

doi:10.11835/j.issn.1000-582X.2021.01.010

时空感知下基于结构相似度的 Web 服务质量预测

夏 会¹, 高 旻², 邹 淑¹

(1. 重庆理工大学 会计学院, 重庆 400054; 2. 重庆大学 大数据与软件学院, 重庆 400044)

摘要:随着云计算等新型服务计算的兴起, Web 服务数量日益增长, 相同或相似功能的 Web 服务也逐渐增多。为了向用户推荐更高质量的服务, 精确地预测 Web 服务的 QoS 值成为亟待解决的重要问题。传统的协同过滤方法已经被广泛应用于 QoS 预测和 Web 服务推荐中, 但因为数据稀疏和噪声问题导致 QoS 预测性能不好。为提高 QoS 预测的性能, 文中通过分析用户-服务 QoS 矩阵的时空特征, 提出了一种基于全局和局部结构相似度的稀疏矩阵分解模型。该方法将 QoS 矩阵的相邻时间相似用户的网络环境相似性这一特征融入到矩阵分解中, 并利用分解的因子对 QoS 矩阵进行低秩填充。这种方式在一定程度上消除了数据稀疏和噪声的影响。在真实 Web 服务调用数据集上进行实验, 结果表明, 该方法在预测精度上优于典型的协同过滤算法(相比于 NMF, 其 MAE 值最大下降了 3.25%, RMSE 值最大下降了 6.65%; 相比于 SVD, 其 MAE 值最大下降了 3.67%, RMSE 值最大下降了 7.01%), 能够有效地解决数据稀疏和噪声的问题。

关键词: QoS 预测; 时空感知; 数据稀疏; 矩阵分解; 服务计算

中图分类号: TP301

文献标志码: A

文章编号: 1000-582X(2020)01-088-09

A structure similarity based quality prediction approach for Web service in the spatial-temporal scenario

XIA Hui¹, GAO Min², ZOU Shu¹

(1. School of Accounting, Chongqing University of Technology, Chongqing 400054, P. R. China;

2. School of Big Data and Software, Chongqing University, Chongqing 400044, P. R. China)

Abstract: With the rapid development of new service computing types like cloud computing, the amount of Web services becomes increasingly massive, as is also the case for Web services with the same or similar functions. One of the most important issues in high-quality Web services recommendation is to identify the QoS value of Web services. Traditional collaborative filtering approaches have been widely employed in QoS prediction and Web Service recommendation. However, they suffer from the sparse and noisy data issues, which definitely cause the low performance of QoS predictions. In order to attain high prediction performance, the paper proposed a novel sparse matrix factorization approach based on the spatial-temporal

收稿日期: 2020-06-07

基金项目: 重庆市教育委员会科学技术项目(KJQN201801103); 重庆市社会科学规划项目(2018BS68); 重庆市教育委员会人文社会科学项目(20SKGH176)。

Supported by Chongqing Education Commission Science and Technology Program (KJQN201801103), Chongqing Social Science Planning Program (2018BS68), and Chongqing Education Commission Humanities and Social Science Program (20SKGH176).

作者简介: 夏会(1989—), 女, 博士, 主要从事个性化推荐、服务计算、云会计、数据挖掘方向研究, (E-mail)summertulip@126.com。

features of user-service QoS matrix, in which the proposed model took the occurrence of similar network environment between similar users in neighbor moment into consideration, and constructed a global and local structure similarity based sparse matrix factorization machine. With the decomposed factors, we could get a low rank matrix completion, which helped to eliminate the impacts of sparse and noisy data in QoS prediction. To evaluate the performance of the proposed approach, a set of extensive experiments were conducted using real-world dataset. The experimental results show that the proposed model outperforms the traditional collaborative filtering approaches (the MAE value decreases by 3.25% and RMSE value by 6.65% compared with those of NMF; MAE value decreases by 3.67%, and RMSE by 7.01% compared with those of SVD), which indicates that it can effectively resolve sparse and noisy data issues.

Keywords: QoS prediction; spatial-temporal-aware; data sparsity; matrix factorization; service computing

Web 服务是一种平台独立、低耦合、可编程的 Web 应用程序,具有良好的互操作性,广泛用于开发大规模的分布式应用程序。随着信息技术的迅猛发展,特别是云计算等新型服务计算的兴起,软件应用、计算能力、数据等许多资源都被打包成服务,Web 服务的数量呈指数级增长,大量功能相同或相似的 Web 服务不断涌现^[1]。用户因为缺乏相应的专业知识而无法作出选择,及时准确地为用户发现和选择高质量的服务成为服务计算领域亟待解决的问题^[2]。

个性化推荐技术作为缓解信息过量产生和有效信息发现不平衡的有效方法之一^[3],被引入到 Web 服务领域,可以及时高效地为用户提供满足需求的 Web 服务,提高用户的满意度^[4]。个性化服务推荐技术的一个重要问题是如何获得 Web 服务的质量(Quality of Service, QoS)。QoS 作为描述 Web 服务的主要非功能性特征,包括 Web 服务的响应时间、吞吐量、价格和可靠性等^[5]。精确的 QoS 可以有效地提高 Web 服务推荐的性能^[6]。尽管用户可以通过亲自调用 Web 服务来评估 QoS,但是要在短时间对大量候选服务的 QoS 进行准确评估是不太可能的^[7]。与此同时,网络环境的不确定性以及恶意的用户反馈可能导致用户调用 Web 服务时的 QoS 是虚假的,包含噪声的^[8]。实际应用中,用户对 Web 服务的 QoS 矩阵通常是稀疏的,包含噪声的,对 Web 服务推荐影响较大。由于 Web 服务的 QoS 属性受到用户和 Web 服务所在的位置、访问时间以及网络环境等因素的影响,当前的研究多在已有的基于用户或服务协同过滤推荐^[5,9-10]的基础上,利用时间和位置的特征^[2,11-14],对 QoS 的预测结果进行优化,以更好地进行协同过滤推荐。

笔者在上述工作的基础上,总结 Web 服务的时空特征为:短时间内,相似用户通常具有相似的网络环境和网络性能,进而有较大可能在相同的 Web 服务上观察到相似的性能。因此,得出用户-服务的 QoS 矩阵具有结构相似性。

1) 全局结构相似性。考虑到用户的网络行为在短时间(如 s、min、h)内通常变化不明显(在 2 h 内用户一直使用视频类软件看电影),即存在用户在较短时间内使用相同的或者类似的 Web 服务的现象,相邻两个时间段(s、min、h)内的用户-服务 QoS 矩阵的全局结构是相似的^[14]。这种全局结构相似性是时间相似性的一种体现。

2) 局部结构相似性。相邻用户通常处于同一自治系统,或者处于相邻自治系统内,其在相邻时间段内具有相同或相似网络环境的可能性较大,有可能对相同服务的 QoS 反馈是相似的^[15]。QoS 矩阵在相邻用户上具有局部的结构相似性,这种局部结构相似性本质上反映了用户位置的相似性。

文中将上述特征融入到对 QoS 的预测上,提出了基于全局和局部结构相似性的稀疏矩阵分解模型(Global and Local Structure Similarity based Sparse Matrix Factorization Machine, GLMF),以达到提升 QoS 预测精度的目标。①将 QoS 矩阵的时空特征与矩阵分解相结合,提出了一种基于全局和局部结构相似性的稀疏矩阵分解模型。在矩阵分解时,通过保留其与前一时间 QoS 值的全局相似性信息,以及当前时刻用户的局部相似性信息,可以显著提高 QoS 的预测性能。②使用真实的 Web 服务 QoS 数据集对提出的方法进行了实验评估。结果验证了所提方法具有较高的预测性能,表明了该方法可以很好地处理数据稀疏和噪声的问题。

基于协同过滤方法的 Web 服务推荐得到广泛的研究和应用,并取得较好的效果。协同过滤推荐技术主要分为基于记忆的协同过滤推荐^[16]、基于模型的协同过滤推荐。基于记忆的方法又可分为基于用户的^[17]和基于物品的^[18]两种方法。基于记忆的协同过滤推荐需要根据历史数据得到相似用户或服务,基于相似用户或服务,实现对 Web 服务的 QoS 值的预测。基于模型的方法^[19]通过历史数据的学习,建立一个全局模型,确定用户-物品间的隐含关系,实现对 QoS 值的预测。Zheng 等^[20]将基于用户和基于物品的方法进行混合,实现 Web 服务的可靠性预测。Wang 等^[10]基于用户相似性和基站相似度实现移动网络环境下 Web 服务的个性化推荐,并在此基础上对 QoS 值进行预测。卢凤等^[9]基于用户和服务的时空相似性特征,对相似性计算方法进行了改进,构建了 Web 服务推荐系统的框架,提高了对 QoS 协同预测的精确度。王磊等^[5]为避免协同过滤推荐精度受数据稀疏的影响,提出基于 K-means 聚类的 Slope One 算法,能够有效地提升 QoS 预测的精度。基于模型的方法能发现用户-服务之间隐含的关系,具有更高的预测精度^[21]。为进一步提高 QoS 的预测精度,Zhang 等^[14]基于 Web 服务的时间感知属性对用户-服务-时间矩阵进行分解,有效地填充了未知的 QoS 值。Chen 等^[14]将用户或服务聚为一类,认为同类的用户或服务共享某种隐含特征,并基于此实现对 QoS 值的预测。Yang 等^[11]基于用户和服务的位置信息对用户-服务矩阵进行分解,有效地解决数据稀疏和冷启动的问题,Liu 等^[13]也利用位置属性对协同过滤推荐的结果进行优化。Ryu 等^[15]则在位置信息的基础上,基于偏好可传播性对用户-服务矩阵进行分解,有效地解决了冷启动的问题。唐明董等^[12]在位置信息的基础上,采用因子分解方法进行矩阵分解,有效地解决了数据稀疏和冷启动的问题。

目前,解决数据稀疏等问题取得了较多成果,然而其主要工作均建立在用户和服务的地理位置相似性上,对其时间属性的应用并不充分。文中将充分应用用户或服务的时间空间属性,挖掘用户、服务之间的隐含关系,提升 QoS 的预测性能。

1 基于全局和局部结构相似性的稀疏矩阵分解模型 (GLMF)

先介绍 GLMF 方法的总体流程,然后对 GLMF 的实现细节进行描述,包括模型和算法步骤。GLMF 方法的总体流程,如图 1 所示,GLMF 通过记录 $t+1$ 时刻用户-服务 QoS 矩阵的位置相似性和相邻时间 QoS 矩阵的时间相似性,采用稀疏矩阵分解的方法对 $t+1$ 时刻未知的 QoS 值进行预测。

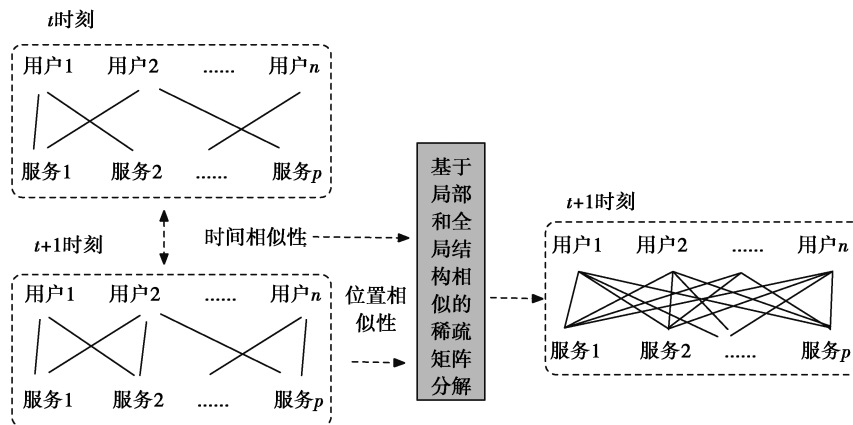


图 1 GLMF 的总体流程

Fig. 1 The general flow of GLMF

1.1 QoS 矩阵的构建

对未知 QoS 值的预测需要获得历史 Web 服务调用所产生的 QoS 数据,主要通过服务用户的反馈以及 QoS 的监测系统产生。前者收集用户对 Web 服务的反馈,如响应时间、可用性及信誉等;后者通过部署在服务器上的监测系统收集 Web 服务的 QoS 属性。所有服务调用产生的 QoS 记录可以用一个矩阵来表示,即用户-服务的 QoS 矩阵,每个 QoS 记录可能包含多个 QoS 参数。下面给出各个时刻 t 下的用户-服务 QoS 矩阵的形式化表示。

1) S 代表 Web 服务集合, $S = \{s_1, s_2, \dots, s_p\}$; U 用于代表调用 Web 服务的所有用户的集合, $U = \{u_1, u_2, \dots, u_n\}$ 。这里假设所有时刻的用户和服务不变。

2) \mathbf{R}^t 代表 t 时刻用户调用服务产生 QoS 记录的矩阵, 即用户-服务 QoS 矩阵, $\mathbf{R}^t = \{r_{ij}^t | u_i \in U, s_j \in S\}$ 。其中, r_{ij}^t 代表 t 时刻用户 u_i 调用服务 s_j 产生的 QoS 记录, 该记录可以是响应时间、吞吐量、反馈等。若用户未调用过该服务, 则 QoS 记录为空。真实场景中, 每个时刻用户通常仅使用部分 Web 服务, 产生的用户-服务 QoS 矩阵应当十分稀疏。文中要解决的问题可以形式化表示为基于稀疏矩阵 \mathbf{R}^t 和 \mathbf{R}^{t+1} , 实现对矩阵 \mathbf{R}^{t+1} 的低秩填充。

1.2 模型简介

对 $t+1$ 时刻的矩阵 \mathbf{R}^{t+1} 进行典型的低秩矩阵分解, 其损失函数为

$$F = \operatorname{argmin}_{\mathbf{A}^{t+1}, \mathbf{B}^{t+1}} (\|\mathbf{R}^{t+1} - \mathbf{A}^{t+1} (\mathbf{B}^{t+1})^T\|_F^2 + \lambda_1 (\|\mathbf{A}^{t+1}\|_F^2 + \|\mathbf{B}^{t+1}\|_F^2)), \quad (1)$$

其中, $\mathbf{R}^{t+1} (N \times P)$, $\|\cdot\|_F^2$ 表示矩阵的 F 范数, $\|\mathbf{A}^{t+1}\|_F^2 + \|\mathbf{B}^{t+1}\|_F^2$ 是正则项, 用于避免过拟合, \mathbf{A}^{t+1} 和 \mathbf{B}^{t+1} 又可称之为用户和服务的隐特征矩阵。假设共有 M 个隐式特征, 则 $\mathbf{A}^{t+1} (N \times M)$, $\mathbf{B}^{t+1} (P \times M)$ 。

由于相邻时刻网络环境相差不大, 相似用户调用服务的 QoS 属性可能具有相似性, 且相邻时刻的 QoS 矩阵也具有相似性。上述特征反映在 QoS 矩阵上分别是局部结构相似性(位置相似性)和全局结构相似性(时间相似性)。将其融入到式(1)中得到修正后的损失函数为

$$F = \operatorname{argmin}_{\mathbf{A}^{t+1}, \mathbf{B}^{t+1}} (\|\mathbf{R}^{t+1} - \mathbf{A}^{t+1} (\mathbf{B}^{t+1})^T\|_F^2 + \lambda_1 (\|\mathbf{A}^{t+1}\|_F^2 + \|\mathbf{B}^{t+1}\|_F^2) + \lambda_2 (\|\mathbf{A}^{t+1} (\mathbf{B}^{t+1})^T - \mathbf{R}^t\|_F^2) + \lambda_3 (\|\mathbf{L} (\mathbf{A}^{t+1} (\mathbf{B}^{t+1})^T)\|_F^2)). \quad (2)$$

其中, 第 3 项通过最小化 $t+1$ 时刻与 t 时刻 QoS 矩阵差异性的方式保留 t 时刻 QoS 矩阵的全局结构特征; 第 4 项通过拉普拉斯矩阵 \mathbf{L} 保留 $t+1$ 时刻的局部结构特征, $\mathbf{L} = \mathbf{D} - \mathbf{W}$ 。 \mathbf{W} 是 \mathbf{R}^{t+1} 的相似度矩阵(采用欧氏距离进行表示, 距离越大, 相似度越小), \mathbf{D} 是对角矩阵(其对角线上的元素来自于对 \mathbf{W} 各行的求和)。

基于式(2)计算梯度为

$$\begin{aligned} \frac{\partial F}{\partial \mathbf{A}^{t+1}} &= -\mathbf{R}^{t+1} \mathbf{B}^{t+1} + \mathbf{A}^{t+1} (\mathbf{B}^{t+1})^T \mathbf{B}^{t+1} + \lambda_1 \mathbf{A}^{t+1} + \\ &\lambda_2 (\mathbf{A}^{t+1} (\mathbf{B}^{t+1})^T - \mathbf{R}^t) \mathbf{B}^{t+1} + \lambda_3 \mathbf{L}^T \mathbf{L} \mathbf{A}^{t+1} (\mathbf{B}^{t+1})^T \mathbf{B}^{t+1}, \\ \frac{\partial F}{\partial \mathbf{B}^{t+1}} &= -(\mathbf{R}^{t+1})^T \mathbf{A}^{t+1} + \mathbf{B}^{t+1} (\mathbf{A}^{t+1})^T \mathbf{A}^{t+1} + \lambda_1 \mathbf{B}^{t+1} + \\ &\lambda_2 ((\mathbf{A}^{t+1} (\mathbf{B}^{t+1})^T - \mathbf{R}^t))^T \mathbf{A}^{t+1} + \lambda_3 \mathbf{B}^{t+1} (\mathbf{A}^{t+1})^T \mathbf{L}^T \mathbf{L} \mathbf{A}^{t+1}. \end{aligned}$$

采用乘性迭代法则对 \mathbf{A}^{t+1} 和 \mathbf{B}^{t+1} 进行寻优:

$$\mathbf{A}^{t+1} = \mathbf{A}^{t+1} \frac{\mathbf{C}}{\mathbf{D}}, \quad (3)$$

$$\mathbf{B}^{t+1} = \mathbf{B}^{t+1} \frac{\mathbf{E}}{\mathbf{F}}, \quad (4)$$

其中, $\mathbf{C} = \mathbf{R}^t + \mathbf{1} \mathbf{B}^{t+1} + \lambda_2 \mathbf{R}^t \mathbf{B}^{t+1}$;

$$\mathbf{D} = \mathbf{A}^{t+1} (\mathbf{B}^{t+1})^T \mathbf{B}^{t+1} + \lambda_1 \mathbf{A}^{t+1} + \lambda_2 \mathbf{A}^{t+1} (\mathbf{B}^{t+1})^T \mathbf{B}^{t+1} + \lambda_3 \mathbf{L}^T \mathbf{L} \mathbf{A}^{t+1} (\mathbf{B}^{t+1})^T \mathbf{B}^{t+1};$$

$$\mathbf{E} = (\mathbf{R}^{t+1})^T \mathbf{A}^{t+1} + \lambda_2 (\mathbf{R}^t)^T \mathbf{A}^{t+1};$$

$$\mathbf{F} = \mathbf{B}^{t+1} (\mathbf{A}^{t+1})^T \mathbf{A}^{t+1} + \lambda_1 \mathbf{B}^{t+1} + \lambda_2 \mathbf{B}^{t+1} (\mathbf{A}^{t+1})^T \mathbf{A}^{t+1} + \lambda_3 \mathbf{B}^{t+1} (\mathbf{A}^{t+1})^T \mathbf{L}^T \mathbf{L} \mathbf{A}^{t+1}.$$

1.3 基于全局和局部结构相似性的稀疏矩阵分解算法

算法 1. 基于时空感知的稀疏矩阵分解算法。

输入: 用户-服务 QoS 矩阵 \mathbf{R}^{t+1} 和 \mathbf{R}^t ; 隐式特征数 M ; 误差标准 ϵ 和 δ ; 迭代次数 count。

输出: QoS 预测值填充后的 QoS 矩阵 $\widehat{\mathbf{R}^{t+1}}$ 。

① 随机初始化分解因子 $\mathbf{A}^{t+1} (N \times M)$ 和 $\mathbf{B}^{t+1} (P \times M)$;

② cnt=0;

③ 按照式(5)和式(6)计算误差度量指标 MAE 和 RMSE;

④ while cnt < count || MAE < ϵ || RMSE < δ

- ⑤按照式(3)和式(4)对 \mathbf{A}^{t+1} 和 \mathbf{B}^{t+1} 更新;
- ⑥按照式(5)和式(6)计算误差度量指标 MAE 和 RMSE;
- ⑦end while
- ⑧ $\widehat{R}^{t+1} = \mathbf{A}^{t+1} (\mathbf{B}^{t+1})^T$ 。

2 实验验证与分析

为了验证方法的有效性,采用真实的 Web 服务 QoS 数据进行实验评估^[14]。该数据集采集自 wsdream.com 网站,包含 4 532 个 Web 服务在 64 个相邻时间段内被 142 个用户调用的 QoS 记录,主要是吞吐量(throughput)和响应时间(response time,或 round trip time,RTT)两方面。值得注意的是,QoS 有多种属性,包括响应时间、吞吐量等客观属性,也包括用户对服务质量的反馈等主观属性。对于客观属性数据,需要对其归一化以消除不同量纲的影响;对于主观属性数据,可以利用用户平均值与整体平均值进行纠偏(这里假定整体均值是中立的)。

文中使用服务的 RTT 数据来评估所提出的方法。实验采用的数据集是用户端的轻量级中间件自动收集用户调用服务时产生的真实 QoS 记录,由于所处网络环境、设备的不同,其 QoS 记录必然是包含噪声的。此外,真实环境下的 RTT 矩阵是稀疏的,在实验时为了模拟实际的 Web 服务应用场景,通过随机移除 RTT 矩阵的一部分数据实现对矩阵的稀疏化。采用稀疏度(移除数据的大小/总矩阵的大小)来衡量矩阵的稀疏程度。稀疏化后的矩阵是可看做训练数据,而被移除的数据可看做测试数据。

2.1 评估指标

通过计算预测的 RTT 值与实际 RTT 值之间的偏差来评估文中方法在预测 Web 服务 RTT 的准确性。常用的误差度量指标:平均绝对误差 MAE(mean absolute error)和均方根误差 RMSE(root mean squared error)。二者的计算公式如下:

$$\text{MAE} = \frac{1}{|\text{SP}|} \sum_{r_{i,j}^{t+1} \in \text{SP}} |r_{i,j}^{t+1} - \widetilde{r}_{i,j}^{t+1}|, \quad (5)$$

$$\text{RMSE} = \sqrt{|r_{i,j}^{t+1} - \widetilde{r}_{i,j}^{t+1}|^2}, \quad (6)$$

其中,SP 是被移除的数据,即用于测试的数据集,|SP|是数据集的大小。MAE 和 RMSE 的值越小,代表预测的精度越高。

2.2 对比方法的选择

为了比较 GLMF 方法与其他方法在预测性能上的优劣,实验中将其与以下 2 个经典的预测方法进行了对比:

1)非负矩阵分解(non-negative matrix factorization,NMF)。NMF 约束分解后的矩阵分量为非负的。在 QoS 预测中,这种假设是合理的。

2)奇异值分解(single value decomposition,SVD)。SVD 是实现矩阵低秩近似的典型方式之一。基于历史数据的潜在信息寻找一个属性空间,用户调用服务的 QoS 值由用户和服务在此属性空间的点集求得。

3)Kmeans+Slope One 算法。基于项目(服务)的协同过滤推荐中,Slope One 算法可以有效地实现缺值预测。然而,Slope One 算法在实现时未考虑到项目与项目之间的相似性。因此,在空缺值预测之前,使用 K-means 算法对项目进行聚类以提高 QoS 预测精度。

2.3 参数 λ_1 、 λ_2 、 λ_3 对预测性能的影响

参数 λ_1 所在项是为了防止过拟合而设置的,参数 λ_2 和 λ_3 分别用于保留 QoS 矩阵的全局结构相似性和局部结构相似性。文中矩阵的稀疏度均设置为 0.9,隐式特征数目设置为 30。从图 2、图 3 和图 4 可以发现,参数 λ_1 、 λ_2 、 λ_3 的取值在一定程度上会影响到预测的精度。图 2 中 MAE 和 RMSE 起初随着参数 λ_1 的增加快速下降,当 λ_1 大于 0.000 5 时,整体反而上升。图 3 中 MAE 和 RMSE 起初随着参数 λ_2 的增加快速下降,当 λ_2 大于 0.5 时,基本保持平稳或者缓慢下降。图 4 中 MAE 和 RMSE 起初随着 λ_3 的增加基本保持不变,当 λ_3 大于 0.000 5 时,反而快速上升。由此,参数 λ_1 的值可固定为 0.000 5;参数 λ_2 和 λ_3 可分别设置为 0.5、0.000 5。

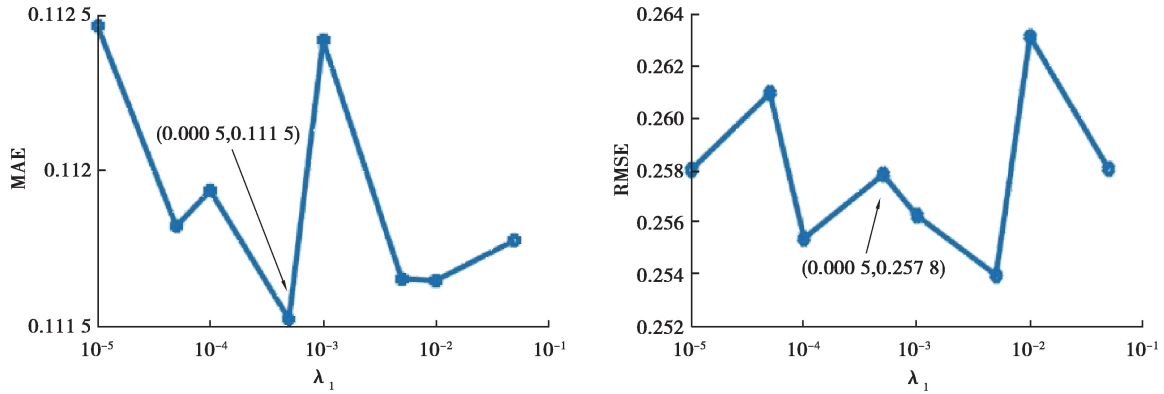


图 2 参数 λ_1 对预测性能的影响
 Fig. 2 The effect of parameter λ_1 on prediction performance

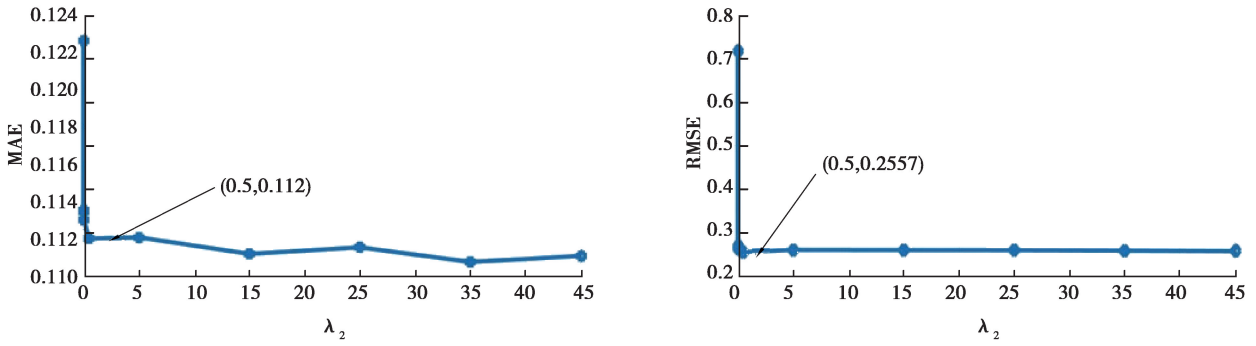


图 3 参数 λ_2 对预测性能的影响
 Fig. 3 The effect of parameter λ_2 on prediction performance

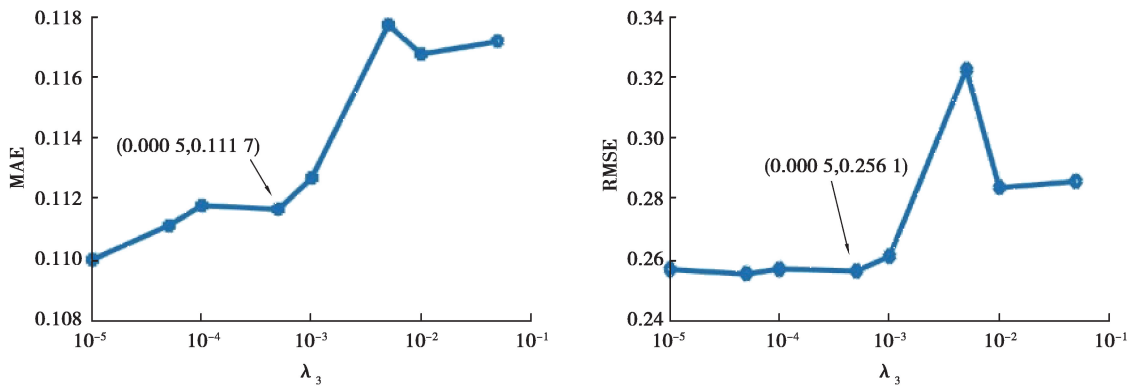


图 4 参数 λ_3 对预测性能的影响
 Fig. 4 The effect of parameter λ_3 on prediction performance

2.4 QoS 矩阵的稀疏度对预测性能的影响

在真实场景中,各个时刻的 QoS 矩阵均是稀疏的,其稀疏程度不同,预测性能是不同的。文中模型对时刻 t 和 $t+1$ 的 QoS 矩阵 (R^t 和 R^{t+1}) 进行处理,矩阵 R^t 和 R^{t+1} 的稀疏度均会对预测性能有影响。具体结果如图 5 和图 6 所示。图 5 中固定矩阵 R^t 的稀疏度为 0.9,矩阵 R^{t+1} 的稀疏度范围为 $[0.55, 0.95]$,每步增长 0.05。由图 5 可知,MAE 和 RMSE 的值随着稀疏度的增加整体上不断增加,且整体增加幅度不大。这意味着适当的收集 QoS 记录,减小数据稀疏度可以提高预测的精度。图 6 中固定矩阵 R^{t+1} 的稀疏度为 0.9,矩阵 R^t 的稀疏度范围为 $[0.55, 0.95]$,每步增长为 0.05。由图 6 可知,随着稀疏度的增加,MAE 值出现缓慢的增

长, RMSE 值基本保持不变。这意味着前一时刻矩阵的稀疏度对预测的性能有影响, 但影响效果十分有限, 表明文中方法具有良好的可扩展性, 少量的全局结构信息可以达到较好的预测性能。

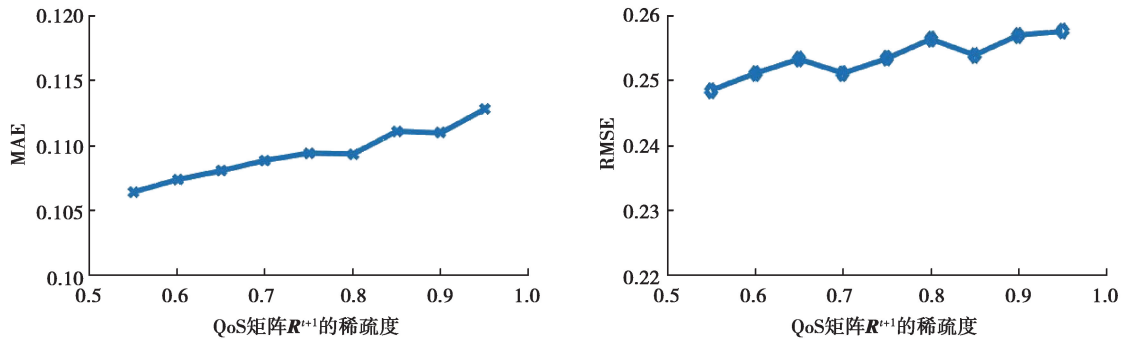


图 5 QoS 矩阵 R^{t+1} 稀疏度对预测性能的影响

Fig. 5 The effect of the sparseness of QoS matrix R^{t+1} on prediction performance

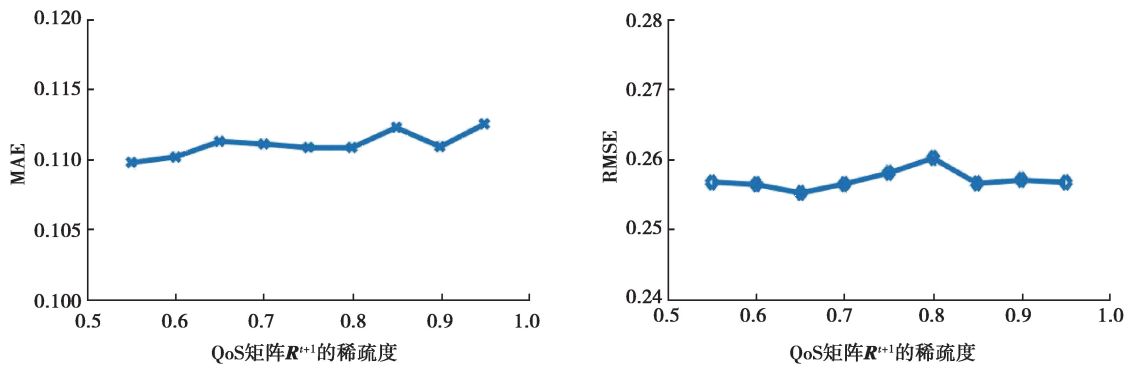


图 6 QoS 矩阵 R^t 稀疏度对预测性能的影响

Fig. 6 The effect of the sparseness of QoS matrix R^t on prediction performance

2.5 隐特征数目对预测性能的影响

隐特征个数体现了对 QoS 矩阵认知的程度。隐特征个数越高意味着对 QoS 矩阵的认知越高, 增加了计算的复杂度; 隐特征个数越少, 虽然降低了计算的复杂度, 但是减少了对 QoS 矩阵的认知。图 7 展示了隐特征数目对预测性能的影响, 其中 QoS 矩阵 R^t 和 R^{t+1} 的稀疏度均设置为 0.9, 隐特征数目的范围为 $[5, 100]$, 每步增长 5。随着隐特征数目的增加, RMSE 整体上呈现显著的降低, MAE 整体上变化不大。这表明隐特征数目的增加, 一定程度上加强了对 QoS 矩阵的认知, 导致预测性能总体变好。然而, 在 $[20, 40]$ 范围内, MAE 和 RMSE 值均出现一定的波动。因为并不是所有增加的隐特征都能有效加强对 QoS 矩阵的认知。文中均衡计算性能和预测性能, 最终在实验中设置隐特征数目为 30。

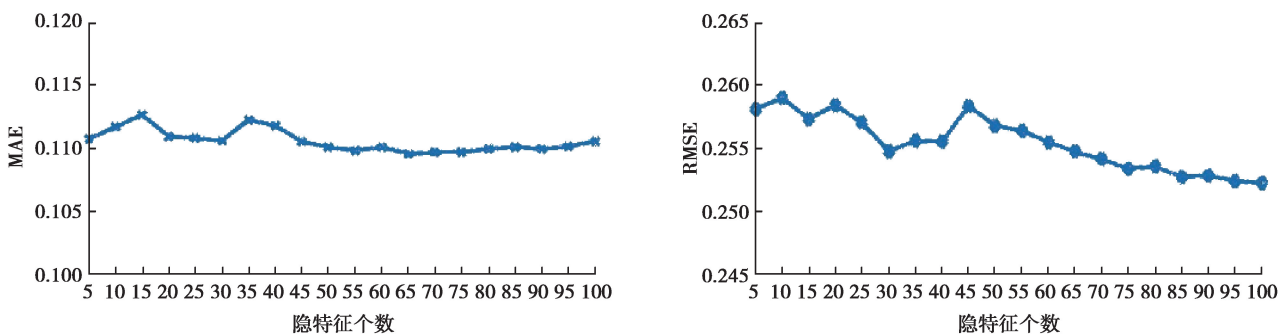


图 7 隐特征个数对预测性能的影响

Fig. 7 The effect of the number of underlying features on prediction performance

2.6 不同方法预测性能的对比

表 1 显示了 GLMF 与各种对比方法的预测结果。其中, Kmeans+Slope One 算法是基于记忆的协同过滤方法, 这种方法在数据稠密时能取得较好的效果, 却并不适用于数据稀疏的环境, 因此预测性能最低。NMF、SVD 和 GLMF 均利用矩阵分解理论对用户-服务 QoS 矩阵进行建模, 发现用户-服务的隐含特征, 提高了预测的性能。由表 1 可知, 随着矩阵稀疏度的增加, 无论何种方法, 其 MAE 和 RMSE 值均增加, 预测性能均下降。值得一提的是, GLMF 方法在高稀疏度情境下, 其 MAE 和 RMSE 值均小于其他方法, 具有更优的预测性能。相比于 NMF, GLMF 的 MAE 值最大下降了 3.25%, RMSE 值最大下降了 6.65%; 相比于 SVD, GLMF 的 MAE 值最大下降了 3.67%, RMSE 值最大下降了 7.01%。与 NMF 和 SVD 相比, GLMF 充分利用了用户-服务 QoS 矩阵的全局和局部结构特征, 进一步提高了预测精度。

表 1 GLMF、NMF、SVD 和 Kmeans+Slope One 算法的预测性能

Table 2 The prediction performance of GLMF、NMF、SVD and Kmeans+Slope One algorithms

| 稀疏度 | GLMF | | NMF | | SVD | | Kmeans+Slope One | |
|---------|---------|---------|---------|---------|---------|---------|------------------|---------|
| | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| 0.900 0 | 0.111 0 | 0.254 1 | 0.111 0 | 0.266 2 | 0.116 4 | 0.273 6 | 0.805 5 | 0.866 4 |
| 0.910 0 | 0.110 8 | 0.256 5 | 0.111 6 | 0.267 4 | 0.116 8 | 0.274 9 | 0.814 0 | 0.871 6 |
| 0.920 0 | 0.111 4 | 0.257 9 | 0.112 3 | 0.269 3 | 0.117 0 | 0.275 7 | 0.773 4 | 0.847 4 |
| 0.930 0 | 0.111 9 | 0.259 6 | 0.112 9 | 0.270 9 | 0.117 0 | 0.276 1 | 0.764 2 | 0.842 7 |
| 0.940 0 | 0.110 9 | 0.261 6 | 0.113 4 | 0.272 1 | 0.117 1 | 0.277 2 | 0.829 8 | 0.880 4 |
| 0.950 0 | 0.112 3 | 0.264 6 | 0.114 1 | 0.273 8 | 0.117 2 | 0.277 8 | 0.802 2 | 0.854 6 |
| 0.960 0 | 0.113 1 | 0.259 4 | 0.114 6 | 0.275 0 | 0.117 2 | 0.278 2 | 0.790 8 | 0.865 4 |
| 0.970 0 | 0.113 1 | 0.257 1 | 0.115 2 | 0.276 6 | 0.117 2 | 0.279 3 | 0.803 4 | 0.856 4 |
| 0.980 0 | 0.113 3 | 0.262 0 | 0.116 0 | 0.278 3 | 0.117 2 | 0.279 9 | 0.779 8 | 0.835 9 |
| 0.990 0 | 0.112 8 | 0.261 1 | 0.116 6 | 0.279 7 | 0.117 1 | 0.280 8 | 0.764 3 | 0.853 3 |

3 结 论

文中针对用户-服务间的 QoS 预测问题, 充分利用时空特征, 即相邻时间段内, 相似用户具有相似的网络环境, 更大可能具有相似的 QoS 属性, 提出了一种基于全局和局部结构相似性的稀疏矩阵分解算法。该算法将 QoS 预测问题归结为矩阵的低秩近似问题, 通过乘性迭代的方式寻找矩阵的因子, 最终实现对 QoS 矩阵的低秩填充。实验结果表明, 基于全局和局部结构相似性的稀疏矩阵分解模型 (GLMF) 与经典的矩阵分解方法、基于记忆的协同过滤方法相比, QoS 预测性能有一定的提高, 能有效解决数据稀疏、噪声等问题。未来拟将文中方法进一步应用于更多场景下, 如为基于情景的 Web 服务推荐 (主要包括基于 QoS 的服务聚集、服务组合优化和服务的异常检测等) 提供支撑。

参考文献:

- [1] Xiong R B, Wang J, Zhang N, et al. Deep hybrid collaborative filtering for Web service recommendation[J]. Expert Systems With Applications, 2018, 110: 191-205.
- [2] Chen S L, Peng Y X, Mi H B, et al. A cluster feature based approach for QoS prediction in web service recommendation[C]// 2018 IEEE Symposium on Service-Oriented System Engineering (SOSE), March 26-29, 2018, Bamberg. IEEE, 2018: 246-251.
- [3] Gao M, Wu Z F, Jiang F. UserRank for item-based collaborative filtering recommendation[J]. Information Processing Letters, 2011, 111(9): 440-446.
- [4] Anithadevi N, Sundarambal M. A design of intelligent QoS aware web service recommendation system[J]. Cluster

- Computing, 2019, 22(S6): 14231-14240.
- [5] 王磊, 瞿佳明. 基于协同过滤和 Slope One 算法的 Web 服务可靠性预测[J]. 计算机工程与科学, 2018, 40, 284(08): 58:65.
Wang L, Qu J M. Web service reliability prediction via collaborative filtering and slope one algorithm[J]. Computer Engineering & Science, 2018, 40, 284(8): 58-65. (in Chinese)
- [6] Zheng Z B, Ma H, Lyu M R, et al. QoS-aware web service recommendation by collaborative filtering[J]. IEEE Transactions on Services Computing, 2011, 4(2): 140-152.
- [7] 任丽芳. 动态环境下的 Web 服务推荐与组合方法研究[D]. 山西:山西大学 2018.
Ren L F. Research on web services recommendation and compoition in dynamic environment [D]. shanxi: ShanXi University, 2018. (in Chinese)
- [8] 任迪, 万健, 殷昱煜, et al. 基于贝叶斯分类的 Web 服务质量预测方法研究[J]. 浙江大学学报(工学版), 2017, 51(6): 1242-1251.
Ren D, Wan J, Yin Y Y, et al. Web services QoS prediction method based on Bayes Classification[J]. Journal of Zhejiang University (Engineering Science), 2017, 51(6): 1242-1251. (in Chinese)
- [9] 卢凤, 李海荣, 韩艳. 基于时空相似度感知的 Web 服务 QoS 协同过滤推荐[J]. 计算机工程, 2017, 43(4): 28-33,38.
Lu F, Li H R, Han Y. QoS collaborative filtering recommendation of web service based on spatial and temporal similarity sensing[J]. Computer engineering, 2017, 43(4): 28-33,38. (in Chinese)
- [10] Wang L B, Sun Q B, Wang S G, et al. Web service QoS prediction approach in mobile Internet environments[C]//2014 IEEE International Conference on Data Mining Workshop, December 14, 2014, Shenzhen, China. IEEE, 2014: 1239-1241.
- [11] Yang Y T, Zheng Z B, Niu X D, et al. A location-based factorization machine model for web service QoS prediction [J]. IEEE Transactions on Services Computing, 2019: 1.
- [12] 唐明董, 张婷婷, 杨亚涛, 等. 基于因子分解机的质量感知 Web 服务推荐方法[J]. 计算机学报, 2018, 41, 426(06): 114-127.
Tang M D, Zhang T T, Yang Y T, et al. Qos-aware web service recommendation based on factorization machines [J]. Chinese Journal of Computers, 41, 426(06): 114-127. (in Chinese)
- [13] Liu J X, Tang M D, Zheng Z B, et al. Location-aware and personalized collaborative filtering for web service recommendation[J]. IEEE Transactions on Services Computing, 2015, 9:689-699.
- [14] Zhang Y, Zheng Z, Lyu M R. WSPred: a time-aware personalized QoS prediction framework for web services[C]// Proceedings of the 22nd IEEE International Symposium on Software Reliability Engineering, Hiroshima, Japan, 2011: 210-219.
- [15] Zhang Y L, Zheng Z B, Lyu M R. WSPred: a time-aware personalized QoS prediction framework for web services[C]// 2011 IEEE 22nd International Symposium on Software Reliability Engineering, November 29-December 2, 2011. Hiroshima, Japan. IEEE, 2011:210-219.
- [16] Ryu D, Lee K, Baik J. Location-based web service QoS prediction via preference propagation to address cold start problem[J]. IEEE Transactions on Services Computing, 2018: 1.
- [17] Adomavicius G, Alexander T. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6): 734-749.
- [18] Gunes I, Kaleli C, Bilge A, et al. Shilling attacks against recommender systems: a comprehensive survey[J]. Artificial Intelligence Review, 2014, 42(4): 767-799.
- [19] Xia H, Fang B, Gao M, et al. A novel item anomaly detection approach against shilling attacks in collaborative recommendation systems using the dynamic time interval segmentation technique[J]. Information Sciences, 2015, 306: 150-165.
- [20] Roughan M, Zhang Y, Willinger W, et al. Spatio-temporal compressive sensing and Internet traffic matrices (extended version)[J]. ACM Transactions on Networking, 2012, 20(3): 662-676.
- [21] Zheng Z, Lyu M R. Collaborative reliability prediction of service-oriented systems// ACM/IEEE International Conference on Software Engineering, Cape Town, South Africa, 2010: 35-44.
- [22] Zheng Z B, Lyu M R. Personalized reliability prediction of web services[J]. ACM Transactions on Software Engineering and Methodology, 2013, 22(2): 1-25.