,该方法获得了较好的预测效果;SILIC et al.^[29]也提出一种混合的协同过滤方法预测服务的可靠性,主要根据用户的地理位置和服务的地理位置(根据IP地址区分)、服务负载和服务级别(根据服务的计算消耗区分)4个方面的内容。用这4个方面的线性组合预测服务可靠性,一方面,服务提供者可能不愿意分享服务的负载和计算消耗值;另一方面,该方法只考虑了局部影响而忽略了用户间和服务间的全局关系。基于记忆的协同预测方法仅从用户或者服务角度采用已有的QoS信息进行预测,忽略了位置信息等一些客观因素,且当计算规模增大时,该方法的计算效率有所降低。本研究结合用户和服务的位置信息,从全局和局部两个角度对未知QoS进行高效的个性化预测。

随着Web服务规模的不断扩大,大部分研究人员开始采用一种更高效的预测方法,即基于模型的预测。最具有代表性的模型方法是矩阵因子分解法^[30],该方法已经被证明能够有效、准确地处理大规模的服务数据集。该方法将用户和服务分别映射到一个低维的特征空间,用两者的乘积代表原来的QoS矩阵。将每一个用户和服务表示成一个低维潜在的特征向量(或者叫潜在因子向量)^[31],将特征向量里的每一个特征视为关于服务或者用户兴趣偏好的任何一个方面或者其他隐含的特征。当矩阵被分解完成之后,根据得到的两个低维矩阵乘积与原QoS矩阵的拟合得到一个最小化的目标函数,该函数的目标就是使预测的QoS值尽量接近真实已知的QoS值^[32]。经过函数的优化学习,最终得到准确性比较高的QoS预测矩阵。ZHENG et al.^[25]采用矩阵因子分解的方法,将用户和服务的QoS矩阵分解为两个低维的用户特征矩阵和服务特征矩阵,然后结合当前用户的邻居用户集合的QoS信息预测未知的QoS信息,该方法能够处理较大的QoS数据集,且具有较高的计算效率;YIN et al.^[33]在矩阵因子分解模型中引入用户的位置信息作为正则项,根据位置信息对用户进行筛选,计算满足一定距离区间内的用户之间的相似程度,根据相似性排序得到当前用户的邻居用户集合,对建立的函数模型进行优化学习,得到QoS预测矩阵。该方法提高了QoS预测的效率和准确率,但没有考虑服务的位置信息。

虽然传统的协同过滤方法得到了较好的QoS预测结果,但仍存在一定弊端。大部分研究仅采用已有的QoS信息进行预测,忽略了用户和服务的位置信

息等一些客观因素的影响。即使一小部分研究考虑了位置信息,但它们没有将全局的QoS信息影响与局部的区域影响相结合。相比之下,本研究将全局的QoS信息与用户服务的局部区域信息相结合进行预测,采用综合的服务数据集,证明本研究方法具有有效性。

2 问题描述

本研究的目的是将历史QoS信息与用户和服务的区域信息相结合,进行个性化的QoS信息预测。在实际应用中,可能存在大量的用户和服务。假设用户数量为 M ,服务数量为 N ,用户使用服务的QoS信息可以表示为 $M \times N$ 的矩阵 R ,矩阵中每个元素 $r_{i,j}$ 为 u_i 用户调用 s_j 服务的QoS信息, i 为用户的序号值, $i=1,2,\dots,f,\dots,M$, j 为服务的序号值, $j=1,2,\dots,k,\dots,N$ 。未被用户使用的服务的QoS信息为空,即要预测的未知值。本研究的目标是创建一个预测模型,有效而准确地填充这些未知值。

为了便于理解,图1给出一个现实世界中Web服务调用场景,该场景能够充分说明考虑用户和服务区域信息的必要性。假设 u_a 用户分别调用来自两个不同国家的两个功能相同的Web服务 s_2 和 s_3 ,由于这两个国家的经济发展水平存在很大的差异,导致其服务器端的设备水平和网络传输效率差异很大。因此, u_a 用户使用这两个服务的QoS感知也差异很大, s_2 服务的响应时间为0.010s, s_3 服务的响应时间为1s,相差100倍。同样地,对于同一个服务 s_1 ,如果同时被来自两个不同国家的 u_b 用户和 u_c 用户调用,由于受其国家发展水平不同的影响, u_b 用户的IT设施和网络拓扑结构比 u_c 用户的高端,导致其感知服务的QoS水平不同。 u_b 用户接收 s_1 服务的响应时间为0.020s, u_c 用户接收 s_1 服务的响应时间为0.400s,相差20倍。因此,考虑用户和服务所在区域(国家或者省市地区)的影响对于提高QoS预测的准确性具有重要意义。

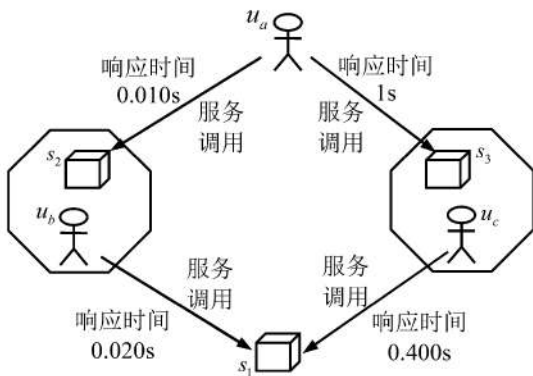


图1 Web服务调用场景

Figure 1 Scenario of Web Service Invocation

3 用户相似度和服务相似度的计算

由于本研究考虑了区域信息,因此,有必要计算

区域内用户或服务之间的相似度,即根据QoS历史信息计算用户或服务之间的相似度。相似用户主要是指两个用户在同一组服务上具有比较相似的QoS历史信息,相似服务主要是指两个服务被同一组用户使用具有比较相似的QoS历史信息^[34]。对于新用户和新服务,由于缺少QoS信息,本研究采用区域内平均值的方法计算其相应的QoS信息。该方法选取同一区域内的所有其他用户和服务的QoS信息均值作为新用户和新服务的QoS信息,从而能够有效地处理QoS预测中面临的冷启动问题。

不同的QoS属性,其数据类型可能不同,如响应时间和成本属于数值型QoS,可靠性和可用性属于比率型QoS。对于比率型的QoS信息,其变化范围是有限的,如变化范围为 $[0,100\%]$ 。然而,对于数值型的QoS信息,其可能存在不同的数值范围,比较典型的数值型属性是响应时间,比如一个用户使用服务的响应时间变化范围是 $[0s, 1s]$,而另一个用户的响应时间变化范围是 $[10s, 20s]$ 。

QoS信息取值范围的不同可能影响相似度的计算,因此,本研究采用高斯法^[15]对QoS信息进行规范化处理,将QoS信息映射到 $[0,1]$ 之间。计算公式为

$$r'_{i,j} = 0.5 + \frac{r_{i,j} - \bar{r}_i}{2 \times 3\sigma_i} \quad (1)$$

其中, $r'_{i,j}$ 为对 $r_{i,j}$ 进行规范化处理之后的QoS信息; \bar{r}_i 为 u_i 用户对所有服务反馈的QoS信息的平均值; σ_i 为 u_i 用户对所有服务反馈的QoS信息的标准差, $3\sigma_i$ 为3- σ 规则,在高斯法中,如果取 3σ ,则可以使99%的数据都映射到 $[0,1]$ 之间。对于经过规范化处理之后在 $[0,1]$ 范围之外的数值,若小于0赋值为0,大于1赋值为1。

对QoS信息进行规范化处理之后,本研究采用皮尔逊相关系数的相似性度量方法^[35]计算用户相似度和服务相似度,其取值范围为 $[0,1]$,该值越大,说明两个用户或服务之间的相似度越大。计算公式为

$$Sim(i,f) = \frac{\sum_{s \in S_{i,f}} (r_{i,s} - \bar{r}_i)(r_{f,s} - \bar{r}_f)}{\sqrt{\sum_{s \in S_{i,f}} (r_{i,s} - \bar{r}_i)^2} \sqrt{\sum_{s \in S_{i,f}} (r_{f,s} - \bar{r}_f)^2}} \quad (2)$$

$$Sim(j,k) = \frac{\sum_{u \in U_{j,k}} (r_{u,j} - \bar{r}_j)(r_{u,k} - \bar{r}_k)}{\sqrt{\sum_{u \in U_{j,k}} (r_{u,j} - \bar{r}_j)^2} \sqrt{\sum_{u \in U_{j,k}} (r_{u,k} - \bar{r}_k)^2}}$$

其中, $Sim(i,f)$ 为 u_i 用户与 u_f 用户之间的相似度, $S_{i,f}$ 为 u_i 用户和 u_f 用户共同使用过的Web服务集合, $r_{i,s}$ 为 u_i 用户对 s 服务反馈的QoS信息, $r_{f,s}$ 为 u_f 用户对 s 服务反馈的QoS信息, \bar{r}_i 为 u_i 用户对所有服务反馈的QoS信息的平均值, $Sim(j,k)$ 为 s_j 服务与 s_k 服务之间的相似度, $U_{j,k}$ 为共同使用过 s_j 服务和 s_k 服务的用户集合, $r_{u,j}$ 为 u 用户对 s_j 服务反馈的QoS信息, $r_{u,k}$ 为 u 用户对 s_k 服务反馈的QoS信息, \bar{r}_j 为所有用户对 s_j 服务反馈的QoS信息的平均值, \bar{r}_k 为所有用户对 s_k 服务反馈的QoS信息的平均值。

4 基于用户和服务区域信息的QoS预测过程

在已有研究中,只采用已有的QoS信息进行预测,忽略了一些客观因素影响,如用户和服务的位置信息。为了提高QoS预测的准确性,本研究将全局QoS信息与用户和服务区域信息相结合,提出一种基于用户和服务区域信息的QoS预测方法QPRIUS(QoS prediction based on region information of users and services)。首先,结合全局QoS信息与局部区域信息构建预测模型,然后采用随机梯度下降法对模型优化学习,最终得到预测结果。考虑用户和服务区域因素有助于提高QoS预测的准确性,采用区域内平均值方法计算新用户和新服务的QoS信息,这样能够有效解决冷启动问题。以下将给出QPRIUS方法的详细过程。

4.1 QoS预测框架

图2给出QPRIUS方法的框架。用户相似度和服务相似度的计算是QoS预测的前期准备工作,见图2中虚线框所示,以区域为单位分别计算同一区域内的用户相似度和服务相似度,得到两个相似度矩阵 Sim_U 和 Sim_S ,这两个矩阵是构造QoS预测模型的基础。

在QoS预测过程中, $M \times N$ 的用户-服务QoS矩阵

R 经过初始的矩阵因子分解,得到低维的用户特征矩阵 U 和服务特征矩阵 S 。分解过程为

$$R = [r_{i,j}]_{M \times N} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1N} \\ r_{21} & r_{22} & \dots & r_{2N} \\ \dots & \dots & \dots & \dots \\ r_{M1} & r_{M2} & \dots & r_{MN} \end{bmatrix} \quad (3)$$

矩阵因子分解为

$$U = [u_{i,j}]_{d \times M} = \begin{bmatrix} U_1 & U_2 & \dots & U_M \\ u_{11} & u_{12} & \dots & u_{1M} \\ u_{21} & u_{22} & \dots & u_{2M} \\ \dots & \dots & \dots & \dots \\ u_{d1} & u_{d2} & \dots & u_{dM} \end{bmatrix} \quad (4)$$

$$S = [s_{i,j}]_{d \times N} = \begin{bmatrix} S_1 & S_2 & \dots & S_N \\ s_{11} & s_{12} & \dots & s_{1N} \\ s_{21} & s_{22} & \dots & s_{2N} \\ \dots & \dots & \dots & \dots \\ s_{d1} & s_{d2} & \dots & s_{dN} \end{bmatrix} \quad (5)$$

$$U_i^T S_j \rightarrow r_{i,j} \xrightarrow{\text{优化}} U_i'^T S_j' \rightarrow r'_{i,j} \quad (6)$$

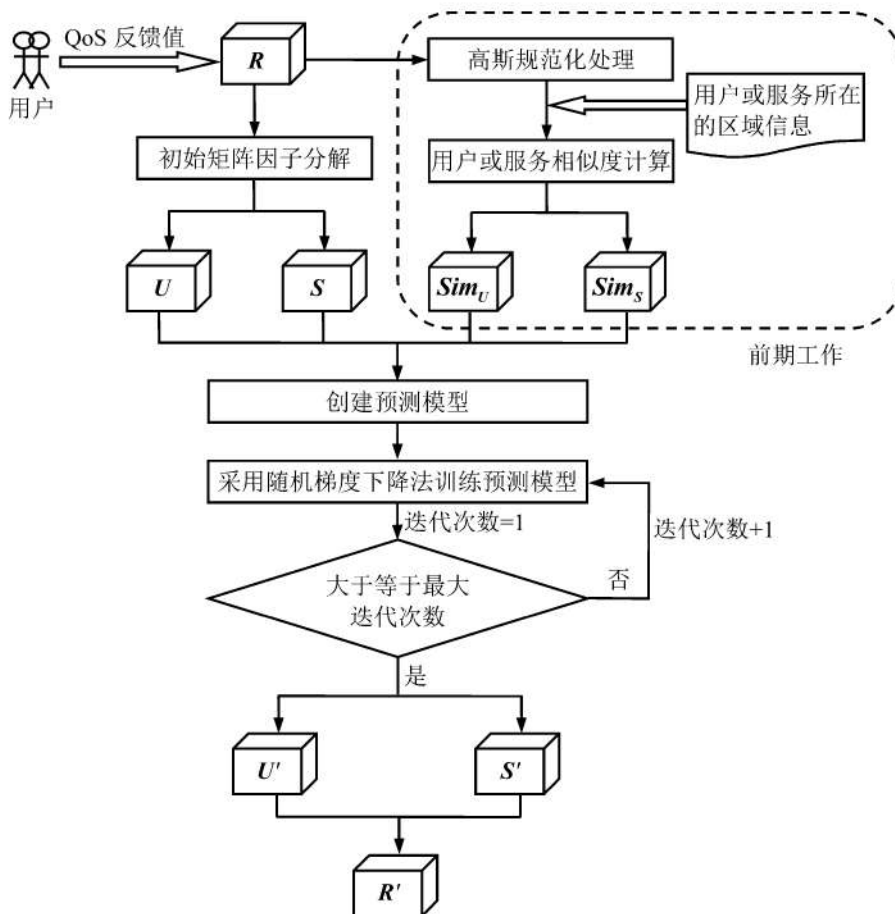


图2 QPRIUS方法框架
Figure 2 Architecture of QPRIUS

其中, U 为 $d \times M$ 的用户特征矩阵, 包含 M 个潜在的特征向量; S 为 $d \times N$ 的服务特征矩阵, 包含 N 个潜在的特征向量; d 为维度, 特征向量的长度 (即特征个数)。 U 中的每一列 U_i 为 u_i 用户的特征向量, S 中的每一列 S_j 为 s_j 服务的特征向量。

经过以上过程, QoS 预测模型由 U 、 S 、 Sim_U 和 Sim_S 共 4 部分组成。采用随机梯度下降法^[36] 对模型进行优化学习, 优化的目标就是使预测值尽量逼近真实的已知 QoS 信息。在模型的优化过程中, 特征矩阵 U 和 S 被不断更新, 当达到一定的迭代次数之后, 得到最终的用户特征矩阵 U' 和服务特征矩阵 S' , 可以用 $U_i'^T S_j'$ 预测任意未知的 QoS 信息, 上角标 T 为矩阵转置符号。

4.2 QPRIUS 方法的详细过程

创建 QoS 预测模型, 该模型结合了 QoS 历史信息以及用户和服务的区域信息, 然后对模型进行优化。

4.2.1 创建 QoS 预测模型

本研究预测模型中采用矩阵因子分解方法, 该方法基于降维技术。奇异值分解^[37] 是比较著名的降维技术, 能够将原始矩阵较好地分解为相应的低秩矩阵。

如前所述, 用户-服务矩阵 R 被分解为两个矩阵 U 和 S , 即 $R \approx U^T S$, $U \in \mathbb{R}^{d \times M}$, $S \in \mathbb{R}^{d \times N}$, $d \ll \min(M, N)$ 。基于此, 本研究的 QoS 预测模型 P 建立过程如 (7) 式, 该最小化函数的目的是使预测的 QoS 信息矩阵与已知的 QoS 信息矩阵之间的差值尽量小。

$$\begin{aligned} & \min P(R, U, S) \\ &= \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^N I_{i,j} (r_{i,j} - U_i^T S_j)^2 + \frac{\lambda_1}{2} \|U\|_F^2 + \frac{\lambda_2}{2} \|S\|_F^2 + \\ & \quad \frac{\alpha}{2} \sum_{i=1}^M \sum_{f \in A(i)^*} Sim(i, f) \|U_i - U_f\|_F^2 + \\ & \quad \frac{\beta}{2} \sum_{j=1}^N \sum_{k \in A(j)^*} Sim(j, k) \|S_j - S_k\|_F^2 \end{aligned} \quad (7)$$

$$A(i)^* = \{l_u \mid Sim(i, l_u) \geq \mu, l_u \in A(i)\}$$

$$A(j)^* = \{l_s \mid Sim(j, l_s) \geq \mu, l_s \in A(j)\}$$

其中, $I_{i,j}$ 为标识变量, 当 u_i 用户调用过 s_j 服务时 $I_{i,j} = 1$, 当 u_i 用户未调用过 s_j 服务时 $I_{i,j} = 0$; $\|\cdot\|_F^2$ 为正则项, 即 Frobenius 范数; λ_1 为用户特征矩阵正则项因子的权重系数, λ_2 为服务特征矩阵正则项因子的权重系数, λ_1 和 λ_2 都是正值, 为了减少模型的复杂度, 通常取 $\lambda_1 = \lambda_2 = \lambda$; α 为用户区域正则项的权重系数; β 为服务区域正则项的权重系数; μ 为相似度阈值; U_f 为 u_f 用户的特征向量; S_k 为 s_k 服务的特征向量; $A(i)^*$ 为 u_i 用户所在区域内相似度超过阈值 μ 的用户邻居集合, $A(i)$ 为 u_i 用户所在区域内的用户集合, $A(j)^*$ 为 s_j 服务所在区域内相似度超过阈值 μ 的服务邻居集合, $A(j)$ 为 s_j 服务所在区域内的服务集合; l_u 为 $A(i)$ 中的任意用户, l_s 为 $A(j)$ 中的任意服务。

根据正则化理论^[38], 4 个正则项被加入模型中, U 和 S 的正则项能够避免过拟合问题^[39], α 和 β 用于

调节用户和服务的区域因子在 QoS 预测过程中的影响程度。为了提高预测的准确性, 排除噪音的影响, 对于同一区域内的用户和服务分别计算, 得到其较高质量的邻居集合。设定一个 μ 值, 当两个用户 (或服务) 之间的相似度超过 μ 时, 才会对它们的特征向量差值进行 Frobenius 范数处理, 而不是对所有的用户进行处理, 这样有助于提高计算效率和准确率。加入区域正则项的目的是使同一区域内相似度较高的用户 (或服务) 之间的特征向量 (QoS 信息) 更加接近。也就是说, 当用户 (或服务) 之间的相似度较大时, 在优化过程中, 为了使目标函数值尽量小, 只能使特征向量差值的 Frobenius 范数 $\|U_i - U_f\|$ 和 $\|S_j - S_k\|$ 尽量小。因此, 就会缩小用户 (或服务) 特征向量之间的距离, 使其更加接近, 这样有助于提高 QoS 预测的准确性。

4.2.2 模型的优化

对 (7) 式进行训练优化, 在矩阵因子分解方法中, 最常用的优化方法是随机梯度下降法, 其基本思想是让目标函数中的变量 U_i 和 S_j 沿着函数负梯度的方向移动。由于矩阵因子分解的目标函数是凸函数^[33], 因此, 负梯度的方向就是朝着函数值减少的方向移动, 直到移动到函数的极小值点。以用户为例, 根据 MA et al.^[40] 的研究, 在计算用户特征向量的偏导数过程中, 用户邻居集合 $A(i)^*$ 内的用户彼此互为邻居。因此, 在对区域正则项计算偏导数时相当于对特征向量 U_i 计算两次; 同理, 服务特征向量 S_j 偏导数的计算也是如此。计算得到的偏导数为

$$U_{i+1} = U_i - \eta_U \frac{\partial P}{\partial U_i}$$

$$S_{j+1} = S_j - \eta_S \frac{\partial P}{\partial S_j}$$

$$\begin{aligned} \frac{\partial P}{\partial U_i} = & \sum_{j=1}^N I_{i,j} (r_{i,j} - U_i^T S_j) (-S_j) + \lambda_1 U_i + \\ & 2\alpha \sum_{f \in A(i)^*} Sim(i, f) (U_i - U_f) \end{aligned}$$

$$\begin{aligned} \frac{\partial P}{\partial S_j} = & \sum_{i=1}^M I_{i,j} (r_{i,j} - U_i^T S_j) (-U_i) + \lambda_2 S_j + \\ & 2\beta \sum_{k \in A(j)^*} Sim(j, k) (S_j - S_k) \end{aligned} \quad (8)$$

其中, U_{i+1} 为 U_i 经过一次训练之后得到的矩阵; S_{j+1} 为 S_j 经过一次训练之后得到的矩阵; η_U 为矩阵 U 的学习步长, η_S 为矩阵 S 的学习步长, η_U 和 η_S 也是学习速率, 均为正值; P 为 (7) 式的预测模型的简写。

在训练过程中, 由于 (8) 式中的负值部分可能使特征矩阵中出现负值, 而在矩阵的更新过程中负值无法正确表示用户和服务的特征信息, 保持矩阵中每一个特征项的非负性具有非常重要的意义。因此, 本研究将使用非负矩阵因子分解的方法预测 QoS 信息。以特征矩阵 U_i 的训练过程为例, 具体步骤为

$$U_{i+1} = U_i - \eta_U \frac{\partial P}{\partial U_i}$$

$$\begin{aligned}
&= U_i - \eta_U \left[\sum_{j=1}^N I_{i,j} (r_{i,j} - U_i^T S_j) (-S_j) + \lambda_1 U_i + \right. \\
&\quad \left. 2\alpha \sum_{f \in A(i)^*} \text{Sim}(i,f) (U_i - U_f) \right] \\
&= U_i + \eta_U \sum_{j=1}^N I_{i,j} r_{i,j} S_j - \eta_U \left[\sum_{j=1}^N I_{i,j} (U_i^T S_j) S_j + \right. \\
&\quad \left. \lambda_1 U_i + 2\alpha \sum_{f \in A(i)^*} \text{Sim}(i,f) (U_i - U_f) \right] \quad (9)
\end{aligned}$$

为了保持矩阵的非负性,将 η_U 调整为^[41]

$$\eta_U = \frac{U_i}{\sum_{j=1}^N I_{i,j} (U_i^T S_j) S_j + \lambda_1 U_i + 2\alpha \sum_{f \in A(i)^*} \text{Sim}(i,f) (U_i - U_f)} \quad (10)$$

这个转化过程能够抵消(9)式中的负值部分,因此,将(10)式代入(9)式,训练过程转换为

$$U_{i+1} = \frac{U_i \sum_{j=1}^N I_{i,j} r_{i,j} S_j}{\sum_{j=1}^N I_{i,j} (U_i^T S_j) S_j + \lambda_1 U_i + 2\alpha \sum_{f \in A(i)^*} \text{Sim}(i,f) (U_i - U_f)} \quad (11)$$

同样地, η_S 被调整为(12)式,特征矩阵 S_j 的训练过程见(13)式。

$$\eta_S = \frac{S_j}{\sum_{i=1}^M I_{i,j} (U_i^T S_j) U_i + \lambda_2 S_j + 2\beta \sum_{k \in A(j)^*} \text{Sim}(j,k) (S_j - S_k)} \quad (12)$$

$$S_{j+1} = \frac{S_j \sum_{i=1}^M I_{i,j} r_{i,j} U_i}{\sum_{i=1}^M I_{i,j} (U_i^T S_j) U_i + \lambda_2 S_j + 2\beta \sum_{k \in A(j)^*} \text{Sim}(j,k) (S_j - S_k)} \quad (13)$$

在QoS预测模型的优化过程中,变量的偏导数是梯度最陡的方向,也就是说,函数模型在朝着目标值下降最快的方向运行。随机梯度下降算法的迭代终止条件比较多,常见的主要有3种情况:

(1)设置固定的迭代次数;

(2)给目标函数设定一个固定的阈值,当目标函数的值小于阈值时,停止迭代过程;

(3)给目标函数值的变化范围设定一个阈值,当前后两次目标函数值变化的绝对值小于阈值时,停止迭代过程。

本研究采用第(1)种设置方法,根据实验效果设置一定的迭代次数。

经过(11)式和(13)式的迭代优化过程,最终可以得到新的用户特征矩阵 U' 和服务特征矩阵 S' ,从而计算得到QoS信息的预测矩阵 $R' \approx U'^T S'$ 。

4.3 复杂性分析

QPRIUS方法的计算复杂性主要存在于预测模型的优化过程中,即用户特征矩阵 U 和服务特征矩阵 S

的更新过程。由(3)式~(6)式可知,QoS矩阵经过矩阵因子分解方法得到两个 d 维的特征矩阵 U 和 S 。由(11)式和(13)式可以看出,其计算复杂性与QoS矩阵密度 MD (即矩阵 R 中QoS信息不为0的数据个数)、用户数量 M 、服务数量 N 、矩阵的维度 d 以及用户和服务的邻居数量相关,假设 l_u^* 为满足相似度要求的用户的平均邻居数, l_s^* 为满足相似度要求的服务的平均邻居数量。

以用户特征矩阵 U 的更新过程为例,根据(11)式,矩阵 U 的每个列向量 U_i 乘上分式

$$\frac{\sum_{j=1}^N I_{i,j} r_{i,j} S_j}{\sum_{j=1}^N I_{i,j} (U_i^T S_j) S_j + \lambda_1 U_i + 2\alpha \sum_{f \in A(i)^*} \text{Sim}(i,f) (U_i - U_f)}$$

即可得到更新后的列向量 U_{i+1} 。 U_i 为已知量,分式的

分子部分 $\sum_{j=1}^N I_{i,j} r_{i,j} S_j$ 中 $r_{i,j}$ 为QoS矩阵的值,变量 $I_{i,j}$ 标识 u_i 用户是否调用过 s_j 服务, S_j 为 d 维列向量,因此,该部分的计算复杂度 $O(\cdot)$ 与QoS矩阵密度 MD 、维度 d 相关,即 $O(MD \cdot d)$ 。同理,分母中 $\sum_{j=1}^N I_{i,j} (U_i^T S_j) S_j$ 的计算复杂度也是 $O(MD \cdot d)$, $\lambda_1 U_i$ 的计算复杂度为常数阶 $O(1)$; $2\alpha \sum_{f \in A(i)^*} \text{Sim}(i,f) (U_i - U_f)$ 的计算复杂度与用户数量 M 、用户平均邻居数量 l_u^* 、维度 d 相关,即 $O(2 \cdot M \cdot l_u^* \cdot d)$ 。综合来看,分式整体的计算复杂度为 $O(MD \cdot d + 2 \cdot M \cdot l_u^* \cdot d)$ 。同样,对于服务特征矩阵 S 的更新过程(13)式,其计算复杂度为 $O(MD \cdot d + 2 \cdot N \cdot l_s^* \cdot d)$ 。

通常情况下,用户和服务的邻居个数远小于用户和服务本身的个数,并且存在 $M \ll MD, N \ll MD$,因此,迭代一次时整个过程的复杂度可以用 $O(MD \cdot d)$ 表示。 d 值越大,在模型训练过程中需要更多的时间成本,不利于实际应用。因此,本研究将 d 统一设为10。可以发现,在 d 值一定的条件下,QPRIUS方法的计算复杂度与矩阵密度线性相关。因此,本研究使用的方法具有较高的计算效率,并且能够处理大规模的数据集。

5 实验和分析

本研究通过实验验证QoS预测方法QPRIUS,实验使用Matlab R2010b编写程序,运行环境为Windows XP,双核CPU2.4GHz,2G内存。

实验主要分为两部分内容:第1部分对方法中所含的参数进行实验分析,第2部分将本研究的QoS预测方法与其他研究进行对比和分析。参数实验分析主要包括QoS预测模型中的特征矩阵权重系数 λ_1 和 λ_2 、用户区域权重系数 α 、服务区域权重系数 β 、矩阵密度 MD 和相似度阈值 μ 。通常情况下取 $\lambda_1 = \lambda_2$,因此,在以下实验中统称为 λ 。通过参数实验,可以分析预测模型中不同的参数取值对QoS预测效果的影响,同时发现较优的参数值。在QoS预测方法的对比实验中,将QPRIUS方法与其他方法进行对比分析。

5.1 实验设计

5.1.1 数据集

本研究使用的实验数据为 WSDream-QoSDataSet2^[42], 见表1。由表1可知, 该数据包含1 974 675条339个用户对5 825个Web服务调用之后反馈的QoS信息, 包括响应时间和吞吐量两个属性。此外, 每个用户和服务都有一个区域信息, 即所在国家名称。

表1 WSDream-QoSDataSet2数据集
Table 1 Data Set of WSDream-QoSDataSet2

统计信息	数值或取值范围
用户数量	339
服务数量	5 825
响应时间	0 ~ 20s
吞吐量	1 ~ 1 000kbps

在实验中, 本研究在QoS矩阵中随机去掉一部分数据, 从而生成不同密度的矩阵进行实验。例如, 当MD的取值为20%时, 也就是说保留原始QoS矩阵中20%的值作为训练数据, 其余80%的数据作为测试数据, 即作为判断预测准确性的期望值。本实验仅对响应时间和吞吐量两个属性进行研究, 对于其他的QoS属性本方法同样适用。

5.1.2 评价指标

本研究采用两个常用的评价指标测量QoS预测方法的准确性, 分别是平均绝对误差(MAE)和均方根误差(RMSE)^[43], 计算公式为

$$MAE = \frac{\sum_{i,j} |\tilde{r}_{i,j} - \hat{r}_{i,j}|}{n_p} \quad (14)$$

$$RMSE = \sqrt{\frac{\sum_{i,j} (\tilde{r}_{i,j} - \hat{r}_{i,j})^2}{n_p}}$$

其中, $\tilde{r}_{i,j}$ 为QoS的实际值, $\hat{r}_{i,j}$ 为QoS的预测值, n_p 为被

预测的QoS个数。MAE指标反映了预测值与实际值之间的平均误差, RMSE指标对较大误差的反应比较灵敏, 用来测试误差程度的大小, MAE值和RMSE值越小, 预测准确性越高。

5.2 参数分析实验

本研究对QoS预测方法中涉及的5个参数进行实验分析, 包括 λ 、 α 、 β 、MD和 μ 。以下实验以响应时间数据为例, 分别对这些参数的影响进行分析, 并选取恰当的值。对于其他的QoS属性, 可采用同样的方法进行分析。

5.2.1 参数 λ 的影响分析

本研究在预测模型中引入用户和服务特征矩阵正则项, 一方面可以防止模型的过度拟合, 另一方面从全局QoS角度对模型优化进行调节。本实验讨论模型中权重系数 λ 对实验结果的影响, 先不考虑用户和服务的局部区域信息, 即 α 和 β 取值为0, MD = 50%, $d = 10$, $\mu = 0.400$ 。本实验将 λ 的取值分别设定为1E-5、1E-4、1E-3、1E-2、1E-1和1E0, 进行300次迭代优化过程, 计算MAE值和RMSE值, 取10次结果的平均值, 实验结果见图3。

由图3可知, 随着 λ 增大, MAE和RMSE的变化趋势基本一致。当 $\lambda = 1E-3$ 时, MAE值为0.361, RMSE值为1.019, 均为最低值, 此时的预测效果最优。实验结果表明, λ 取值大小对模型的预测效果具有一定影响。如果取值过小, 正则项因子的范数在模型优化过程中获得较大值, 使对已知数据的逼近误差越来越小, 容易形成过度拟合, 增大了对预测数据的逼近误差。也就是说, 当 λ 取值过小时, 正则项因子的加入失去了调节作用; 如果 λ 取值过大, 导致正则项因子的范数越来越低, 增大了对已知数据的逼近误差, 最终影响未知数据的预测结果。对本研究来说, 当 λ 取1E-3时, 实验效果最优, 在后面的实验中将使用此值。

5.2.2 参数 α 的影响分析

参数 α 决定用户区域因素在QoS预测中所占的权重, 如果 α 特别小, 在QoS预测中使用用户-服务QoS矩阵; 如果 α 特别大, 用户的区域信息控制整个预测过程, 影响预测准确性。本实验讨论 α 的取值对预测

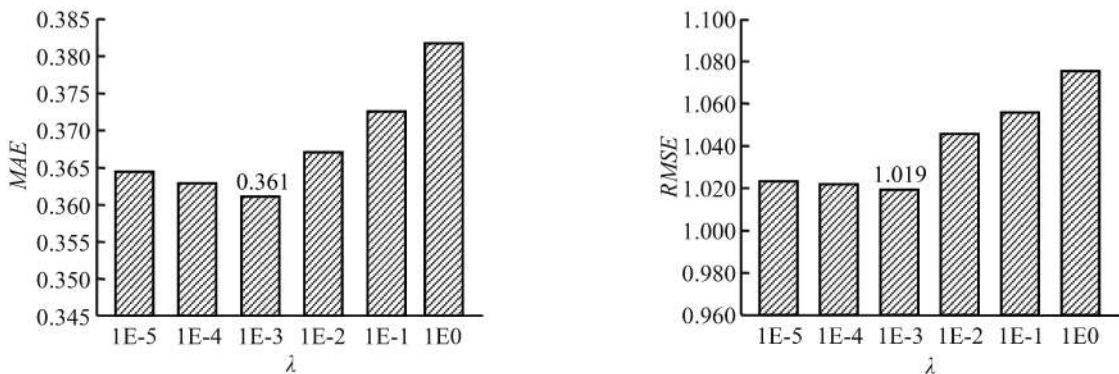
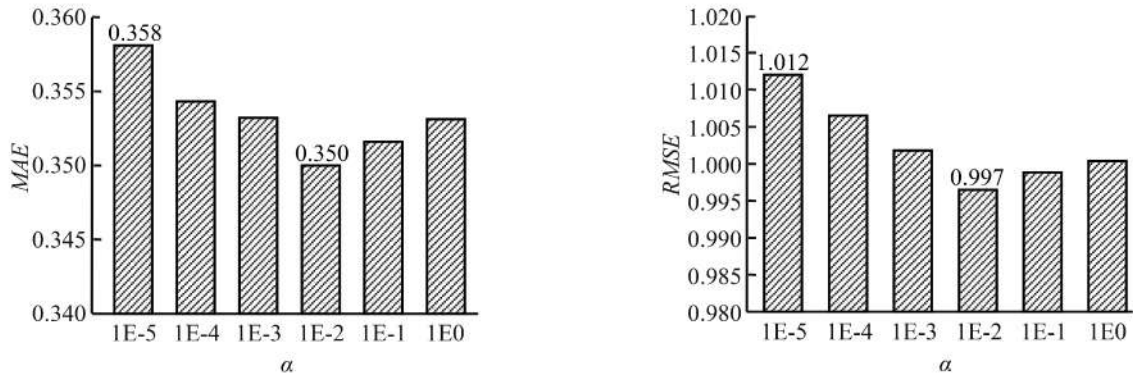
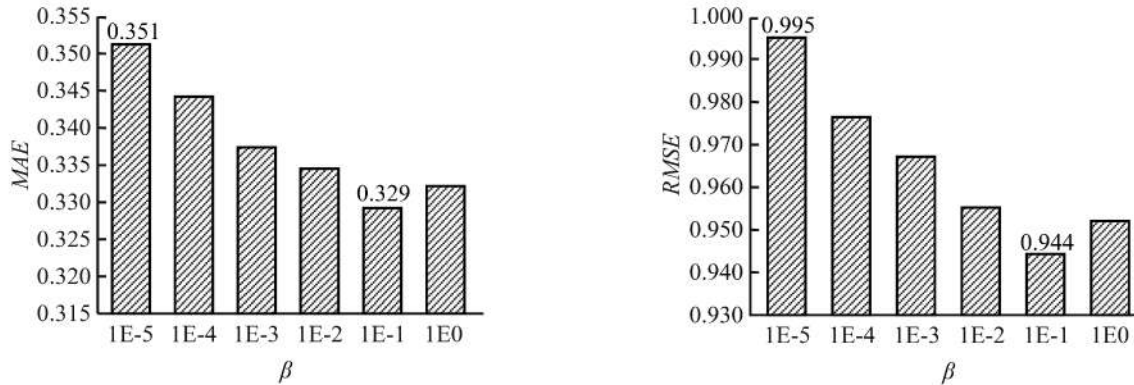


图3 不同 λ 取值下的实验结果

Figure 3 Experimental Results under Various λ

图4 不同 α 取值下的实验结果Figure 4 Experimental Results under Various α 图5 不同 β 取值下的实验结果Figure 5 Experimental Results under Various β

结果的影响,将 α 的取值分别设为1E-5、1E-4、1E-3、1E-2、1E-1和1E0, $\lambda = 1E-3$, $\beta = 0$, $MD = 50\%$, $d = 10$, $\mu = 0.400$,计算10次结果的MAE和RMSE平均值,计算结果见图4。

由图4可知,当 $\alpha = 1E-5$ 时,MAE值为0.358, RMSE值为1.012,比图3中 $\lambda = 1E-3$ 时的最优MAE值和RMSE值小。说明加入用户区域信息后,模型的预测效果有所提高。当 $\alpha = 1E-2$ 时,MAE值为0.350, RMSE值为0.997,均为最小值,此时的预测准确性最佳。实验结果表明,加入用户局部区域因素有助于提升QoS预测效果。若用户区域因素在预测模型中的权重过大,会过度陷入局部因素,忽略全局影响,降低预测准确性。 α 取1E-2时预测效果最优,后面的实验将采用此值。

5.2.3 参数 β 的影响分析

与参数 α 类似, β 决定服务区域因素在QoS预测中所占的权重。本实验讨论 β 取值变化对QoS预测效果的影响,将 β 值分别设为1E-5、1E-4、1E-3、1E-2、1E-1和1E0, $\lambda = 1E-3$, $\alpha = 0$, $MD = 50\%$, $d = 10$, $\mu = 0.400$,实验运行10次取平均值,计算结果见图5。

由图5可知, $\beta = 1E-1$ 时,MAE值为0.329, RMSE值为0.944,优于图3中MAE和RMSE的最小值,说明考虑

服务的区域因素有助于提高QoS预测效果。当 $\beta = 1E-1$ 时,MAE值为0.329, RMSE值为0.944,均为最小值。实验结果表明,加入服务区域信息有助于提升QoS预测效果,但是当 β 取值过大时,预测模型同样会陷入局部影响。当 $\beta = 1E-1$ 时预测效果最佳,后面的实验将采用此值。

与5.2.2的实验结果对比可以发现,在同样的权值下,相对于用户的局部区域信息,服务的局部区域信息对QoS预测的影响更大。通常情况下,当用户向Web服务注册中心提交服务请求后,注册中心根据用户请求查找满足其需求的服务并提供相应的绑定信息。用户直接与服务提供者进行交互并调用服务,服务在提供者的服务器端运行并提供用户需要的信息,用户通过电脑接收服务结果。服务器在整个服务运行中起主导作用,而用户端只负责接收运行结果。因此,服务所在的区域发展水平对其QoS信息具有较大的影响,实验结果充分说明了这一点。

5.2.4 矩阵密度的影响分析

在实际中,不可能每个用户都调用所有服务并反馈QoS信息,因此,QoS矩阵往往存在一定的稀疏度,即矩阵密度MD。QoS信息越多,矩阵密度越大,反之则越小。本实验将MD分别设为10%、20%、30%、

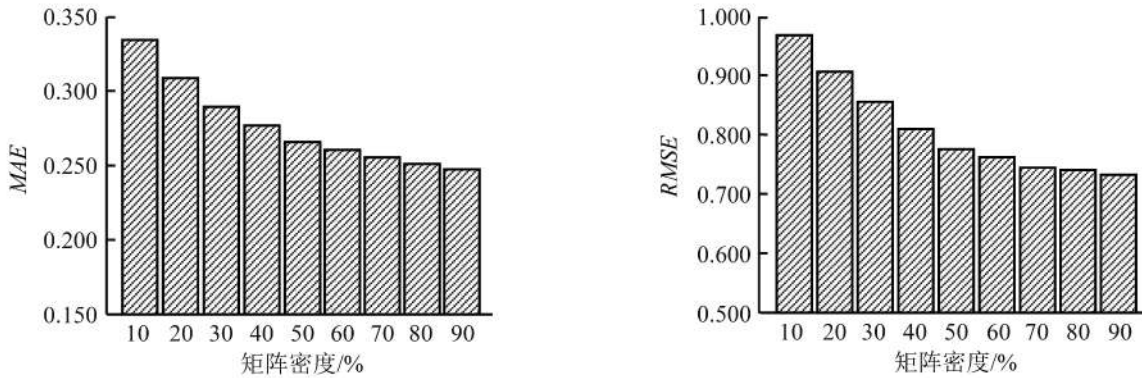


图 6 不同矩阵密度下的实验结果
Figure 6 Experimental Results under Various Matrix Density

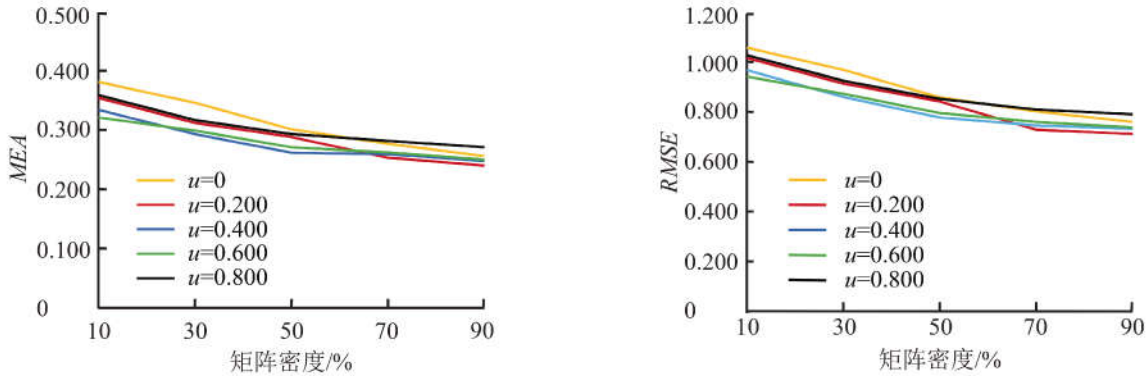


图 7 不同相似度阈值时的实验结果
Figure 7 Experimental Results under Various Similarity Threshold

40%、50%、60%、70%、80%和90%， $\lambda = 1E-3, \alpha = 1E-2, \beta = 1E-1, d = 10, \mu = 0.400$ ，实验运行10次取平均值，计算结果见图6。

由图6可知，随着矩阵密度的增加，预测效果越来越好。当MD取值从10%增加到50%时，MAE和RMSE下降趋势较为明显，随着MD增大，MAE和RMSE下降趋势相对较缓。实验结果说明已知的QoS信息越多，其对未知QoS信息预测的准确性帮助越大。矩阵密度在由10%逐渐增加到50%时，其对预测准确性的影响比较明显，进一步说明鼓励用户尽量多地提供QoS信息反馈值非常重要。

5.2.5 相似度阈值的影响分析

为了获取当前用户或服务更高质量的邻居集合，本研究设定一个相似度阈值对相似用户或服务进行过滤。本实验讨论在不同的矩阵密度下阈值 μ 的不同取值对预测结果的影响， μ 的取值分别设定为 0、0.200、0.400、0.600 和 0.800，MD 取值为 10%、30%、50%、70%和90%， $\lambda = 1E-3, \alpha = 1E-2, \beta = 1E-1, d = 10$ ，实验运行10次取平均值，计算结果见图7。

由图7可知，当 μ 固定时，随着MD增加，MAE和RMSE逐渐下降，这一结果与5.2.4实验结果相吻合。此外，在不同的MD下，对应不同的最优阈值 μ 。例

如，当MD = 10% 时， $\mu = 0.600$ 的预测效果比较好；当MD取值为30%和50%时， $\mu = 0.400$ 的预测效果比较好；当MD取值为70%和90%时，预测效果在 $\mu = 0.200$ 时比较好。

可以发现，随着MD增大， μ 越小对预测效果的准确性越有利；MD越小， μ 则越大越好。说明当QoS信息比较少时，用户或服务之间的相似度结果存在较大噪音，只有取相似度比较大的邻居才有利于QoS信息的预测。较低的相似度并不能正确地描述和区分用户或服务之间的关系，进而影响预测结果。当MD较大时，具有丰富的QoS信息，因此，用户或服务之间的相似度计算能够比较恰当地描述用户或服务之间的关系，即使是较低的相似度也能比较准确地表示其关系。综上所述，当MD较小时，设置较大的阈值有利于提高预测的准确性；当MD较大时，设置较小的阈值对QoS预测比较有利。当MD = 50% 时，最优阈值为0.400，该值被应用于本研究的其他实验中。

5.3 与其他方法的对比分析

本研究将QoS预测方法QPRIUS与NIMF^[20]方法和Colbar^[25]方法进行对比。NIMF方法主要采用矩阵因子分解的方法，结合当前用户的邻居用户QoS信息对未知的QoS信息进行预测。Colbar方法在矩阵因子分

解模型中引入用户的位置信息作为正则项,根据位置信息对用户进行筛选,然后计算满足一定距离内的用户之间的相似度,得到当前用户的邻居集,再对建立的模型进行学习,得到QoS预测矩阵。

图8和图9给出本研究方法与NIMF和Colbar两种方法在不同矩阵密度下的对比。NIMF方法没有考虑区域信息的影响,Colbar方法只考虑了用户的位置信息而没有考虑服务的位置信息的影响。因此,将本研究只考虑用户区域信息的方法QPRIU($\alpha \neq 0, \beta = 0$)与Colbar方法进行对比,然后再将同时考虑用户和服务区域因素时的方法QPRIUS与Colbar方法和NIMF方法进行对比。

根据申利民等^[20]和ZHENG et al.^[25]的原有计算结果,Colbar方法中计算了MD取值为10%、20%、30%和40%的预测结果,NIMF方法中计算了MD取值为5%、10%、15%和20%的预测结果。为了方便对比,本研究将矩阵密度设定为5%、10%、15%、20%、30%和40%,计算响应时间和吞吐量两个QoS属性的MAE和RMSE值,取10次结果的平均值,并与NIMF方法和Colbar方法分别进行对比和分析。

首先,分析当本研究仅考虑用户区域信息时的QPRIU方法与Colbar方法的对比。由图8和图9可知,

在同样的矩阵密度条件下,对于响应时间和吞吐量两个属性来说,本研究的QPRIU方法计算得到的MAE和RMSE均比Colbar方法小。说明在同样考虑用户区域信息的情况下,本研究的计算方法比Colbar方法的预测准确度更高一些。主要是因为本研究模型在优化过程中保持了矩阵的非负性,并且是根据区域划分对用户进行分类并计算相似度,避免了将两个相邻区域的用户因为地理位置很近但是硬件设施差别很大而被认为是互为邻居的情况。由于Colbar方法无法正确处理这种情况,从而导致其计算结果存在一定误差。

其次,当本研究同时考虑用户和服务区域信息时与NIMF方法和Colbar方法进行对比。由图8和图9可知,在同样的矩阵密度条件下,对于响应时间和吞吐量两个属性来说,QPRIUS方法预测结果明显优于NIMF方法和Colbar方法。该实验结果充分说明本研究方法的优越性和QoS预测的准确性,以及充分考虑用户区域信息和服务区域信息的必要性。

6 结论

本研究提出一种基于用户和服务区域信息的Web服务质量预测方法QPRIUS。在所创建的QoS预

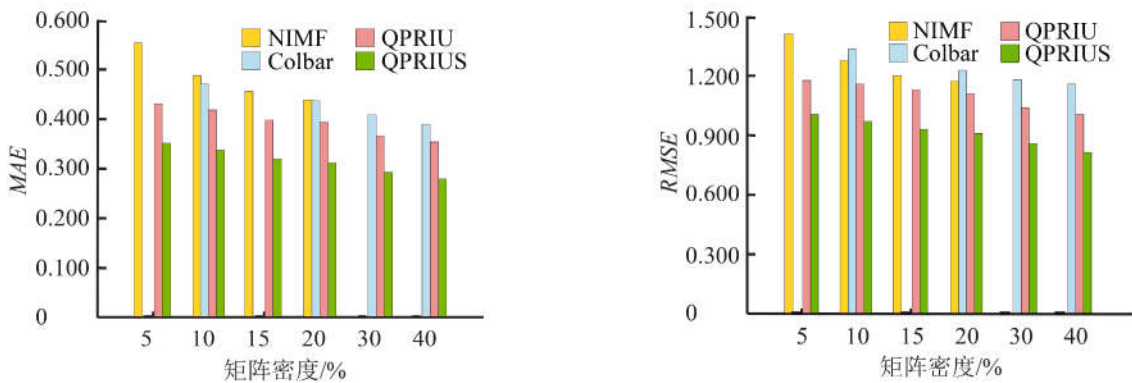


图8 4种方法对比——响应时间

Figure 8 Comparisons of Four Methods for Response Time

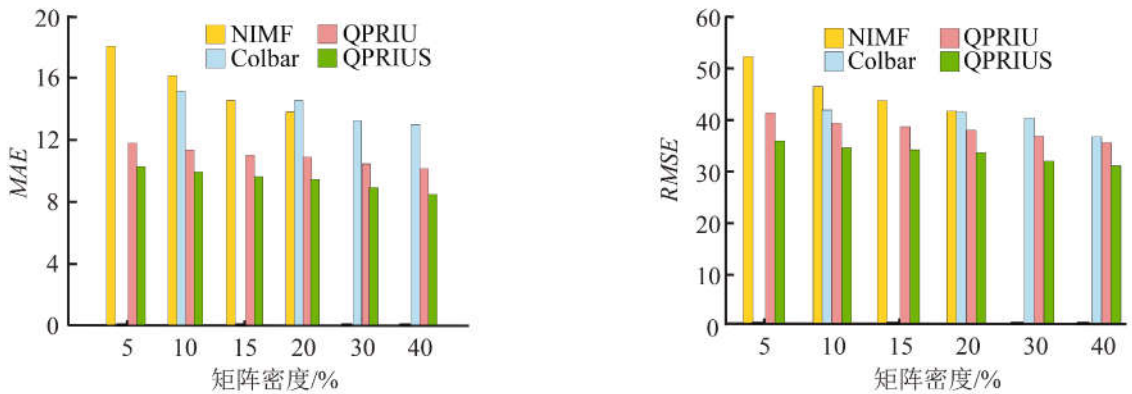


图9 4种方法对比——吞吐量

Figure 9 Comparisons of Four Methods for Throughput

测模型中,不仅考虑全局的QoS信息,而且加入用户和服务的局部区域信息,通过模型的迭代优化过程,最终得到准确性较高的QoS预测矩阵。研究结果表明,①在QoS预测模型中加入用户和服务的区域影响因素有助于提高QoS预测的准确性。相较而言,服务所在的区域情况对QoS信息的预测效果影响较大。主要是由于当用户提交其服务请求之后,服务在提供者的服务器端运行并提供给用户需要的信息。因此,服务器端在服务运行过程中起主导作用,其所在区域的发展水平对服务运行的QoS信息影响很大。②已有QoS信息的密度对QoS预测效果具有一定影响,密度越大,预测准确性越高,说明鼓励用户尽量多地提供QoS信息反馈值非常重要。③为了提高模型的优化效率,获取高质量的邻居集合,在计算区域内用户之间和服务之间相似度时引入相似度阈值,当QoS信息的密度较小时,较大的 μ 值有利于提高预测的准确性;当QoS信息的密度较大时,较低的 μ 值对QoS信息的预测比较有利。

本研究具有较高的QoS预测准确性,与其他方法对比,准确性提高了25%。同时,本研究有效地解决了新用户和新服务的冷启动问题。对于较稀疏的矩阵,也能够得到较好的预测效果。研究结果对大量基于QoS的研究工作具有重要的支持作用,有利于为其提供较为准确的QoS信息,为面向服务计算的发展提供一些有益的启示。

本研究还存在一些不足可以在后续研究予以拓展。在今后的研究中可以考虑用户对服务的满意度和信任度等主观因素对QoS预测的影响,这将有利于进一步提高QoS预测的准确性。此外,本研究针对静态的QoS信息,有些QoS属性是动态变化的,需要对其进行实时预测,这有待于今后的进一步研究。

参考文献:

- [1] 李金忠,曾小荟,夏洁武,等.基于QoS的Web服务选择方法研究新进展.《计算机应用研究》,2015,32(8):2245-2253.
LI Jinzhong, ZENG Xiaohui, XIA Jiewu, et al. Recent advances in approaches of Web service selection based on QoS. *Application Research of Computers*, 2015, 32(8): 2245-2253.
- [2] 申文果,张秀娟,谢礼珊.网络企业服务质量的测量及其影响的实证研究.《管理科学》,2007,20(1):38-45.
SHEN Wenguo, ZHANG Xiujuan, XIE Lishan. Empirical research of measurement and impact of e-service quality. *Journal of Management Science*, 2007, 20(1): 38-45.
- [3] 马力,邱志洋,陈彦萍,等.基于QoS的语义Web服务选择.《计算机科学》,2017,44(3):226-230,246.
MA Li, QIU Zhiyang, CHEN Yanping, et al. Semantic Web service selection based on QoS. *Computer Science*, 2017, 44(3): 226-230, 246.
- [4] CHEN X, ZHENG Z B, LIU X D, et al. Personalized QoS-aware Web service recommendation and visualization. *IEEE Transactions on Services Computing*, 2013, 6(1): 35-47.
- [5] 欧阳超,陈志泊,孙国栋. Web服务组合QoS优化中的改进遗传算法.《计算机工程》,2017,43(8):231-235,242.
OUYANG Chao, CHEN Zhibo, SUN Guodong. Improved genetic algorithm for Web service composition QoS optimization. *Computer Engineering*, 2017, 43(8): 231-235, 242.
- [6] WANG X Y, ZHU J K, SHEN Y H. Network-aware QoS prediction for service composition using geolocation. *IEEE Transactions on Services Computing*, 2015, 8(4): 630-643.
- [7] MARDUKHI F, NEMATBAKHS N, ZAMANIFAR K, et al. QoS decomposition for service composition using genetic algorithm. *Applied Soft Computing*, 2013, 13(7): 3409-3421.
- [8] 唐瑞春,方丽华,庞敏,等. QoS预测机制的个性化云媒体服务选择.《北京邮电大学学报》,2016,39(6):42-46.
TANG Ruichun, FANG Lihua, PANG Min, et al. Personalized cloud-media services selection by QoS prediction. *Journal of Beijing University of Posts and Telecommunications*, 2016, 39(6): 42-46.
- [9] 任迪,万健,殷昱煜,等.基于贝叶斯分类的Web服务质量预测方法研究.《浙江大学学报(工学版)》,2017,51(6):1242-1251.
REN Di, WAN Jian, YIN Yuyu, et al. Web services QoS prediction method based on Bayes classification. *Journal of Zhejiang University (Engineering Science)*, 2017, 51(6): 1242-1251.
- [10] WU J, JI L C, LIANG T T, et al. Topic model based collaborative QoS prediction. *Applied Mathematics & Information Sciences*, 2014, 8(5): 2545-2555.
- [11] ZHENG Z B, MA H, LYU M R, et al. QoS-aware Web service recommendation by collaborative filtering. *IEEE Transactions on Services Computing*, 2011, 4(2): 140-152.
- [12] 徐文庭,殷昱煜,王菊仙,等.基于CART与SlopeOne的服务质量预测算法.《计算机集成制造系统》,2017,23(5):1080-1090.
XU Wenting, YIN Yuyu, WANG Juxian, et al. QoS prediction based on CART and SlopeOne. *Computer Integrated Manufacturing Systems*, 2017, 23(5): 1080-1090.
- [13] 李欣,于渤.服务质量评价特征及服务补救策略.《管理科学》,2004,17(3):72-75,76.
LI Xin, YU Bo. The characteristics of service quality evaluation and tactics of service recovery. *Journal of Management Science*, 2004, 17(3): 72-75, 76.
- [14] WANG P, CHAO K M, LO C C. On optimal decision for QoS-aware composite service selection. *Expert Systems with Applications*, 2010, 37(1): 440-449.
- [15] WANG S G, HSU C H, LIANG Z J, et al. Multi-user web service selection based on multi-QoS prediction. *Information Systems Frontiers*, 2014, 16(1): 143-152.
- [16] XU Y S, YIN J W, DENG S G, et al. Context-aware QoS prediction for Web service recommendation and selection. *Expert Systems with Applications*, 2016, 53: 75-86.
- [17] SUN S X, ZHAO J. A decomposition-based approach for service composition with global QoS guarantees. *Information Sciences*, 2012, 199(15): 138-153.
- [18] LIU Z Z, JIA Z P, XUE X, et al. Reliable Web service composition based on QoS dynamic prediction. *Soft Computing*, 2015, 19(5): 1409-1425.
- [19] SKOGRUD H, BENATALLAH B, CASATI F. Model-driven

- trust negotiation for Web services. *IEEE Internet Computing*, 2003, 7(6): 45-52.
- [20] 申利民, 陈真, 李峰. 考虑数据变化范围的Web服务服务质量协同预测方法. *计算机集成制造系统*, 2017, 23(1): 215-224.
SHEN Limin, CHEN Zhen, LI Feng. Web service QoS collaborative prediction approach considering range of QoS data. *Computer Integrated Manufacturing Systems*, 2017, 23(1): 215-224.
- [21] 江晓苏, 魏延, 邱炳发. QoS感知的Web服务个性化推荐. *计算机技术与发展*, 2015, 25(12): 85-90.
JIANG Xiaosu, WEI Yan, QIU Bingfa. QoS-aware Web services personalized recommendations. *Computer Technology and Development*, 2015, 25(12): 85-90.
- [22] 俞东进, 殷昱煜, 吴萌萌, 等. 基于混合协同过滤的Web服务QoS预测方法. *浙江大学学报(工学版)*, 2014, 48(11): 2039-2045, 2060.
YU Dongjin, YIN Yuyu, WU Mengmeng, et al. QoS prediction for Web services based on hybrid collaborative filtering. *Journal of Zhejiang University (Engineering Science)*, 2014, 48(11): 2039-2045, 2060.
- [23] KOREN Y, BELL R. Advances in collaborative filtering // RICCI F, ROKACH L, SHAPIRA B, et al. *Recommender Systems Handbook*. New York: Springer, 2011: 145-186.
- [24] WU J, CHEN L, FENG Y P, et al. Predicting quality of service for selection by neighborhood-based collaborative filtering. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2013, 43(2): 428-439.
- [25] ZHENG Z B, MA H, LYU M R, et al. Collaborative Web service QoS prediction via neighborhood integrated matrix factorization. *IEEE Transactions on Services Computing*, 2013, 6(3): 289-299.
- [26] TANG M D, ZHANG T T, LIU J X, et al. Cloud service QoS prediction via exploiting collaborative filtering and location-based data smoothing. *Concurrency and Computation: Practice & Experience*, 2015, 27(18): 5826-5839.
- [27] SHAO L S, ZHANG J, WEI Y, et al. Personalized QoS prediction for Web services via collaborative filtering // *IEEE International Conference on Web Services*. IEEE, 2007: 439-446.
- [28] DING S, YANG S L, ZHANG Y T, et al. Combining QoS prediction and customer satisfaction estimation to solve cloud service trustworthiness evaluation problems. *Knowledge - Based Systems*, 2014, 56: 216-225.
- [29] SILIC M, DELAC G, KRKA I, et al. Scalable and accurate prediction of availability of atomic Web services. *IEEE Transactions on Services Computing*, 2014, 7(2): 252-264.
- [30] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems. *Computer*, 2009, 42(8): 30-37.
- [31] 叶卫根. 基于协同过滤的个性化推荐算法研究. 江苏: 江南大学, 2016: 15-16.
YE Weigen. *Research on personalized recommendation algorithms based on collaborative filtering*. Jiangsu: Jiangnan University, 2016: 15-16.
- [32] LUO X, ZHOU M C, XIA Y N, et al. An efficient non-negative matrix-factorization based approach to collaborative filtering for recommender systems. *IEEE Transactions on Industrial Informatics*, 2014, 10(2): 1273-1284.
- [33] YIN J W, LO W, DENG S G, et al. Colbar: a collaborative location-based regularization framework for QoS prediction. *Information Sciences*, 2014, 265(5): 68-84.
- [34] DUAN L J, TIAN H. Collaborative Web service discovery and recommendation based on social link. *Future Internet*, 2017, 9(4): 1-12.
- [35] 苏凯, 马良荔, 刘霞, 等. 一种用于服务质量预测的混合协同过滤方法. *武汉理工大学学报(信息与管理工程版)*, 2015, 37(2): 207-211.
SU Kai, MA Liangli, LIU Xia, et al. A hybrid collaborative filtering approach for service QoS prediction. *Journal of Wuhan University of Technology (Information & Management Engineering)*, 2015, 37(2): 207-211.
- [36] KOREN Y. Collaborative filtering with temporal dynamics. *Communications of the ACM*, 2010, 53(4): 89-97.
- [37] 潘善亮, 张迎新, 俞晓锋. 基于OWL-S的Web服务匹配研究. *计算机工程*, 2010, 36(12): 48-51.
PAN Shanliang, ZHANG Yingxin, YU Xiaofeng. Research on Web service matching based on OWL-S. *Computer Engineering*, 2010, 36(12): 48-51.
- [38] BOYD S, VANDENBERGHE L. *Convex optimization*. Cambridge, UK: Cambridge University Press, 2004: 302-318.
- [39] JAIN A K, DUIN R P W, MAO J C. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2000, 22(1): 4-37.
- [40] MA H, ZHOU D Y, LIU C, et al. Recommender systems with social regularization // *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, 2011: 287-296.
- [41] BERRY M W, BROWNE M, LANGVILLE A N, et al. Algorithms and applications for approximate nonnegative matrix factorization. *Computational Statistics & Data Analysis*, 2007, 52(1): 155-173.
- [42] 黎懋靛. 基于协同过滤和QoS预测的Web服务推荐方法研究. 重庆: 重庆大学, 2015: 41-42.
LI Maoliang. *Research on Web service recommendation algorithms based on collaborative filtering and QoS prediction*. Chongqing: Chongqing University, 2015: 41-42.
- [43] 李国盛, 王娜. 一种轻量级的Web服务QoS预测机制. *计算机应用研究*, 2016, 33(11): 3311-3314, 3333.
LI Guosheng, WANG Na. Lightweight prediction mechanism for Web service QoS. *Application Research of Computers*, 2016, 33(11): 3311-3314, 3333.

Personalized QoS Prediction for Web Services Based on the Region Information of Users and Services

LU Chenghua^{1,2}, KOU Jisong¹

1 College of Management and Economics, Tianjin University, Tianjin 300072, China

2 College of Pearl River, Tianjin University of Finance and Economics, Tianjin 301811, China

Abstract: Concerning the QoS-based studies, the QoS information is all assumed to be known. However, in practical application, much QoS information is unknown and have personalized characters. Previous studies have just employed the global QoS information to make predictions and do not consider the impact of user/service local region information on the QoS values. Therefore, it is of important theoretical value and practical significance to make personalized QoS prediction from both global and local perspectives.

From the perspective of personalized QoS and the influence from users and services, this study proposes a novel QoS prediction approach based on the region information of users and services (QPRIUS). First, a prediction model is established by taking advantage of both the global QoS information and the local region information of users and services. Then the model is optimized iteratively by means of stochastic gradient descent. Finally, a satisfactory prediction result can be obtained. Experiments utilize the synthetic data WSDream-QoSDataSet2 and employ the related software package of machine learning in Matlab R2010b to train the prediction model. Different parameters in the model are analyzed about their effects on the QoS prediction accuracy. Furthermore, this study is compared with the methods of NIMF and Colbar.

The experimental results indicate that it is beneficial to improve the accuracy of QoS prediction by considering the region of users and services. Parameter experiments show that the effect of service region is larger than that of user region because there is some difference in the running condition for services in the server-side. The density of existing QoS information has certain influence on prediction of unknown QoS. The larger the density is, the better the prediction accuracy will be. In the comparative experiments with the methods of NIMF and Colbar, the *MAE* and *RMSE* criteria indicate that the proposed method achieves high QoS prediction accuracy. This method uses the intra-regional mean one to compute the QoS information of new users and services and can effectively solve the cold start problem.

The findings contribute to provide more accurate QoS information and have important supporting action for some QoS-based studies such as service selection, service recommendation, service composition and so on. Moreover, the findings provide some reality enlightenment for the development of service-oriented computing.

Keywords: Web service; quality of service; region information; QoS prediction; personalized prediction

Received Date: December 31st, 2017 **Accepted Date:** April 6th, 2018

Funded Project: Supported by the National Natural Science Foundation of China(71631003)

Biography: LU Chenghua is a Ph. D candidate in the College of Management and Economics at Tianjin University and is a lecturer in the College of Pearl River at Tianjin University of Finance and Economics. Her research interests include management information system and data mining. E-mail: iamluchenghua@sina.com

KOU Jisong, doctor in management, is a professor in the College of Management and Economics at Tianjin University. His research interests cover information system and information management, data mining, and knowledge management. His representative paper titled "Dynamics of fitness sharing evolutionary algorithms for coevolution of multiple species" was published in the *Applied Soft Computing* (Issue 3, 2010). E-mail: jskou@tju.edu.cn

□