

基于结构化噪声矩阵补全的 Web 服务 QoS 预测

陈蕾^{1,2}, 杨庚^{1,2}, 陈正宇¹, 肖甫¹, 许建¹

(1. 南京邮电大学 计算机学院, 江苏 南京 210003;

2. 南京邮电大学 宽带无线通信与传感网技术教育部重点实验室, 江苏 南京 210003)

摘要: 随着面向服务计算技术的快速发展, 越来越多具有相同或相似功能的 Web 服务被部署在网络上。用户进行服务选择之前, 通常需要根据历史调用信息对未使用过的服务 QoS 进行预测。由于历史调用信息收集过程缺乏有效的监督和约束机制, 所采样的 QoS 信息往往容易受到结构化噪声污染, 从而导致现有方法预测性能急剧下降。为了克服这个困难, 通过将 Web 服务 QoS 预测问题建模为 L_{2,1} 范数正则化矩阵补全问题, 提出了一类基于结构化噪声矩阵补全的 Web 服务 QoS 预测方法。真实数据集上的实验结果表明, 该方法不仅能精确地辨识出 QoS 采样矩阵中噪声行所在位置, 而且能对缺失 Web 服务 QoS 进行有效预测。

关键词: Web 服务; QoS 预测; 矩阵补全; 算子分裂; 结构化噪声

中图分类号: TP311

文献标识码: A

Web services QoS prediction via matrix completion with structural noise

CHEN Lei^{1,2}, YANG Geng^{1,2}, CHEN Zheng-yu¹, XIAO Fu¹, XU Jian¹

(1. School of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China;

2. Key Laboratory of Broadband Wireless Communication and Sensor Network Technology of Ministry of Education, Nanjing University of Post and Telecommunications, Nanjing 210003, China)

Abstract: With the rapid development of service-oriented computing, more and more Web services with the same or similar function are deployed on the Internet. Usually, before selecting the most suitable service, users need to predict QoS of unused services from the service invoking history. Due to the lack of effective supervision and constraint mechanisms, some number of the rows in the QoS sample matrix is often contaminated by the structural noise, which leads to a sharp decrease for QoS prediction performance. In order to address this problem, an efficient Web services QoS prediction approach via matrix completion with structural noise is proposed by formulating Web services QoS prediction problem as a L_{2,1}-norm regularized matrix completion problem. The proposed approach can not only exactly detect the position where the data is contaminated, but also effectively predict the missing QoS values. Finally, experimental results performed on a real public dataset demonstrate the feasibility of our proposed approach.

Key words: Web service; QoS prediction; matrix completion; operator splitting; structural noise

1 引言

面向服务计算 (SOC, service-oriented computing) 作为一种新型计算范型, 为分布式应用和企业

信息系统集成提供了一种灵活的解决方案, 已成为计算机软件领域最热门的研究主题之一。随着面向服务计算技术的快速发展, 网络上出现了大量功能相同或相似而服务质量 (QoS, quality of service) 不

收稿日期: 2014-04-19; 修回日期: 2014-10-16

基金项目: 国家自然科学基金资助项目 (61272084, 61201163, 61272422, 61373137); 江苏省自然科学基金资助项目 (BK2011072, BK20130096); 江苏省高校自然科学研究重大基金资助项目 (11KJA520002); 高等学校博士学科点专项科研基金资助项目 (20113223110003)

Foundation Items: The National Natural Science Foundation of China (61272084, 61201163, 61272422, 61373137); The Natural Science Foundation of Jiangsu Province (BK2011072, BK20130096); The Key Natural Science Research Program of Jiangsu (11KJA520002); The Research Fund for the Doctoral Program of High Education (20113223110003)

同的 Web 服务。用户在选择满足其个性化需求的 Web 服务时,不仅要考虑服务的功能属性,而且要考虑服务的非功能属性,即 QoS。服务选择时将 QoS (如响应时间、可用性、可靠性、吞吐量等)考虑在内,对成功构建面向服务的应用具有非常重要的意义。为此,众多研究人员展开了关于 QoS 感知的 Web 服务选择和组合方法的研究^[1-4],所有这些研究都基于一个共同的前提,那就是假设可供用户选择的服务其 QoS 信息不仅全部已知而且真实可信。但是,在实际应用中这种假设并不总是成立。一方面,由于服务 QoS 的个性化和服务资源的海量性,使用户所能获取的 QoS 信息不可能是全面的;另一方面,由于网络环境的开放性和不稳定性,使用户所获取的 QoS 信息将不可避免地存在一些噪声。因此,在开放网络环境下提供有效的 Web 服务 QoS 预测是进行 Web 服务选择和组合的重要前提和保障。

近年来,许多学者展开了 Web 服务 QoS 预测的研究^[5-10]。但是所有这些 Web 服务 QoS 预测方法都仅能处理 QoS 采样信息不含噪或仅含少量高斯随机噪声的情形,而在实际的应用场景中,由于服务 QoS 信息收集过程缺乏有效的监督和约束机制,少数不诚实的 Web 服务提供者可能出于不当牟利的目的伪装或者串通用户恶意地篡改竞争对手和自己的 QoS 信息。如果将用户反馈的历史调用信息建模为一个用户-服务 QoS 采样矩阵(采样矩阵的行代表用户,列代表 Web 服务,采样矩阵的元素代表相应的用户调用 Web 服务时所对应的个性化 QoS),那么这些不诚实用户所反馈的 QoS 信息就表现为 QoS 采样矩阵中的行结构化噪声。这种结构化噪声的存在使现有 Web 服务 QoS 预测方法的预测性能急剧下降。因此,在进行实际的 Web 服务 QoS 预测时,如何在结构化噪声采样情形下为 Web 服务用户提供尽可能准确地个性化 QoS 预测就成为一个亟待解决的难题。

为了解决结构化噪声采样情形下 Web 服务 QoS 预测问题,本文首先利用用户-服务 QoS 矩阵的近似低秩特性^[10],通过引入分类器设计领域中的 L2,1 范数正则化技术^[11]将结构化噪声采样情形下的 Web 服务 QoS 预测问题建模为一类 L2,1 范数正则化矩阵补全问题,所引入的 L2,1 范数能有效平滑此类结构化噪声。其次,由于 L2,1 范数正则化矩阵补全问题是一类典型的非光滑凸优化问题,该问题

并不存在显式的解析解,因此本文进一步引入机器学习流行的算子分裂技术^[12],提出一类基于算子分裂技术的结构化噪声矩阵补全算法(OSMCSN, operator splitting based matrix completion with structural noise)迭代求解该问题。真实数据集上的实验结果表明,所提出的 OSMCSN 方法一方面能精确地辨识出 QoS 采样矩阵中噪声行所在位置,从而可以实现对操纵和篡改 QoS 信息的不诚实行为进行预警;另一方面可以充分利用所辨识出的噪声行位置信息,在进行缺失 QoS 预测前预先将采样矩阵中的噪声行移除,从而能在不含噪情形下对缺失的 Web 服务 QoS 进行预测,最后的实验结果表明采用本方法所获得的 QoS 预测性能明显优于现有的 Web 服务 QoS 预测方法,从而使在开放网络环境下准确发现满足用户需求的可信 Web 服务更加成为可能,进而为高可信服务应用系统的构建提供重要的技术支撑。

2 相关研究工作

作为服务计算领域的热门研究主题之一,QoS 感知的 Web 服务选择与组合问题引起了研究者的广泛关注,并提出了许多有效的算法和解决思路。在服务选择方面,Xia 等引入服务聚类思想将具有相同功能的 Web 服务按照其 QoS 进行自适应分类^[1],提出了一种 QoS 感知的 Web 服务选择算法。王尚广等则引入自适应粒子群优化算法提出了一种基于全局 QoS 约束分解的动态服务选择方法来支持快速和动态的高可信服务应用系统集成^[2]。在服务组合方面,邓水光等提出了一种将正向搜索与逆向搜索相结合并追求 QoS 最优的快速服务自动组合方法^[3]。Parejo 等则提出了一种基于 GRASP 和 Path-Relinking 的启发式 QoS 感知 Web 服务组合算法^[4]。尽管上述这些 QoS 感知 Web 服务选择和组合方法在各自的数据集上分别得到了有效性验证,但其有效性都基于一个共同的先验假设,即所有备选服务的个性化 QoS 信息对用户而言都是已知的和准确的。而事实上通过本文引言部分的分析可知这个假设并不总是成立。

为此,诸多研究者开始研究如何根据服务的历史调用信息对未使用过的服务 QoS 进行预测。现有的 Web 服务 QoS 预测方法主要有 2 类:一类是基于协同过滤的 QoS 预测方法^[5-8],该类方法以采样到的用户-服务 QoS 矩阵为基础,通过采用特定的

相似性度量方法分别计算用户之间、Web 服务之间的相似程度对目标用户进行个性化的 QoS 预测。协同过滤是一种广泛使用的信息检索方法，最初被用于个性化推荐系统中预测用户的兴趣偏好，从而获得用户潜在感兴趣的产品或服务。协同过滤有许多成功应用的案例：Amazon 的商品推荐、Pandora 的音乐推荐、Netflix 的电影推荐、Google 的新闻推荐以及 Facebook 的朋友推荐等。受到协同过滤技术在个性化推荐系统中得到良好应用的启发，Shao 等率先在文献[5]中提出了一种基于协同过滤的个性化 QoS 预测方法 (UPCC)，该方法首先以用户的历史 QoS 信息为基础计算用户之间的相似程度，然后根据相似用户之间的经验对某些用户需要的服务进行个性化的 QoS 预测。然而，由于作者在相似度计算中使用了 Pearson 相关系数，该系数一般用于度量数据之间的线性相关关系，并不能很好地解释 QoS 数据之间的相似性。后来，作者对该工作做了进一步优化^[6]，增加了相似服务对 QoS 预测的权重，同时改进了用户线性相似度的计算方法，采用了一种简单有效的基于欧几里德距离的相似度量方法，较好地解决了用户较少时有效计算用户相似性的问题。另一方面，Zheng 等在文献[7]中不仅提出了一种用户协同的 QoS 信息收集机制，而且提出了一种基于邻域的 Web 服务 QoS 预测方法 (UIPCC)，UIPCC 方法同时考虑了用户相似性和服务相似性，该文献的实验结果表明该算法取得了较好的个性化 QoS 预测效果。其后，Wu 等提出了另外一种基于邻域的协同过滤方法 (ADF) 来预测缺失的 Web 服务 QoS^[8]，该方法一方面采用一种改进的余弦相似性度量准则来移除 QoS 值差异对服务相似性计算的影响，另一方面引入数据平滑技术对训练集中的用户进行聚类操作，并利用各聚类中用户的历史 QoS 信息对预测系统的输入数据进行数据平滑处理，同时采用一种相似性融合机制来处理采样数据的稀疏问题，这些措施有效地提高了 ADF 方法的预测精度。另一类 Web 服务 QoS 预测方法是基于矩阵分解的预测方法^[9,10]，该类方法也是以收集到的 QoS 采样矩阵为基础，并结合 QoS 矩阵的近似低秩特性，利用矩阵分解技术将 QoS 采样矩阵分解为低秩服务特征矩阵和低秩用户特征矩阵对 QoS 进行预测。Zheng 等在文献[9]中提出了一种融合邻近信息的矩阵分解协同方法 (NIMF) 来预测 Web 服务 QoS，该方法巧妙地将局部的用户相似性邻域

信息融入包含全局关联信息的矩阵分解模型，消除了基于协同过滤的 QoS 预测方法由于用户-服务 QoS 矩阵的稀疏性而带来的计算相似服务和相似用户不准确的影响。Yu 等则在文献[10]中同时利用 QoS 矩阵的近似低秩特性和用户/服务的相似性聚类特性，采用非负矩阵三分解算法将采样 QoS 矩阵分解为 3 部分：用户聚类特征矩阵、服务聚类特征矩阵和用户聚类/服务聚类关联矩阵，从而提出一种新颖的核范数正则化矩阵分解 (TNR-MF) 方法进行 Web 服务 QoS 预测。尽管所有这些方法在各自选取的数据集上取得了良好的 QoS 预测性能，但是它们无一例外地仅考察了无噪声 QoS 采样情形下缺失 QoS 信息的预测，并没有考虑采样 QoS 矩阵中可能存在虚假和错误信息的情形。事实上，在实际的应用场景中，由于历史调用信息的反馈缺乏有效的监督和约束机制，少数不诚实的 Web 服务提供商可能出于不当牟利的目的伪装或者串通用户恶意地篡改竞争对手和自己的 QoS 信息，从而使所采样的 QoS 矩阵受到结构化噪声的污染。因此，如何在结构化噪声采样情形下有效进行 Web 服务 QoS 的个性化预测就成为了亟待解决的重要问题。

3 预备知识

定义 1 矩阵 $L_{2,1}$ 范数^[13] 矩阵 $X \in \mathbb{R}^{n_1 \times n_2}$ 的 $L_{2,1}$ 范数定义为

$$\|X\|_{2,1} = \sum_{i=1}^{n_1} \|X^{(i)}\|_2 = \sum_{i=1}^{n_1} \left(\sum_{j=1}^{n_2} X_{ij}^2 \right)^{1/2}$$

定义 2 矩阵函数次梯度与次微分^[13] 令 $F(X)$ 是 $\mathbb{R}^{n_1 \times n_2}$ 上的实值凸函数，如果 $Z \in \mathbb{R}^{n_1 \times n_2}$ 满足 $F(X) \geq F(X_0) + \langle Z, X - X_0 \rangle$, $\forall X \in \mathbb{R}^{n_1 \times n_2}$ ，则称 Z 为矩阵函数 $F(X)$ 在 X_0 处的次梯度，所有在 X_0 处的次梯度的集合记为 $\partial F(X_0)$ ，同时称 $\partial F(X_0)$ 为 $F(X)$ 在 X_0 处的次微分。

定义 3 奇异值阈值算子^[14] 假设秩为 r 的矩阵 $X \in \mathbb{R}^{n_1 \times n_2}$ 的奇异值分解为 $X = U \Sigma V^T$ ，其中 $\Sigma = \text{diag}(\{\sigma_i | 1 \leq i \leq r \text{ 且 } \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0\})$ ，如果 $\tau \geq 0$ ，则 τ 对应的奇异值阈值算子定义为

$$D_\tau(X) = U S_\tau(\Sigma) V^T \quad (1)$$

其中， $S_\tau(\Sigma) = \text{diag}(\{\max(0, \sigma_i - \tau)\})$ 。

定理 1^[14] 对任意的 $\tau, \mu > 0$ ， $Z \in \mathbb{R}^{n_1 \times n_2}$ ，有

$$D_{\tau/\mu}(Z) = \arg \min_{X \in \mathbb{R}^{n_1 \times n_2}} \left\{ \tau \|X\|_* + \frac{\mu}{2} \|X - Z\|_F^2 \right\}.$$

证明 略, 详见文献[14].

定理 2^[12] 近邻前向后向分裂算法 (PFBS, proximal forward backward splitting algorithm) 假设 F_1, F_2 是 2 个下半连续的凸函数, F_2 在 $\mathbb{R}^{n_1 \times n_2}$ 中可微且对某个常数 $\beta > 0$ 具有 β -Lipschitz 连续梯度, 即 $\|\nabla F_2(U) - \nabla F_2(V)\|_F \leq \beta \|U - V\|_F$, 则对于凸优化问题

$$\min_{X \in \mathbb{R}^{n_1 \times n_2}} F_1(X) + F_2(X) \quad (2)$$

有如下性质:

- 1) 如果 $F_1 + F_2$ 是强制的, 即 $\lim_{\|X\|_F \rightarrow +\infty} F_1(X) + F_2(X) = +\infty$, 则问题 2) 至少存在一个解;
- 2) 如果 $F_1 + F_2$ 是严格凸的, 则问题 2) 至多存在一个解;
- 3) 如果 F_1, F_2 满足条件 1) 和 2), 则问题 2) 存在唯一解, 且对任意的初始值 X^0 及 $0 < \delta < 2/\beta$ 用如下方法生成的迭代序列 X^{k+1} 收敛到问题 2) 的唯一解

$$X^{k+1} = \text{prox}_{\delta F_1(X)}(X^k - \delta \nabla F_2(X^k)) \quad (3)$$

证明 略, 详见文献[12].

定理 3 对任意的 $\tau, \mu > 0, W \in \mathbb{R}^{n_1 \times n_2}$, 函数 $H(X) = \tau \|X\|_{2,1} + \frac{\mu}{2} \|X - W\|_F^2$ 存在全局最小点 X^* : $(X^*)^{(i)} = \max\{\|W^{(i)}\|_2 - \tau/\mu, 0\} W^{(i)} / \|W^{(i)}\|_2, i = 1, 2, \dots, n_1$, 其中 $X^{(i)}$ 表示矩阵 X 的第 i 行, $\|\cdot\|_2$ 表示向量 L2 范数.

证明 显然, 函数 $H(X)$ 是 $\mathbb{R}^{n_1 \times n_2}$ 上的凸函数, 因此根据凸函数的性质易知函数 $H(X)$ 一定存在全局最小点, 不妨记其为 X^* , 即

$$\begin{aligned} X^* &= \arg \min_{X \in \mathbb{R}^{n_1 \times n_2}} \tau \|X\|_{2,1} + \frac{\mu}{2} \|X - W\|_F^2 \\ &= \arg \min_{X \in \mathbb{R}^{n_1 \times n_2}} \tau \sum_{i=1}^{n_1} \|X^{(i)}\|_2 + \frac{\mu}{2} \sum_{i=1}^{n_1} \|X^{(i)} - W^{(i)}\|_2^2 \end{aligned} \quad (4)$$

由于 $n_1 \times n_2$ 阶矩阵 X^* 可分解为 n_1 个 n_2 阶行向量 $\{(X^*)^{(i)} | i = 1, 2, \dots, n_1\}$, 而且式 (4) 的目标优化函数具有变量可分离性, 因此对于 $i = 1, 2, \dots, n_1$, 有

$$(X^*)^{(i)} = \arg \min_{X^{(i)} \in \mathbb{R}^{n_2}} \tau \|X^{(i)}\|_2 + \frac{\mu}{2} \|X^{(i)} - W^{(i)}\|_2^2 \quad (5)$$

为描述方便, 不妨将行向量 $W^{(i)}$ 记为 v , 将 $X^{(i)}$ 记为 u , 将 $(X^*)^{(i)}$ 记为 u_0 , 因此式 (5) 可简化为

$$u_0 = \arg \min_{u \in \mathbb{R}^{n_2}} \tau \|u\|_2 + \frac{\mu}{2} \|u - v\|_2^2 \quad (6)$$

根据次微分的定义, 易知函数 $f(u) = \tau \|u\|_2 + \frac{\mu}{2} \|u - v\|_2^2$ 在 u_0 的次微分为

$$\partial(f(u_0)) = \begin{cases} \mu(u_0 - v) + \frac{\tau u_0}{\|u_0\|_2}, & u_0 \neq 0 \\ \{\mu(u_0 - v) + y : y \in \mathbb{R}^{n_2}, \|y\|_2 \leq \tau\}, & u_0 = 0 \end{cases}$$

由次微分性质可知, 如果 u_0 是凸函数 $f(u)$ 的全局最优点当且仅当 $0 \in \partial(f(u_0))$, 因此式 (6) 的解 u_0 必须满足

$$\begin{cases} \mu(u_0 - v) + \frac{\tau u_0}{\|u_0\|_2} = 0, & u_0 \neq 0 \\ \mu(u_0 - v) + y = 0 \text{ 且 } \|y\|_2 \leq \tau, & u_0 = 0 \end{cases} \quad (7)$$

即

$$\begin{cases} \mu(u_0 - v) + \frac{\tau u_0}{\|u_0\|_2} = 0, & u_0 \neq 0 \\ \mu \|v\|_2 \leq \tau, & u_0 = 0 \end{cases} \quad (8)$$

由式 (8) 可知

- 1) 当 $\|v\|_2 \leq \tau/\mu$ 时, $u_0 = 0$
- 2) 当 $\|v\|_2 > \tau/\mu$ 时, $u_0 \neq 0$ 。此时有

$$\mu(u_0 - v) + \frac{\tau u_0}{\|u_0\|_2} = 0$$

即 $u_0 = \frac{\mu}{\mu + \tau/\|u_0\|_2} v$, 进而可得

$$\|u_0\|_2 = \left\| \frac{\mu}{\mu + \tau/\|u_0\|_2} v \right\|_2 = \frac{\mu}{\mu + \tau/\|u_0\|_2} \|v\|_2$$

化简后得

$$\|u_0\|_2 = \|v\|_2 - \tau/\mu \quad (9)$$

因此可按式(10)计算 u_0 。

$$\begin{aligned} u_0 &= \|u_0\|_2 \frac{u_0}{\|u_0\|_2} \\ &= \|u_0\|_2 \left(\frac{\mu}{\mu + \tau/\|u_0\|_2} v \middle/ \left\| \frac{\mu}{\mu + \tau/\|u_0\|_2} v \right\|_2 \right) \quad (10) \\ &= \|u_0\|_2 \frac{v}{\|v\|_2} \end{aligned}$$

将式 (9) 代入式 (10) 后可得

$$u_0 = (\|v\|_2 - \tau/\mu) \frac{v}{\|v\|_2} \quad (11)$$

综合 1), 2), 易知

$$\mathbf{u}_0 = \begin{cases} (\|\mathbf{v}\|_2 - \tau/\mu) \frac{\mathbf{v}}{\|\mathbf{v}\|_2}, & \|\mathbf{v}\|_2 > \tau/\mu \\ 0, & \|\mathbf{v}\|_2 \leq \tau/\mu \end{cases}$$

即

$$\mathbf{u}_0 = \max\{\|\mathbf{v}\|_2 - \tau/\mu, 0\} \frac{\mathbf{v}}{\|\mathbf{v}\|_2} \quad (12)$$

由于之前用 \mathbf{v} 表示 $\mathbf{W}^{(i)}$, \mathbf{u}_0 表示 $(\mathbf{X}^*)^{(i)}$, 因此式 (12) 可写为

$$(\mathbf{X}^*)^{(i)} = \max\{\|\mathbf{W}^{(i)}\|_2 - \tau/\mu, 0\} \frac{\mathbf{W}^{(i)}}{\|\mathbf{W}^{(i)}\|_2}$$

由此, 命题得证。

4 Web 服务 QoS 预测问题建模

假设存在 n_1 个用户组成的用户集合 $U = \{u_i | i = 1, 2, \dots, n_1\}$ 和 n_2 个 Web 服务组成的服务集合 $S = \{ws_j | j = 1, 2, \dots, n_2\}$ 。同时假设存在一个 $n_1 \times n_2$ 的用户-服务 QoS 采样矩阵 \mathbf{R} , 其元素 R_{ij} 表示用户 u_i 实际调用服务 ws_j 时所反馈的 QoS 信息, 如果用户 u_i 没有调用过服务 ws_j , 则令 $R_{ij} = \text{null}$ 。本文拟采用矩阵补全技术来预测缺失的 Web 服务 QoS 值。由矩阵补全的基本理论可知^[15], 一般情况下, 从矩阵的部分采样元素预测其余的未知元素是不可能的, 除非该矩阵具有低秩或近似低秩特性。已有的研究表明^[9,10], 由于 QoS 具有较高的时间和空间关联性, 使用户-服务 QoS 矩阵通常表现出良好的近似低秩特性。

通常, 标准矩阵补全问题被描述为

$$\min_{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}} \text{rank}(\mathbf{X}) \quad \text{s.t. } P_\Omega(\mathbf{R}) = P_\Omega(\mathbf{X}) \quad (13)$$

其中, Ω 为采样元素的索引集合且 $\Omega \subseteq [n_1] \times [n_2]$ ($[n_1] = \{1, \dots, n_1\}$, $[n_2] = \{1, \dots, n_2\}$), $P_\Omega(\cdot)$ 为正交投影算子, 表示当 $(i, j) \in \Omega$ 时, R_{ij} 为采样元素, 即

$$[P_\Omega(\mathbf{R})]_{ij} = \begin{cases} R_{ij}, & (i, j) \in \Omega \\ 0, & \text{其他} \end{cases}$$

在式 (13) 中, 由于 $\text{rank}(\cdot)$ 函数的非凸性和非光滑性, 导致该问题求解成为 NP 难问题, 在所有已知的求解算法中, 其求解的复杂度均随着矩阵维数的增长呈平方倍指数关系增长。为了使矩阵补全问题易于求解, Candes 等^[15]利用凸光滑的矩阵核范

数代替非凸的矩阵秩函数, 将问题 (13) 松弛为如下凸优化问题

$$\min_{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}} \|\mathbf{X}\|_* \quad \text{s.t. } P_\Omega(\mathbf{R}) = P_\Omega(\mathbf{X}) \quad (14)$$

目前, 针对上述标准的矩阵补全问题及其变体提出了许多有效算法。这些算法大致可分为 2 类: 一类是基于核范数最小化的迭代算法, 如 SVT 算法^[14], FPCA 算法^[16]、IALM 算法^[17]等; 另一类是基于矩阵分解的优化算法, 包括 LMaFit 算法^[18]和 RTRMC 算法^[19]等。

现有的这些矩阵补全算法或者完全没有考虑采样矩阵的含噪问题, 或者只是简单地假设采样矩阵受到随机高斯噪声污染, 并不能解决 Web 服务 QoS 预测问题中遇到的结构化噪声污染问题。在 Web 服务 QoS 预测问题中, 由于缺乏有效的监督和约束机制, 不可避免地存在少数不诚实的 Web 服务提供商可能出于不当牟利的目的伪装或者串通用户恶意地篡改竞争对手和自己的 QoS 信息, 从而使所采样的用户-服务 QoS 矩阵中出现相应的行结构化噪声。为了有效地平滑这类结构化噪声, 将其带来的不利影响尽可能降到最低, 本文通过引入矩阵 L2,1 范数正则化项到标准矩阵补全问题, 对其目标函数施加相应的正则化约束, 从而将 Web 服务 QoS 预测问题建模为如下的基于 L2,1 范数正则化的结构化噪声矩阵补全问题。

$$\begin{aligned} \min_{\mathbf{X}, \mathbf{Z} \in \mathbb{R}^{n_1 \times n_2}} \|\mathbf{X}\|_* + \lambda \|\mathbf{Z}\|_{2,1} \\ \text{s.t. } P_\Omega(\mathbf{R}) = P_\Omega(\mathbf{X} + \mathbf{Z}) \end{aligned} \quad (15)$$

其中, $\mathbf{R} \in \mathbb{R}^{n_1 \times n_2}$ 为采样矩阵, \mathbf{X} 为优化矩阵, \mathbf{Z} 为噪声矩阵, $\|\mathbf{Z}\|_{2,1}$ 为平滑结构化噪声的矩阵 L2,1 范数正则化项, λ 是一个用来平衡结构化噪声和矩阵低秩程度的可调参数。

5 基于算子分裂技术的结构化噪声矩阵补全算法 (OSMCSN)

为了求解式 (15), 进一步将其松弛为如下无约束优化问题

$$\min_{\mathbf{X}, \mathbf{Z} \in \mathbb{R}^{n_1 \times n_2}} \left\{ \mu (\|\mathbf{X}\|_* + \lambda \|\mathbf{Z}\|_{2,1}) + \frac{1}{2} \|P_\Omega(\mathbf{R} - \mathbf{X} - \mathbf{Z})\|_F^2 \right\} \quad (16)$$

松弛的目的一方面是为了使约束优化问题转换为无约束优化问题后更加易于求解, 另一方面是

为了使松弛后得到的新问题除了能够处理结构化噪声,还能处理其他类型的噪声,例如高斯随机噪声等。

为描述方便,不妨定义

$$L(\mathbf{X}, \mathbf{Z}) = \mu(\|\mathbf{X}\|_* + \lambda\|\mathbf{Z}\|_{2,1}) + \frac{1}{2}\|P_\Omega(\mathbf{R} - \mathbf{X} - \mathbf{Z})\|_F^2 \quad (17)$$

则问题(16)可等价描述为

$$(\mathbf{X}^*, \mathbf{Z}^*) = \arg \min_{\mathbf{X}, \mathbf{Z} \in \mathbb{R}^{\eta_1 \times \eta_2}} L(\mathbf{X}, \mathbf{Z}) \quad (18)$$

上述问题不存在解析解,但可以采用优化领域常用的交替迭代法对其进行求解,即迭代求解如下2个子问题

$$\begin{cases} \mathbf{X}^{k+1} = \arg \min_{\mathbf{X} \in \mathbb{R}^{\eta_1 \times \eta_2}} L(\mathbf{X}, \mathbf{Z}^k) \\ \mathbf{Z}^{k+1} = \arg \max_{\mathbf{Z} \in \mathbb{R}^{\eta_1 \times \eta_2}} L(\mathbf{X}^{k+1}, \mathbf{Z}) \end{cases}$$

下面将分别给出这2个子问题的求解方案。

子问题1求解 $\mathbf{X}^{k+1} = \arg \min_{\mathbf{X} \in \mathbb{R}^{\eta_1 \times \eta_2}} L(\mathbf{X}, \mathbf{Z}^k)$

将式(17)代入子问题1后,得

$$\begin{aligned} \mathbf{X}^{k+1} &= \arg \min_{\mathbf{X} \in \mathbb{R}^{\eta_1 \times \eta_2}} \{ \mu(\|\mathbf{X}\|_* + \lambda\|\mathbf{Z}^k\|_{2,1}) + \\ &\quad \frac{1}{2}\|P_\Omega(\mathbf{R} - \mathbf{X} - \mathbf{Z}^k)\|_F^2 \} \\ &= \arg \min_{\mathbf{X} \in \mathbb{R}^{\eta_1 \times \eta_2}} \mu\|\mathbf{X}\|_* + \frac{1}{2}\|P_\Omega(\mathbf{R} - \mathbf{X} - \mathbf{Z}^k)\|_F^2 \end{aligned}$$

不妨令 $F_2(\mathbf{X}) = \frac{1}{2}\|P_\Omega(\mathbf{R} - \mathbf{X} - \mathbf{Z}^k)\|_F^2$, $F_1(\mathbf{X}) = \mu\|\mathbf{X}\|_*$, 易知 F_1, F_2 同为下半连续的凸函数, F_2 在 $\mathbb{R}^{\eta_1 \times \eta_2}$ 中可微且具有1-Lipschitz连续梯度, 并且 F_1, F_2 还同时满足定理2的性质1)和性质2), 因此根据定理2性质3), 可知子问题1存在唯一解, 且对任意的初始值 \mathbf{X}^0 及 $0 < \delta_x < 2$ 采用如下方法生成的迭代序列 \mathbf{X}^{k+1} 收敛到子问题1的唯一解

$$\mathbf{X}^{k+1} = \arg \min_{\mathbf{X} \in \mathbb{R}^{\eta_1 \times \eta_2}} \{ \delta_x \mu\|\mathbf{X}\|_* + \frac{1}{2}\|(\mathbf{X} - \mathbf{X}^k - \delta_x P_\Omega(\mathbf{R} - \mathbf{X}^k - \mathbf{Z}^k))\|_F^2 \}$$

此时, 如果令 $\mathbf{U}^k = \mathbf{X}^k + \delta_x P_\Omega(\mathbf{R} - \mathbf{X}^k - \mathbf{Z}^k)$, 则上式可简化为

$$\mathbf{X}^{k+1} = \arg \min_{\mathbf{X} \in \mathbb{R}^{\eta_1 \times \eta_2}} \mu\delta_x \|\mathbf{X}\|_* + \frac{1}{2}\|\mathbf{X} - \mathbf{U}^k\|_F^2$$

根据定理1可得 $\mathbf{X}^{k+1} = D_{\mu\delta_x}(\mathbf{U}^k)$

因此, 子问题1可按式(19)迭代求解。

$$\begin{cases} \mathbf{U}^k = \mathbf{X}^k + \delta_x P_\Omega(\mathbf{R} - \mathbf{X}^k - \mathbf{Z}^k) \\ \mathbf{X}^{k+1} = D_{\mu\delta_x}(\mathbf{U}^k) \end{cases} \quad (19)$$

子问题2求解 $\mathbf{Z}^{k+1} = \arg \max_{\mathbf{Z}} L(\mathbf{X}^{k+1}, \mathbf{Z})$

将式(17)代入子问题2后, 得

$$\begin{aligned} \mathbf{Z}^{k+1} &= \arg \min_{\mathbf{Z} \in \mathbb{R}^{\eta_1 \times \eta_2}} \{ \mu(\|\mathbf{X}^{k+1}\|_* + \lambda\|\mathbf{Z}\|_{2,1}) + \\ &\quad \frac{1}{2}\|P_\Omega(\mathbf{R} - \mathbf{X}^{k+1} - \mathbf{Z})\|_F^2 \} \\ &= \arg \min_{\mathbf{Z} \in \mathbb{R}^{\eta_1 \times \eta_2}} \mu\lambda\|\mathbf{Z}\|_{2,1} + \frac{1}{2}\|P_\Omega(\mathbf{R} - \mathbf{X}^{k+1} - \mathbf{Z})\|_F^2 \end{aligned}$$

同理, 根据定理2, 可知

$$\begin{aligned} \mathbf{Z}^{k+1} &= \arg \min_{\mathbf{Z} \in \mathbb{R}^{\eta_1 \times \eta_2}} \{ \mu\lambda\delta_z \|\mathbf{Z}\|_{2,1} + \\ &\quad \frac{1}{2}\|\mathbf{Z} - \mathbf{Z}^k - \delta_z P_\Omega(\mathbf{R} - \mathbf{X}^{k+1} - \mathbf{Z}^k)\|_F^2 \} \end{aligned}$$

如果令 $\mathbf{V}^k = \mathbf{Z}^k + \delta_z P_\Omega(\mathbf{R} - \mathbf{X}^{k+1} - \mathbf{Z}^k)$, 则

$$\mathbf{Z}^{k+1} = \arg \min_{\mathbf{Z} \in \mathbb{R}^{\eta_1 \times \eta_2}} \mu\lambda\delta_z \|\mathbf{Z}\|_{2,1} + \frac{1}{2}\|\mathbf{Z} - \mathbf{V}^k\|_F^2$$

进一步, 根据定理3可得

$$\begin{aligned} (\mathbf{Z}^{k+1})^{(i)} &= \max \{ 1 - \mu\lambda\delta_z / \|(\mathbf{V}^k)^{(i)}\|_2, 0 \} (\mathbf{V}^k)^{(i)}, \\ &\quad i = 1, 2, \dots, n_1 \end{aligned}$$

因此, 子问题2可按如下方式迭代求解

$$\begin{cases} \mathbf{V}^k = \mathbf{Z}^k + \delta_z P_\Omega(\mathbf{R} - \mathbf{X}^{k+1} - \mathbf{Z}^k) \\ (\mathbf{Z}^{k+1})^{(i)} = \max \{ 1 - \mu\lambda\delta_z / \|(\mathbf{V}^k)^{(i)}\|_2, 0 \} (\mathbf{V}^k)^{(i)}, \\ i = 1, 2, \dots, n_1 \end{cases} \quad (20)$$

至此, 可以将子问题1和子问题2的求解算法归纳起来, 整理得到式(16)的迭代求解算法, 并将其命名为基于算子分裂技术的结构化噪声矩阵补全算法(OSMCSN), 详细步骤如算法1所述。

算法1 基于算子分裂技术的结构化噪声矩阵补全算法(OSMCSN)

输入: $P_\Omega(\mathbf{R})$, μ , λ 和 δ_x , δ_z , 最大迭代次数 $Maxiter$

输出: $\mathbf{X}^{\text{opt}}, \mathbf{Z}^{\text{opt}}$

- 1) INITIALIZE $\mathbf{X}^0 = 0, \mathbf{Z}^0 = 0$;
- 2) FOR $k = 0$ to $Maxiter$
- 3) $\mathbf{U}^k = \mathbf{X}^k + \delta_x P_\Omega(\mathbf{R} - \mathbf{X}^k - \mathbf{Z}^k)$;
- 4) $\mathbf{X}^{k+1} = D_{\mu\delta_x}(\mathbf{U}^k)$;
- 5) $\mathbf{V}^k = \mathbf{Z}^k + \delta_z P_\Omega(\mathbf{R} - \mathbf{X}^{k+1} - \mathbf{Z}^k)$;
- 6) FOR $i = 1$ to n_1

- 7) $(\mathbf{Z}^{k+1})^{(i)} = \max\{1 - \mu\lambda\delta_z / \|\mathbf{V}^k\|_2, 0\}(\mathbf{V}^k)^{(i)}$;
- 8) END FOR
- 9) END FOR
- 10) RETURN $\mathbf{X}^{\text{opt}}, \mathbf{Z}^{\text{opt}}$

从算法 1 容易看出, OSMCSN 算法在迭代过程中一方面可使 \mathbf{X} 保持较好的低秩性, 另一方面还可使 \mathbf{Z} 保持良好的行稀疏特性, 而且每一次迭代过程只涉及到一次矩阵的部分奇异值分解, 而对于大型矩阵的部分奇异值分解可以采用成熟的 PROPACK 软件包^[20], 这些特性确保了 OSMCSN 算法的高效率和可扩放性。

6 数值实验及结果分析

本节将首先对 OSMCSN 算法参数进行分析, 然后采用人工合成数据来验证 OSMCSN 算法的性能, 最后将 OSMCSN 算法应用于 Web 服务 QoS 预测。文中所有实验都是在 Intel Core i7-3520M 2.90 GHz CPU, 4.00 GB RAM 的环境下采用 MATLAB 完成。

6.1 OSMCSN 算法参数分析

从算法 1 易知影响 OSMCSN 算法性能的参数有 λ 、 μ 和 δ_x 、 δ_z 。目前还没有展开对这些可调参数如何自适应设置的理论研究, 只能利用所处理问题的先验知识对可调参数进行交叉验证。但尽管如此, 仍然可以对这些参数的设置做一些非形式化分析。

1) 参数 δ_x 和 δ_z 在算法 1 的子问题求解过程中被引入。对于子问题 1, 如果令 $F_2(\mathbf{X}) = (1/2)\|P_\Omega(\mathbf{R} - \mathbf{X} - \mathbf{Z}^k)\|_F^2$, 显然 $F_2(\mathbf{X})$ 在 $\mathbb{R}^{n_1 \times n_2}$ 中可微且存在常数 $\beta=1$, 使

$$\begin{aligned} & \|\nabla F_2(\mathbf{U}) - \nabla F_2(\mathbf{V})\|_F \\ &= \|P_\Omega(\mathbf{U} + \mathbf{Z}^k - \mathbf{R}) - P_\Omega(\mathbf{V} + \mathbf{Z}^k - \mathbf{R})\|_F \\ &= \|P_\Omega(\mathbf{U} - \mathbf{V})\|_F \leq 1\|\mathbf{U} - \mathbf{V}\|_F \end{aligned}$$

即函数 $F_2(\mathbf{X})$ 具有 1-Lipschitz 连续梯度, 根据定理 2 可知, 参数 δ_x 在区间 (0,2) 任意取值, 子问题 1 的解序列均能收敛到唯一解; 同理易知参数 δ_z 也可在区间 (0,2) 任意取值, 子问题 2 的解序列同样能收敛到唯一解。

2) 参数 λ 在 OSMCSN 算法中被用来调节结构化噪声含量大小, 一般而言, 设定 $\lambda \in (0,1)$ 。从式 (16) 可以看出, 在矩阵秩恒定的情况下, 采样矩

阵无结构化噪声或结构化噪声越小, λ 取值应越大, 反之亦然。

3) 参数 μ 在 OSMCSN 算法中有 2 个作用, 一是被用来调节结果矩阵的低秩程度, 二是被用来平衡 $(\|\mathbf{X}\|_* + \lambda\|\mathbf{Z}\|_{2,1})$ 项和 $(1/2)\|P_\Omega(\mathbf{R} - \mathbf{X} - \mathbf{Z})\|_F^2$ 项的大小。从算法 1 步骤 4) 可以看出, 如果希望得到结果矩阵的秩越低, 则参数 μ 的取值应越大, 但是从式 (16) 容易看出, 此时采样矩阵所能容忍的非结构化噪声也越大。通常在实际问题中, 结果矩阵都是低秩或近似低秩的, 同时由于包含结构化噪声的采样矩阵往往不包含或仅包含很少的非结构化噪声, 因此 $(1/2)\|P_\Omega(\mathbf{R} - \mathbf{X} - \mathbf{Z})\|_F^2$ 项的值往往非常小。因此在实际的算法实现中, 并不固定参数 μ 的取值, 而是借鉴 FPCA 算法^[16]的实现技巧, 设置参数 μ 的初值为采样矩阵的 L2 范数, 然后让其以 0.25 的速率迭代到 0.001。这样既满足了结果矩阵的低秩性, 也同时使析出的非结构化噪声符合采样矩阵的实际含噪情形。

6.2 OSMCSN 算法性能测试

本节主要通过合成数据来验证 OSMCSN 算法的性能。考虑到加噪和采样过程的随机性, 本节所有实验结果都是 20 次独立重复实验的平均值, OSMCSN 算法的收敛条件设置为最大迭代次数 150。第一个实验考察 OSMCSN 算法对结构化噪声行位置信息的辨识性能; 第二个实验比较移除结构化噪声行前后 OSMCSN 算法和 SVT、FPCA 算法对缺失元素的补全精度。为此分别针对不同的矩阵维度 $n_1 \times n_2$ 和秩 r , 通过产生 2 个 $n_1 \times r$ 和 $r \times n_2$ 的独立同分布高斯随机矩阵 \mathbf{M}_L 和 \mathbf{M}_R (均值为 0, 方差为 1) 生成秩为 r 的矩阵 $\mathbf{M} = \mathbf{M}_L \mathbf{M}_R$; 然后对矩阵 \mathbf{M} 添加噪声污染程度为 p_n 的结构化噪声 (令 $p_n = m/n_1$, 其中 m 表示受到噪声污染元素的行数), 我们称加噪后的矩阵为矩阵 \mathbf{M}' ; 采样元素均匀随机取自 \mathbf{M}' , 采样元素的下标集合用 Ω 表示, 并记采样率为 p_s , 其中 $p_s = |\Omega|/(n_1 n_2)$, $|\Omega|$ 表示集合 Ω 的基数。实验中选取如下 2 个评价指标来衡量 OSMCSN 算法的性能。

1) 结构化噪声行位置信息识别率

$$r_{\text{line}} = \frac{2 \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

其中,

$$precision = \frac{m_true}{m_all}, \quad recall = \frac{m_true}{m}$$

m_all 表示 OSMCSN 算法输出的噪声矩阵 \mathbf{Z}^{opt} 中含有非零元素行的总数, m_true 表示矩阵 \mathbf{Z}^{opt} 中含有非零元素行所对应的 \mathbf{M}' 矩阵中真实含噪的行数。

2) 缺失元素补全精度

$$r_{miss} = \frac{\|P_{\Omega^c}(\mathbf{X}^{opt} - \mathbf{M})\|_F}{\|P_{\Omega^c}(\mathbf{M})\|_F}$$

其中, \mathbf{X}^{opt} 表示 OSMCSN 算法补全后的结果矩阵, Ω^c 表示缺失元素索引集合。

实验 1 不同情形下 OSMCSN 算法对结构化噪声行位置信息的辨识性能

为了验证不同情形下 OSMCSN 算法对结构化噪声行位置信息的辨识性能, 针对 (n_1, n_2, r, p_s, p_n) 设计了 5 组不同的实验。其中参数 μ 的初值设置为采样矩阵的 L2 范数, 然后让其以 1/4 的速率迭代减小到 0.001。实验结果如表 1 所示, 从表 1 可以看出, OSMCSN 算法能准确识别出结构化噪声所在行的位置, 而有了这些噪声行位置信息, 可以在补全矩阵的缺失元素时, 预先把采样矩阵的含噪数据行去除, 然后在此基础上对不含噪部分的缺失元素进行补全。

实验 2 移除结构化噪声行前后 OSMCSN 算法和 SVT、FPCA 算法对缺失元素的补全精度比较
实验 2 进一步设计了 5 组不同实验来考察移除

结构化噪声行前后 OSMCSN 算法和 SVT、FPCA 算法对缺失元素的补全精度比较。其中, SVT 算法的实现代码下载于 <http://svt.stanford.edu/code.html>, FPCA 算法的实现代码下载于 <http://www1.se.cuhk.edu.hk/~sqma/FPCA.html>。SVT 算法和 FPCA 算法的参数按照算法设计者设定的方法设置, OSMCSN 算法的参数经过交叉验证后选取最优设置。实验结果见表 2。从表 2 可以看出: 1) 移除结构化噪声前, OSMCSN 算法对缺失元素的补全精度明显优于 SVT 和 FPCA 算法; 2) 移除结构化噪声后, 采样矩阵变成了无噪声采样, 此时无论是 OSMCSN 算法还是 SVT 和 FPCA 算法, 对缺失元素的补全精度均得到了几何级数的提高。

6.3 基于 OSMCSN 算法的 Web 服务 QoS 预测

在这一节采用 OSMCSN 算法来预测 Web 服务 QoS。选用香港中文大学 Zheng Zibing 博士提供的 QoS 数据集 (<http://www.wsdream.net>), 现有的 Web 服务 QoS 预测研究大多基于该数据集。该数据集包括 339 个用户调用 5 825 个 Web 服务后所反馈的响应时间 (response-time) 和吞吐量 (throughput) 信息。然而, 由于不是每一次服务调用都是成功的, 因此该数据集所收集的 QoS 值实际上也是不完全的。为了能全面客观地评价本算法的 QoS 预测性能, 从中筛选出 1 300 个已被所有用户成功调用过的服务作为实验数据, 这样得到 2 个 $339 \times 1 300$

表 1 不同情形下 OSMCSN 算法对结构化噪声行位置信息的辨识性能

n_1	n_2	r	p_s	p_n	λ	δ_x	δ_z	r_{line}
300	400	5	0.20	0.10	0.95	1.5	1.5	100%
300	400	5	0.20	0.50	0.80	1.5	1.5	100%
500	500	10	0.35	0.15	0.90	1.5	1.5	100%
600	500	15	0.45	0.40	0.80	1.5	1.5	100%
1 000	1 000	50	0.30	0.30	0.85	1.5	1.5	100%

表 2 移除结构化噪声行前后 SVT、FPCA 与 OSMCSN 算法对缺失元素的补全精度比较

n_1	n_2	r	p_s	p_n	SVT		FPCA		OSMCSN	
					移除噪声行前	移除噪声行后	移除噪声行前	移除噪声行前	移除噪声行前	移除噪声行后
300	400	5	0.45	0.25	3.58×10^{-1}	1.62×10^{-5}	5.66×10^{-1}	2.97×10^{-5}	3.57×10^{-2}	8.71×10^{-6}
500	300	5	0.35	0.15	3.17×10^{-1}	1.57×10^{-5}	5.44×10^{-1}	2.08×10^{-5}	2.33×10^{-2}	9.62×10^{-6}
500	500	10	0.45	0.25	3.32×10^{-1}	1.56×10^{-5}	4.19×10^{-1}	2.31×10^{-5}	2.47×10^{-2}	6.23×10^{-6}
1 000	1 000	15	0.30	0.30	4.98×10^{-1}	9.85×10^{-5}	4.58×10^{-1}	1.02×10^{-5}	2.45×10^{-2}	5.18×10^{-6}
1 500	1 000	10	0.30	0.10	3.04×10^{-1}	2.31×10^{-5}	3.23×10^{-1}	3.28×10^{-5}	1.59×10^{-2}	3.32×10^{-6}

的不包含 null 元素的用户-服务 QoS 矩阵，即响应时间 QoS 矩阵和吞吐量 QoS 矩阵。表 3 概括了这 2 个 QoS 矩阵的统计信息。此外，由于矩阵补全算法适用的前提是目标矩阵必须低秩或者近似低秩，为了验证响应时间 QoS 矩阵和吞吐量 QoS 矩阵的低秩性，分别考察了这 2 个矩阵的奇异值分布情况。图 1 描述了已规一化到[0, 1]区间的响应时间 QoS 矩阵和吞吐量 QoS 矩阵各前 100 个奇异值分布情况，从图 1 可以看出它们的奇异值分布基本呈幂式衰减，从而表明这 2 个 QoS 矩阵是近似低秩的。进一步观察图 1 还能看出，吞吐量 QoS 矩阵的低秩性略优于响应时间 QoS 矩阵的低秩性。

表 3 吞吐量 QoS 矩阵和响应时间 QoS 矩阵统计信息

统计信息	吞吐量 QoS 矩阵	响应时间 QoS 矩阵
用户数量	339	339
服务数量	1 300	1 300
取值区间	0~1 000 kbit/s	0~20 s
均值	28.026 3 kbit/s	0.628 5 s

接下来，将所提出的 OSMCSN 算法分别应用于响应时间 QoS 和吞吐量 QoS 预测，并与现有算法 UIPCC^[7]、ADF^[8]、NIMF^[9]和 TNR-MF^[10]进行比较。选择规一化均方误差 (NMAE, normalized mean absolute error) 作为 QoS 预测性能评价指标

$$NMAE = \frac{\|P_{\Omega^c}(\tilde{M} - M)\|_F}{\|P_{\Omega^c}(M)\|_F}$$

其中， \tilde{M} 表示采用 QoS 预测算法预测出的结果 QoS 矩阵， Ω^c 表示缺失元素索引集合。从规一化均方误差 NMAE 和第 6.2 节缺失元素补全精度 r_{miss} 的定义容易看出，这 2 个定义其实是一致的。为了使实验能更好地契合 Web 服务 QoS 预测的真实情况，首先在 339 个用户中随机选择 34 位用户，在他们所反馈的 QoS 中再随机选取 50%的数据添加随机噪声（其中响应时间 QoS 添加的噪声范围为[0, 2]，吞吐量 QoS 添加的噪声范围为[0, 100]）。再对加噪后的 QoS 矩阵分别随机采样 5%、10%、15%、20%、25%、30%、35%和 40%，从而分别得到包含 10%结构化噪声的 QoS 采样矩阵，然后基于这些采样矩阵分别对不含噪部分的缺失 QoS 进行预测。本节所有实验均独立重复进行 40 次，其中前 20 次添加的是均匀随机噪声，后 20 次添加的是非均匀随机噪声，本小节所有实验结果都是 40 次独立重复实验的平均值。

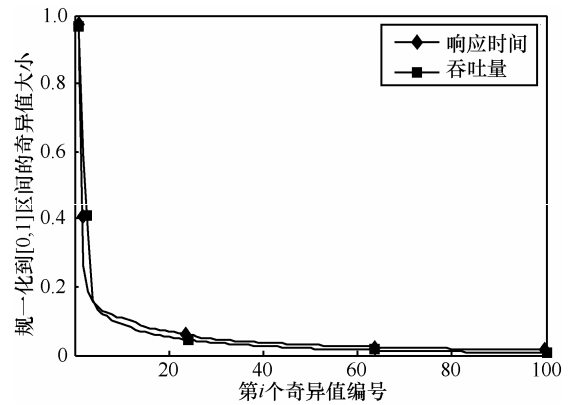


图 1 吞吐量 QoS 矩阵和响应时间 QoS 矩阵的奇异值分布示意

这里需要特别指出的是，预测实验分 2 步进行：第一步采用 OSMCSN 算法识别结构化噪声所在行的位置信息；第二步根据噪声行位置信息在采样矩阵中移除噪声行，然后再采用 OSMCSN 算法对缺失的 QoS 进行预测。表 4 给出了 5%~25%采样率情形下结构化噪声行位置信息的识别结果，图 2 和图 3 则给出了不同算法对吞吐量和响应时间 QoS 预测性能的比较结果。从表 4 容易看出，各种采样率情形下的结构化噪声行位置信息识别率都达到了 100%，这说明应用 OSMCSN 算法可以精确探测出提供虚假 QoS 信息的不诚实用户，从而可以实现对操纵和篡改 QoS 信息的不诚实用户进行预警。同时，从图 2 和图 3 可以看出：1) 随着采样率的提高，所有算法的预测精度也有所提高，这和我们的直觉是相吻合的，因为相对稠密的采样为缺失数据的预测提供了更为丰富的信息；2) 采样率无论高低，OSMCSN 算法在缺失元素补全方面的性能均明显优于其他算法。此外也注意到，虽然图 2 和图 3 的实验结果表明提出的 OSMCSN 算法对于 Web 服务 QoS 的预测性能明显优于其他算法，在一定程度上是可行的，但是，相对于表 2 给出的该算法在合成数据上的精确补全性能还有很大的差距，究其原因，主要在于 Web 服务 QoS 矩阵并不是严格低秩的。

表 4 不同采样率情形下结构化噪声行位置信息识别性能

P_s	吞吐量识别率	响应时间识别率
5%	100%	100%
10%	100%	100%
15%	100%	100%
20%	100%	100%
25%	100%	100%

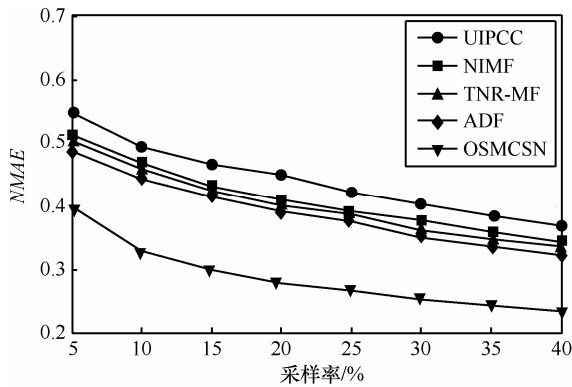


图2 吞吐量 QoS 预测性能比较

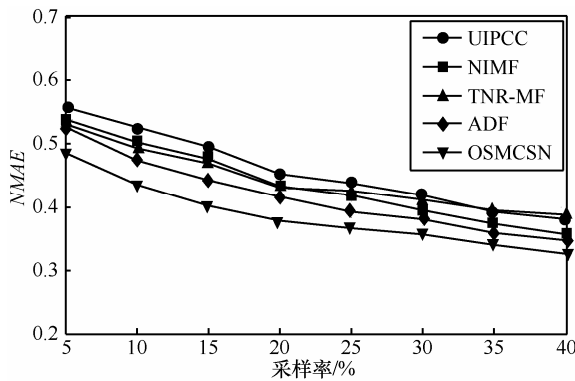


图3 响应时间 QoS 预测性能比较

此外, 为了进一步考察 OSMCSN 算法辨识噪声行位置信息的顽健性, 在采样率为 20% 的情形下对吞吐量 QoS 矩阵执行了不同结构化噪声污染程度的识别实验, 采用的性能评价指标包括 *recall*、*precision* 和 r_{line} 。实验结果如表 5 所示。

表 5 不同结构化噪声污染程度情形下噪声行位置信息识别性能 (采样率为 20% 时的吞吐量 QoS 矩阵)

p_n	r_{line}	<i>precision</i>	<i>recall</i>
5%	100%	100%	100%
10%	100%	100%	100%
15%	100%	100%	100%
20%	98.5%	97.1%	100%
25%	82.3%	71.8%	96.4%

从表 5 可以看出, 当结构化噪声含量低于 15% 时, r_{line} 识别率都能达到 100%; 当结构化噪声含量增加到 20% 时, r_{line} 识别率只达到了 98.5%, 但是其查全率依然达到了 100%, 这意味着仍然可以将那些提供虚假 QoS 信息的不诚实用户全部识别出来; 而当结构化噪声含量达到 25% 时, 无论是查准率还是查全率都达不到 100%, 但是认为这并不妨碍本算法的实用性, 因为在实际应用中, 结构化噪声含量一般都不会超过

20%。最后, 需要特别指出的是, 虽然本文仅仅考察了响应时间和吞吐量 QoS 的预测, 实际上本文方法对其他 QoS 也是适用的。

7 结束语

Web 服务 QoS 预测是进行个性化服务选择和服务组合时一个非常关键的步骤。本文提出了一类基于结构化噪声矩阵补全的 Web 服务 QoS 预测方法。实验结果表明, 该方法不仅可以显著提高 Web 务 QoS 预测的精度, 还能准确探测出采样矩阵中的噪声行位置信息, 从而可以实现对操纵和篡改 QoS 信息的不诚实用户进行预警。但是, 本文目前仅对响应时间和吞吐量这 2 个 QoS 进行了验证, 而实际上有大量非常重要的 QoS 并未涉及到。本文的后续工作将包括对这些重要属性的验证。

参考文献:

- [1] XIA Y, CHEN P, BAO L, *et al.* A QoS-aware web service selection algorithm based on clustering[A]. Proceedings of the 9th International Conference on Web Service[C]. Washington, DC, USA, 2011. 428-435.
- [2] 王尚广, 孙其博, 杨放春. 基于全局 QoS 约束分解的 Web 服务动态选择[J]. 软件学报, 2011, 22(7): 1426-1439.
WANG S G, SUN Q B, YANG F C. Web service dynamic selection by the decomposition of global QoS constraints[J]. Journal of Software, 2011, 22(7): 1426-1439.
- [3] 邓水光, 黄龙涛, 吴斌等. 一种 QoS 最优的语义 Web 服务自动组合方法[J]. 计算机学报, 2013, 36(5): 1015-1030.
DENG S G, HUANG L T, WU B. QoS optimal automatic composition of semantic Web services[J]. Chinese Journal of Computers, 2013, 36(5): 1015-1030.
- [4] PAREJO J A, SEGURA S, FERNANDEZ P. QoS-aware web services composition using GRASP with path relinking[J]. Expert Systems with Applications, 2014, 41(9): 4211-4223.
- [5] SHAO L S, ZHANG J, WEI Y, *et al.* Personalized QoS prediction for Web services via collaborative filtering[A]. Proceedings of the 5th International Conference on Web Service[C]. Salt Lake City, USA, 2007. 439-446.
- [6] 邵凌霄, 周立, 赵俊峰等. 一种 Web Service 的服务质量预测方法[J]. 软件学报, 2009, 20(8): 2063-2073.
SHAO L S, ZHOU L, ZHAO J F, *et al.* Web service QoS prediction approach[J]. Journal of Software, 2009, 20(8): 2062-2073.
- [7] ZHENG Z B, MA H, LYU M R, *et al.* QoS-aware web service recommendation by collaborative filtering[J]. IEEE Transactions on Service Computing, 2011, 4(2): 140-152.
- [8] WU J, CHEN L, FENG Y P, *et al.* Predicting quality of service for selection by neighborhood-based collaborative filtering[J]. IEEE Transactions on Systems, Man and Cybernetics: Systems, 2013, 43(2): 428-439.
- [9] ZHENG Z, MA H, LYU M, *et al.* Collaborative web service QoS prediction via neighborhood integrated matrix factorization[J]. IEEE Transactions on Service Computing, 2013, 6(3): 289-299.
- [10] YU Q, ZHENG Z. B. Trace norm regularized matrix factorization for

- service recommendation[A]. Proceedings of the 11th International Conference on Web Service[C]. Santa Clara, CA, USA, 2013. 34-41.
- [11] CAI X, NIE F P, HUANG H. *et al.* Multi-class L_{2,1}-norm support vector machine[A]. Proceedings of the 11th International Conference on Data Mining[C]. Vancouver, BC, 2011. 91-100.
- [12] COMBETTES P L, WAJS V R. Signal recovery by proximal forward-backward splitting[J]. Multiscale Modeling and Simulation, 2005, 4(4): 1168-1200.
- [13] BOYD S, VANDENBERGHE L. Convex optimization[M]. Cambridge University Press, 2009.
- [14] CAI J F, CANDÈS E J, SHEN Z. A singular value thresholding algorithm for matrix completion[J]. SIAM Journal of Optimization, 2010, 20(4): 1956-1982.
- [15] CANDÈS E J, RECHT B. Exact matrix completion via convex optimization[J]. Foundations of Computational Mathematics, 2009, 9(6): 717-772.
- [16] MA S, GOLDFARB D, CHEN L. Fixed point and Bregman iterative methods for matrix rank minimization[J]. Mathematics Programming Series A, 2011, 128(1): 321-353.
- [17] LIN Z, CHEN M, MA Y. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices[R]. UIUC University, Technical report, UILU-ENG-09-2214, 2010.
- [18] WEN Z W, YIN W T, ZHANG Y. Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm[R]. Rice University, Technical Report TR1007, 2010.
- [19] BOUMAL N, ABSIL P. RTRMC: a riemannian trust region method for matrix completion[A]. Proceedings of the 25th Annual Conference on Neural Information Processing Systems[C]. Granada, Spain, 2011. 406-414.
- [20] LARSEN R M. PROPACK-software for large and sparse SVD calculations[EB/OL]. <http://soi.stanford.edu/~rmunk/PROPACK/>.

作者简介:



陈蕾 (1975-), 男, 江西宜春人, 博士, 南京邮电大学副教授、硕士生导师, 主要研究方向为可信服务计算、机器学习、数据挖掘。



杨庚 (1961-), 男, 江苏建湖人, 博士, 南京邮电大学教授、博士生导师, 主要研究方向为计算机通信与网络、并行与分布式计算、信息安全。

陈正宇 (1978-), 男, 江苏淮安人, 南京邮电大学博士生, 主要研究方向为无线传感网数据融合、服务计算和隐私保护。

肖甫 (1980-), 男, 湖南邵阳人, 博士, 南京邮电大学教授、博士生导师, 主要研究方向为无线传感器网络、云计算和多媒体图像检索等。

许建 (1980-), 男, 江苏徐州人, 南京邮电大学副教授, 主要研究方向为无线传感器网络与信息安全、智能信息处理。