

一种基于全域子空间分解挖掘的 QoS 准确预测方法

张博雅 胡晓辉

(兰州交通大学电子与信息工程学院 兰州 730070)

摘要 QoS 的准确预测是评判和选择最佳 Web 服务的一种重要标准;传统的 QoS 预测方法采用时间平均值和各种参数简单加权的方法,无法对大量 Web 服务下的资源进行准确预测,预测结果模糊;提出一种基于全域子空间分解挖掘的 QoS 准确预测方法,即采用全域分析的思想对所有数据进行预处理,在此基础上,通过子空间分解的方法,在子空间中对数据进行分解分析,提取数据的深层次特征,然后将全域分析的结果与子空间分解分析的结果进行有效的数据融合,从而实现对所分析数据的准确预测;采用一组 Web 节点和拟定度量参数进行了预测实验,结果显示,基于全域子空间分解挖掘的 QoS 预测方法可以精确预测出渐变过程,结果准确,在 QoS 预测中具有广泛的应用价值。

关键词 QoS 预测,全域分析,子空间分解,数据挖掘

中图分类号 TP393 **文献标识码** A

Accurate Prediction Method of QoS Based on Global Subspace Decomposition Mining

ZHANG Bo-ya HU Xiao-hui

(School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract The accurate prediction of QoS is an important criterion for the Web services. In traditional QoS prediction method, when the time-weighted average and simple weight of various parameters are used, but a large number of Web services can not be accurately predicted, and the outcome is vague. An accurate prediction method of QoS based on global subspace decomposition mining was proposed, and the idea of global analysis of all data preprocessing was done, on this basis, through subspace methods, the subspace decomposition analysis of the data was carried out and the depth features of the data were extracted, then the global result of the analysis and subspace decomposition analysis results were fuzzed for effective data fusion and achieving the analytical data valid mining and predictive. A set of Web nodes and metrics were used to do the prediction experiment, and the result shows that with the global subspace decomposition mining method, the gradual process can be predicted with detailed process. The result is accurate and it has wide application value in QoS prediction.

Keywords QoS prediction, Global analysis, Subspace decomposition, Data mining

1 引言

随着网络上服务资源及服务数量的迅速增长,出现了很多功能相似或相同而服务质量(Quality of Service, QoS)不同的 Web 服务。QoS 已经成为评判和选择最佳候选服务的一种重要标准,它一般包括很多不同的非功能性属性,不同的领域关注的 QoS 不同,一般包括:时间和空间的可用性、渠道、安全性、可靠性、所有权等。网络上通过 QoS 预测方法实现对网络服务资源的开发、传输、储存过程的控制和管理,解决 Web 服务中的安全性、可靠性等,提高 Web 服务的高效性和智能化已经成为当今网络通信研究的热门话题之一^[1]。

目前关于 Web 服务的 QoS 预测方法研究处于初级阶段,为了提高服务质量预测的准确性,对传统网络服务资源及服务数据的处理传输过程中出现的丢包、网络堵塞、信息安全等现象进行有效的预防和控制。一般 Web 服务多采用 Int-

Serv、DiffServ 等模式和 SBM 带宽管理方法,对网络服务过程中的漏洞进行抑制和控制,这些方法应用在 Web 服务资源规模相对有限的情况下,根本不能满足网络资源日益庞大的要求。

侯丽敏、张瑞坤等人提出一种基于 Agent 的 QoS 组播路由算法及仿真方法,采用 Agent 选路由,可以使系统具有一定的灵活性与适应性。由于路由器本身不需要具有 QoS 路由功能,不需要维持专门的 QoS 路由表,因此系统简单,但系统可靠性无法得到保证,Agent 的失效会带来整个系统的崩溃^[2]。文献^[3]中,唐明董、姜叶春等人提出一种用户位置感知的 Web 服务 QoS 预测方法,该方法利用一些 QoS 属性受用户位置影响的事实,采用处于同一个系统中用户位置邻近的特点,根据自治系统来计算用户位置和划分相似用户,算法先进、智能,但对于不同情况下的划分,执行结果差异很大,不具有普遍应用的意义^[3]。华哲邦、李萌等人提出一种基于时

到稿日期:2013-07-05 返修日期:2013-10-30 本文受国家自然科学基金项目(61163009),甘肃省高等学校研究生导师科研项目(1104-05)资助。
张博雅(1982-),女,硕士生,主要研究方向为电子服务,E-mail:liguo_2008@126.com;胡晓辉(1963-),男,博士,教授,主要研究方向为智能信息处理、分布计算。

间序列分析的 Web Service QoS 预测方法,实现了相应的 Web Service QoS 自动预测工具,但预测时间短^[4]。文献[5]中,许利军、杨棉绒等人提出了一种 QoS 组播路由的多种群遗传算法,种群初始化采用多种算法生成不同种群,设计了交叉、变异操作,保证系统独立性,但算法种群划分依据不太明确,应用困难。

2 传统的 QoS 预测方法及弊端分析

QoS 的准确预测是评判和选择最佳 Web 服务的一种重要标准,传统的 QoS 预测方法采用时间平均值和各种参数简单加权的方法,主要描述如下:

假设传输拓扑树 T ,对于任意一个节点 n ,其所在的层次为 $l(n)$,层次为 $l(n)$ 的所有节点的集合定义为 $L(n)$;节点 n 的子节点集合为 Dn ; $l(n)$ 层的子节点集合为 $D_{l(n)}$, $l(n)$ 层的平均子节点为 $\bar{D}_{l(n)}$;那么拓扑树的总平均子节点数为 M 层子节点数总和与层次总和的比值,即:

$$\bar{D} = \frac{\sum_{i=1}^{M-1} |D_i|}{\sum_{j=1}^{M-1} |L_j|} \quad (1)$$

QoS 质量监控过程中,Web 资源树与上层节点的最小竞争窗口值以及平均子节点数目相关,并且满足节点所在层的最小竞争窗口值必须大于或等于上层最小竞争窗口值,实现了异步传输控制,异步层最小竞争异步递进值为:

$$CW_{min}^i = \begin{cases} CW_{min}^0 \times (1 + \bar{D}_0)^x, & i=0 \\ CW_{min}^i \times (1 + \bar{D}_i)^x, & 0 < i < M \end{cases} \quad (2)$$

传输树资源 QoS 控制加权跃进系数表示为:

$$\chi = \log_{(1+\bar{D})} M A / CW_{min}^0 \quad (3)$$

式中, CW_{min}^0 为传输树中最底层根节点的预测初始化值;而 CW_{min}^i 为第 i 步的传输控制值;异步递进每一步之间加权控制构成是相同的,其差异主要是由层和层之间的平均子节点数目控制。 χ 为调节系数,对于每层的最小竞争窗口值该调节系数为定值; A 为层最小竞争窗口的上限值,可以调节由于最小竞争窗口过大而引起的信息等待时间太长的问题,其取值为:

$$\begin{aligned} CW_{min}^M &= CW_{min}^{M-1} \times (1 + \bar{D}_{l_{M-1}})^x \\ &= CW_{min}^0 \times (1 + \bar{D}_0)^x \times (1 + \bar{D}_1)^x \times \cdots \times (1 + \bar{D}_{l_{M-1}})^x \\ &= CW_{min}^0 \times [(1 + \bar{D}_0) \times (1 + \bar{D}_1) \times \cdots \times (1 + \bar{D}_{l_{M-1}})]^x \end{aligned} \quad (4)$$

其中

$$\begin{aligned} &[(1 + \bar{D}_0) \times (1 + \bar{D}_1) \times \cdots \times (1 + \bar{D}_{l_{M-1}})] \\ &\leq ((1 + \bar{D}_0) + 1 + \bar{D}_1 + \cdots + 1 + \bar{D}_{l_{M-1}}) / M^M \\ &= (1 + \bar{D})^M \end{aligned} \quad (5)$$

得:

$$\begin{aligned} CW_{min}^M &\leq CW_{min}^0 \times [(1 + \bar{D})^M]^x \\ &= CW_{min}^0 \times A / CW_{min}^0 = A \end{aligned} \quad (6)$$

在 QoS 控制 Web 信息网络拓扑树 T 中,同一层中的节点所携带的资源信息随着所包含子节点数目的不同而不同,因此,对每个节点给予不同的 Web 资源信道占用概率, QoS 异步层最小竞争加权系数为:

$$CW_{min}^n = \begin{cases} ((1 - B_{l(n)})e^{-\alpha_n} + B_{l(n)}) \times CW_{min}^{l(n)}, & \alpha_n > 1 \\ CW_{min}^{l(n)}, & 0 \leq \alpha_n \leq 1 \end{cases} \quad (7)$$

$$B_{l(n)} = (1 / (1 + \bar{D}_{l(n)-1}))^x \quad (8)$$

$$\alpha_n = \begin{cases} |D_n| / (\bar{D}_{l(n)}), & \bar{D}_{l(n)} \neq 0 \\ 0, & \bar{D}_{l(n)} = 0 \end{cases} \quad (9)$$

式中, α_n 为子节点异步迭代系数, $B_{l(n)}$ 为异构层的调整系数。Web 资源节点 QoS 异步预测算法保证了父节点比其子节点拥有更多的信道资源占用率,从而保证了网络资源的利用率,满足了拓扑传输树中的非均衡传输特点。

传统的 QoS 预测方法主要采用求平均值的方法,这类方法简单,易于实现,却没有考虑用户的个性化要求,所有用户得到的 QoS 预测值都相同。Shao L 等人在研究中提出了一种基于协作过滤的方法,该方法在很大程度上避免了用户输出、传递及环境等性能上的差异,但也存在着一些缺点,即在计算用户相似性时^[5],没有考虑用户调用服务时的环境信息,只考虑用户的历史使用信息。

3 全域子空间分解挖掘算法的提出

全域子空间分解挖掘方法是一种基于全局理论设想和多维子空间分解方法的数据挖掘分析方法。采用全域分析的方法,对数据源进行全面的整体趋势和性能分析,采用子空间分解的方法,在多个方面对数据进行分解分析,提取数据的深层次特征,然后将全域分析的结果与子空间分解分析的结果进行有效的数据融合,实现对所分析数据的有效挖掘^[6,7]。

3.1 子空间分解挖掘方法

全域挖掘的第一步是子空间分解挖掘,子空间作为全域的子集合,其分析结果的好坏直接影响着最终的全域分析结果。所以在子空间分析时,需要尽可能地对所有的子空间进行详细的分解挖掘,提取出尽可能多的信息,为下一步的全域挖掘提供数据源。采用全域子空间分解挖掘算法进行数据挖掘预测分析的基本思想如下:

假设数据挖掘预测分析时,在 t 时刻的预测分析数据信息度 $Bel(x_t) = p(x_t | d_0, \dots, t)$ 为 N ,代表随机的预测分析样本数据集,则系统预测分析集合定义为:

$$S_t = \{s_t^j(x_t^j, w_t^j) | j=1, 2, \dots, N\} \quad (10)$$

式中, x_t 为预测分析样本的实际预测分析数据; w_t 为相应的 x_t 的权值向量。

从而,预测分析系统的所有预测分析数据定义为:

$$x_k = f\{x_{k-1}, u_{k-1}, w_{k-1}\} \quad (11)$$

式中, u_k 为预测分析数据的输入模型化,比如预测分析次数、频率等参数; w_k 为预测分析统计值。

则基于子空间分解挖掘算法的预测系统测量方程为:

$$z_k = h(x_k, M, v_k) \quad (12)$$

式中, M 为预测分析时系统的正常状态; v_k 为随机扰动的预测分析数据。

从而,基于子空间分解挖掘的预测分析系统信度表达式定义为:

$$Bel(x_t) = p(x_t | z_t, u_{t-1}, z_{t-1}, \dots, u_0, z_0) \quad (13)$$

根据贝叶斯定理,得到子空间分解挖掘时系统递归计算表达式的定义为^[8]:

$$Bel(x_t) = \frac{p(z_t | x_t, u_{t-1}, z_{t-1}, \dots, z_0) p(x_t | u_{t-1}, \dots, z_0)}{p(z_t | u_{t-1}, d_0, \dots, t-1)} \quad (14)$$

子空间分解挖掘提供了各个子空间详细的数据深层结构

特征,为后面的全域挖掘提供了丰富的数据源。

3.2 全域挖掘方法

在上面子空间分解挖掘的基础上,得到了基于各个子空间的分解挖掘结果,将这些挖掘结果收集起来,形成基于全域分析的数据源。由于子空间分解挖掘中,具备了各个子空间的深层次特征,因此全域挖掘时,将这些深层次的特征融合在一起,形成全域挖掘的结果,具有更好的数据挖掘结果。

全域挖掘时,假设测量数据 z_t 与其过去的测量值 z_{t-1} 相互独立,仅与 x_t 相关,即:

$$p(x_t | z_t, u_{t-1}, \dots, z_0) = p(x_t | x_{t-1}, u_{t-1}) \quad (15)$$

则有:

$$\begin{aligned} Bel(x_t) &= \frac{p(z_t | x_t) p(x_t | u_{t-1}, \dots, z_0)}{p(z_t | u_{t-1}, d_0, \dots, t-1)} \\ &= \eta p(z_t | x_t) \int p(x_t | x_{t-1}, u_{t-1}) Bel(x_{t-1}) dx_{t-1} \end{aligned} \quad (16)$$

即:

$$p(z_t | x_{t-1}, d_0, \dots, t-1) = p(x_t | x_{t-1}, u_{t-1}) Bel(x_{t-1}) dx_{t-1} \quad (17)$$

$$p(z_t | x_t) = p(x_t | x_{t-1}, u_{t-1}) \quad (18)$$

其中,系数 η 定义为:

$$\eta = \frac{1}{p(z_t | u_{t-1}, d_0, \dots, t-1)} \quad (19)$$

通过上面的子空间分解挖掘方法和全域挖掘方法,可以在基于子空间分解挖掘后,将所有有用的数据信息进行融合,达到提取数据深度特征的目的。

4 基于全域子空间分解挖掘的 QoS 预测方法

在上面 QoS 预测模型建立和 QoS 预测方法确立的基础上,进行基于全域子空间分解挖掘的 QoS 预测,将全域子空间分解挖掘的结果输入到 QoS 预测模型,采用设计的 QoS 预测方法,实现精确的 QoS 预测。

全域子空间分解挖掘时,子空间的分解对于最终的挖掘结果影响很大,这里采用正交子空间分解的方法,对数据进行全方位分析。将 QoS 预测数据分解为两个空间向量,一是 QoS 预测期望子空间向量,另一个是 QoS 预测误差子空间向量。设计采用两个空间向量的正交特性进行预测分析处理和特征量的检测和提取^[9-11]。

令 $R_{d \times L}$ 为 $d \times L$ 维的矩阵, L 为每次进行特征向量提取运算的数据长度,对于每一个预测模型,空间矢量表示为:

$$R_1 = \{X_1, X_2, X_3, \dots, X_d\}^T \quad (20)$$

得到相关函数为:

$$R_1^T R_1 = \{X_1, X_2, \dots, X_m\} \{X_1, X_2, \dots, X_m\}^T \quad (21)$$

对上式进行特征值分解得:

$$R_1^T R_1 = V_1 \Sigma_1 V_1^T \quad (22)$$

从 $L+1$ 到 $2L$ 维 QoS 预测分析数据迭代类推得到:

$$R_2^T R_2 = V_2 \Sigma_2 V_2^T \quad (23)$$

$$R_2 = \{X_{d+1}, X_{d+2}, \dots, X_{d+m}\}^T \quad (24)$$

$$R_2^T R_2 = \{X_{d+1}, X_{d+2}, \dots, X_{d+m}\} \{X_{d+1}, X_{d+2}, \dots, X_{d+m}\}^T \quad (25)$$

式中,向量 $V = [V_1, V_2, \dots, V_m] \in R^{m \times m}$ 且正交,从而得到:

$$VV^T = I_M \quad (26)$$

$$\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_m) \in R^{m \times m} \quad (27)$$

上式即为提取的 $R^T R$ 的 QoS 预测分析数据的特征值,这些特征值之间存在大小关系为:

$$\sigma_1 > \sigma_2 > \sigma_3 > \dots > \sigma_{s+1} > \sigma_m \quad (28)$$

向量 V 是 $R_2^T R_2$ 的规范正交基特征向量,特征值 Σ 中含有 QoS 预测信息的所有状态信息,能有效表征研究的 QoS 预测分析数据的状态参量。由上式,还可得:

$$R^T R = [V_s \quad V_n] \begin{bmatrix} \Sigma_s & 0 \\ 0 & \Sigma_n \end{bmatrix} [V_s \quad V_n]^T \quad (29)$$

本文特征提取算法中采用多次迭代求解特征值的方法,用 QoS 有效预测与 QoS 预测误差比 $\Delta \text{SNR}(i)$ 表示 i 次 QoS 预测特征信号的预测性能, $\Delta \text{SNR}(i+1)$ 表示经过 $i+1$ 次迭代后对于 QoS 预测的有效性提高量,循环迭代,当下式成立时:

$$\Delta \text{SNR}(i+1) < \Delta \text{SNR}(i) \quad (30)$$

则循环结束,得出最优化预测性能。

5 系统实验与结果分析

5.1 实验环境描述

为了测试基于全域子空间分解挖掘的 QoS 预测效果,构建如下测试实验环境,详细预测度量参数见表 1。

表 1 度量参数描述

度量类型	度量属性	度量权重	度量标准
费用	2	0.0408	元
反应时间	1	0.1676	s
效率	2	0.0857	%
可用性	1	0.1071	[0,1]
可靠性	1	0.1572	[0,1]
信誉度	2	0.0805	[0,1]
安全度	1	0.1344	[0,1]
执行成功率	1	0.1449	%
稳定性	2	0.0785	[0,1]
可重载性	2	0.0033	[0,1]

度量属性中:

1 代表重要的度量属性,加权系数较大;

2 代表一般的度量属性,加权系数较小。

采用分布下的 Web 资源节点进行实际的 QoS 预测仿真测试,各个 Web 节点参数分布见表 2。

表 2 实验节点参数分布

节点编号	节点容量 (GB)	节点访问量 (万次/日)	节点重要性
节点 1	251.40	49.6552	0.0743
节点 2	354.73	89.9769	0.0578
节点 3	214.44	82.1629	0.0471
节点 4	152.30	64.491	0.0735
节点 5	94.82	81.7974	0.0657
节点 6	96.71	66.0228	0.0841
节点 7	341.11	34.1971	0.1012
节点 8	151.38	28.9726	0.0553
节点 9	270.83	34.1194	0.0931
节点 10	75.43	53.4079	0.0183
节点 11	348.94	72.7113	0.1036
节点 12	189.18	30.929	0.0287
节点 13	430.00	83.8496	0.0267
节点 14	426.82	56.8072	0.0926
节点 15	296.78	37.0414	0.078

5.2 结果分析

度量权重的分布如图 1 所示。

(下转第 224 页)

[2] BlogEngine[OL]. <http://www.dotnetblogengine.net/>

[3] Brian C, Jacob W. Secure programming with static analysis [M]. Addison-Wesley Professional, 2007

[4] Alain D. Interprocedural may-alias analysis for pointers; beyond k -limiting [C]//Proceedings on PLDI, 1994; 230-241

[5] Seth H, Benjamin C, Xie Yi-chen, et al. A system and language for building system-specific, static analyses [C]// Proceedings on PLDI, 2002; 69-82

[6] Heine D L, Lam M S. A practical flow-sensitive and context-sen-

sitive C and C++ memory leak detector [C]//Proceedings on PLDI, 2003; 168-181

[7] 梁彬, 侯看看, 石文昌, 等. 一种基于安全状态跟踪检查的漏洞静态检测方法研究与实施 [J]. 计算机学报, 2009, 32(5): 899-909

[8] Mono[OL]. http://www.mono-project.com/Main_Page

[9] SourceGrid[OL]. <http://sourcegrid.codeplex.com/>

[10] 夏一民, 罗军, 张民选. 基于静态分析的安全漏洞检测技术研究 [J]. 计算机科学, 2006, 33(10): 279-282

(上接第 219 页)

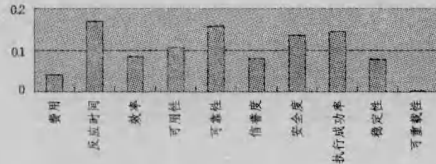


图 1 度量权重分布图

在上面各种度量参数分布和 Web 实验节点分布的基础上, 系统采用 6 个监测点进行实际的预测实验。6 个监测点在发生变化的过程中采用逐渐变化的方法, 采用过程变化量作为衡量系统 QoS 预测效果的标准, 通过渐变过程, 可以体现 QoS 预测效果的精细和准确度。采用传统方法进行预测的结果如图 2 所示。

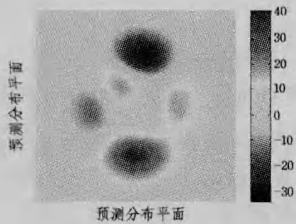


图 2 传统方法预测结果

从图 2 可以看出, 采用传统的预测方法, 可以实现各个变化量的预测, 但是预测的结果在预测分布平面上很模糊, 无法精细地辨别出渐变过程中各个阶段的变化量。

采用基于全域子空间分解挖掘的预测结果如图 3 所示。



图 3 全域子空间分解挖掘的预测结果

从图 3 可以看出, 采用全域子空间分解挖掘的预测方法不但可以很好地将各个变化的监测点预测出来, 而且可以实现渐变过程的精细化预测, 预测结果在预测分布平面上很清晰, 可以很好地辨别出渐变过程各个阶段的变化量。

所以, 相对于传统的模糊预测结果, 采用基于全域子空间分解挖掘的预测方法, 预测结果更加精细, 当然也更加准确。

结束语 研究了一种基于全域子空间分解挖掘的 QoS 准确预测方法。随着现代 Web 资源访问的迅猛增长, 如何评价和预测是一个难点, 所以 QoS 的准确预测是评判和选择最佳 Web 服务的一种重要标准。在传统的 QoS 预测方法中,

经常采用普通时间平均值方法或者对各种评价参数进行一个简单的参数加权方法, 此方法对于少量简单资源可以适用, 但无法对大量 Web 服务下的资源进行准确预测, 预测结果模糊。提出了一种基于全域子空间分解挖掘的 QoS 准确预测方法, 采用全域分析思想和子空间分解的方法, 在全域对数据进行预处理, 在子空间中数据进行分析, 提取数据的深层次特征, 最后将两者进行有效的数据融合, 实现对所分析数据的有效挖掘和准确预测。采用一组 Web 节点和拟定量参数进行预测实验, 结果显示, 采用基于全域子空间分解挖掘的 QoS 预测方法, 可以精确预测出渐变过程, 结果准确, 在 QoS 预测中具有广泛的应用价值。

参考文献

[1] 孔维梁. 基于二维 QoS 模型的 Web 服务组合 [J]. 计算机科学, 2008, 35(11): 131-135

[2] 侯丽敏, 张瑞坤. 基于 Agent 的 QoS 组播路由算法及仿真 [J]. 计算机仿真, 2011, 28(1): 140-144

[3] 唐明董, 姜叶春, 刘建勋. 用户位置感知的 Web 服务 QoS 预测方法 [J]. 小型微型计算机系统, 2012, 33(12): 2664-2667

[4] 华哲邦, 李萌. 基于时间序列分析的 Web Service QoS 预测方法 [J]. 计算机科学与探索, 2013, 7(3): 219-228

[5] 许利军, 杨棉绒. QoS 组播路由的多种群遗传算法 [J]. 科技通报, 2012, 28(5): 171-175

[6] 邹恩, 刘泽华, 方仕勇, 等. 基于混沌遗传算法的组播路由优化研究 [J]. 计算机工程, 2011, 37(3): 155-157

[7] Li Yan, Liu Yao, Zhang Liang-jie, et al. An exploratory study of Web services on the Internet [C] // Proceedings of the 2007 IEEE International Conference on Web Services (ICWS 07). Salt Lake City, UT, USA, 2007, Washington, DC, USA; IEEE Computer Society, 2007; 380-387

[8] Charif-Djebbar Y, Sabouret N. Dynamic service composition and selection through an agent interaction protocol [C] // Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops (WI-IATW 06). Washington, DC, USA; IEEE Computer Society, 2006; 105-108

[9] Shao Ling-shuang. Research on quality of Web services management technologies [D]. Beijing: Peking University, 2009

[10] 曾欢, 张灿, 陈德元. 空间通信网中音视频传输的应用层 QoS 控制与测试方法 [J]. 中国科学院研究生院学报, 2011, 28(1): 108-114

[11] 邵凌霜. 面向 Web Services 的服务质量管理技术研究 [D]. 北京: 北京大学, 2009

[12] 张力娜, 李小林. 一种基于 QoS 偏好的服务选择策略 [J]. 重庆邮电大学学报: 自然科学版, 2013, 25(6): 807-812