













$$R_{u,i} = \lambda \times R_{u,i}(u) + (1-\lambda) \times R_{u,i}(i) \quad (5)$$

其中,参数  $\lambda$  表示预测值依赖相似用户和相似服务的比例.当  $\lambda=1$  时,仅使用基于用户的方法进行预测;当  $\lambda=0$  时,仅使用基于服务的方法进行预测;如果用户  $u$  没有调用过任何服务,而服务  $i$  也没有被任何用户调用过,则使用所有服务 QoS 值的平均值进行预测.

### 3 实验结果与分析

本文在真实的数据集上做了大量实验来比较所提出的方法与目前经典的相关方法的预测精度,同时研究不同的参数对预测精度的影响.本部分实验环境是 IntelliJ IDEA Community Edition 14.1.4,jdk 1.7,scala-sdk 2.10.4,编程语言 scala,实验机器配置为 16G 内存,core i7-4970 3.6GHZ 处理器,ubuntu 14.04 LTS 系统.

#### 3.1 实验数据

在实验中,为评价本文提出的 SOMQP 的服务质量预测方法的有效性,我们使用 WS-Dream 数据集<sup>[26]</sup>,这个数据集中两个矩阵分别包含 339 用户调用 5 825 个服务产生的 1 974 675 个 Web 服务的响应时间(response time,简称 RT)和吞吐量(throughput,简称 TP)的真实记录值,其中,83.7%的响应时间值在[0,1]之间.真实世界中,大部分用户都只调用过少数的服务,所以产生的用户-服务矩阵是一个稀疏矩阵.为此,我们随机移出一部分数据作为训练集,剩余的作为测试集.在 WS-Dream 数据集中,训练集的比例为 0.05,0.1,0.15,0.2,0.25.

#### 3.2 评价指标

为评价服务质量预测的有效性,本文在数据集上划分多个训练集进行实验;同时,也在不同相似用户数和相似服务数的基础上进行验证.本文选择两个度量预测 QoS 值质量的指标是绝对平均误差(MAE)和均方根误差(RMSE),分别定义如下:

$$MAE = \frac{\sum_{u,s} |R(u,s) - P(u,s)|}{N} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{u,s} (R(u,s) - P(u,s))^2}{N}} \quad (7)$$

其中, $R(u,s)$ 表示用户  $u$  调用服务  $s$  得到的实际 QoS 值, $P(u,s)$ 表示预测的值, $N$ 表示预测值的数量.MAE 是通过比较用户调用服务的真实 QoS 值和预测值之间差异值来计算的,它的值越小,表示预测精度越高.

#### 3.3 实验结果与分析

为评价所提出的 SOMQP 方法的预测精度,选择以下经典方法在 MAE 和 RMSE 两个指标上进行对比实验.

- 1) UPCC(基于用户的协同过滤方法).该方法基于相似用户的历史 QoS 值去预测缺失值.
  - 2) IPCC(基于项目的协同过滤方法).该方法基于相似服务的历史 QoS 值去预测缺失值.
  - 3) UIPCC(混合的协同过滤方法).该方法是分别给与 UPCC 和 IPCC 一个权重,用综合的方法去预测缺失的值.
  - 4) KUPCC.借鉴文献[14],用  $K$ -means 方法对用户进行聚类,得到用户关系矩阵并寻找相似用户,然后基于相似用户对缺失 QoS 值进行预测.
  - 5) KIPCC.用  $K$ -means 方法对服务进行聚类,得到服务关系矩阵并寻找相似服务,然后基于相似服务对缺失 QoS 值进行预测.
  - 6) KUIPCC.基于 KUPCC 和 KIPCC 混合策略对缺失 QoS 值进行预测.
  - 7) TOSEM.该方法基于改进的 PCC 来计算用户或服务之间的相似度,进而用综合方法去预测缺失值<sup>[26]</sup>.
  - 8) SUPCC.用 SOM 神经网络算法对用户进行聚类寻找相似用户,基于相似用户对缺失 QoS 值进行预测.
  - 9) SIPCC.用 SOM 神经网络算法对服务进行聚类寻找相似服务,基于相似服务对缺失 QoS 值进行预测.
- 实验中,在 WS-Dream 数据集上设置矩阵密度为 0.05~0.25,以 0.05 的间隔逐渐增加,Top- $k$  相似用户和相似

服务中, $k=2$ ,在公式(5)中的  $\lambda=0.4$ ,分别求得这几种方法在响应时间和吞吐量上的 MAE 和 RMSE,实验结果比较见表 1 和表 2.

**Table 1** Comparison between SOMQP and other methods on RT

表 1 SOMQP 与其他方法在 RT 上的比较

方法	MAE/矩阵密度						RMSE/矩阵密度				
	0.05	0.1	0.15	0.2	0.25	平均	0.05	0.1	0.15	0.2	0.25
UPCC	0.637	0.531	0.496	0.472	0.451	0.517	1.513	1.372	1.343	1.295	1.272
IPCC	0.869	0.854	0.832	0.815	0.786	0.831	1.682	1.668	1.642	1.583	1.512
UIPCC	0.663	0.595	0.541	0.496	0.465	0.552	1.459	1.415	1.377	1.356	1.328
KUPCC	0.538	0.466	0.435	0.415	0.403	0.451	1.579	1.446	1.405	1.358	1.342
KIPCC	0.405	0.379	0.351	0.343	0.336	0.363	1.363	1.276	1.272	1.249	1.241
KUIPCC	0.418	0.393	0.365	0.353	0.345	0.375	1.297	1.221	1.210	1.174	1.168
TOSEM	0.652	0.525	0.481	0.454	0.435	0.509	1.447	1.302	1.258	1.211	1.196
SUPCC	0.491	0.436	0.411	0.392	0.381	0.422	1.520	1.415	1.374	1.327	1.314
SIPCC	0.406	0.371	0.362	0.352	0.346	0.367	1.370	1.285	1.286	1.259	1.254
<b>SOMQP</b>	<b>0.399</b>	<b>0.362</b>	<b>0.358</b>	<b>0.342</b>	<b>0.335</b>	<b>0.359</b>	<b>1.306</b>	<b>1.227</b>	<b>1.209</b>	<b>1.196</b>	<b>1.192</b>

**Table 2** Comparison between SOMQP and other methods on TP

表 2 SOMQP 与其他方法在 TP 上的比较

方法	MAE/矩阵密度						RMSE/矩阵密度				
	0.05	0.1	0.15	0.2	0.25	平均	0.05	0.1	0.15	0.2	0.25
UPCC	26.858	21.688	18.865	17.923	16.698	20.406	65.225	57.476	53.148	51.007	49.019
IPCC	37.341	36.862	35.013	33.695	30.218	34.625	74.322	71.061	68.430	63.697	59.077
UIPCC	27.642	25.425	24.768	23.242	21.571	24.529	62.869	58.116	56.348	54.189	53.662
KUPCC	17.211	15.287	14.239	13.777	13.254	14.754	55.502	50.816	48.399	47.159	45.759
KIPCC	15.837	14.285	13.199	12.734	12.247	13.660	55.208	51.401	48.727	47.315	45.749
KUIPCC	14.847	13.373	12.162	11.719	11.285	12.677	50.337	46.572	41.618	40.351	39.141
TOSEM	26.827	22.648	20.702	19.675	17.954	21.561	62.116	55.222	51.218	50.713	50.055
SUPCC	17.128	15.246	14.093	13.706	12.208	14.476	54.499	50.789	48.116	46.877	45.521
SIPCC	15.775	14.191	13.055	12.635	12.110	13.553	54.029	49.853	46.977	45.773	43.888
<b>SOMQP</b>	<b>14.420</b>	<b>13.044</b>	<b>11.980</b>	<b>11.618</b>	<b>11.154</b>	<b>12.443</b>	<b>46.869</b>	<b>43.328</b>	<b>41.108</b>	<b>39.943</b>	<b>38.514</b>

从实验结果可知:在矩阵密度相同的情况下,本文方法的 QoS 预测精度最好;同时,随着矩阵密度的增大,本文方法的预测精度一直优于其他方法,而且稀疏度越大时,本文方法与其他方法的预测精度的差距越大.表 1 中展示的是本文提出的 SOMQP 和其他经典算法在响应时间上的比较,从实验结果可以得出,SOMQP 方法在 MAE 上比 UIPCC,KUIPCC,TOSEM 这 3 种经典方法分别提高 34.9%,4.3%,29.5%.同样地,表 2 中是在吞吐量数据集上的比较,本文的算法比上述 3 个算法在 MAE 上平均值分别提高 49.5%,1.8%和 42.3%.综上所述,在不同的数据集上,本文提出的 SOMQP 算法的预测精度比这几种经典算法都有一定的提高.该实验结果表明,本文方法可以更好地选取目标用户的相似用户来进行缺失值预测,同时降低那些恶意评价用户对预测精度的影响.

#### (1) $\lambda$ 对预测精度的影响

在本文的方法中,参数  $\lambda$  是决定 SOMQP 方法依靠相似用户和相似服务的权重,若  $\lambda=1$  时,仅使用基于用户的方法进行预测;若  $\lambda=0$  时,仅使用基于服务的方法进行预测.为研究  $\lambda$  对预测结果的影响,本文在 WS-Dream 数据集上,设置矩阵密度为 0.05,相似用户数和相似服务数为 2 的情况下, $\lambda$  的值从 0.1 到 0.9 并以 0.1 的间隔逐渐增加对 MAE 和 RMSE 的影响,实验结果如图 6 所示.

其中,图 6(a)、图 6(b)是在响应时间的实验结果,图 6(c)、图 6(d)是在吞吐量上的实验结果.图 6 的实验结果表明,参数  $\lambda$  的值对 SOMQP 方法的预测精度有很重要的影响.从图 6(a)、图 6(c)中看出,一开始,随着  $\lambda$  值的增大,预测精度一直在提高,当预测精度达到最大值之后,随着  $\lambda$  值的增大,预测精度开始降低;图 6(a)中,在  $\lambda=0.2$  时,预测精度达到最优值;图 6(c)中,在  $\lambda=0.4$  时,预测精度达到最优值.这表明,一个合适的  $\lambda$  值可以达到更好的预测精度.同时,图 6(a)、图 6(c)的结果表明,在不同的数据集上,达到最优预测精度时的  $\lambda$  值不一样的.

#### (2) Top-k 对预测的影响

本文方法中参数 Top-k 决定了选取目标用户的相似用户数或目标服务的相似服务数.为研究 Top-k 值对预



测精度的影响,实验用 WS-Dream 数据集分别在响应时间和吞吐量两个属性上,在矩阵密度为 0.05, $\lambda=0.4$ ,设置 Top- $k$  值为从 1~6 并以 1 的间隔逐渐增加对 MAE 和 RMSE 的影响,结果如图 7 所示。

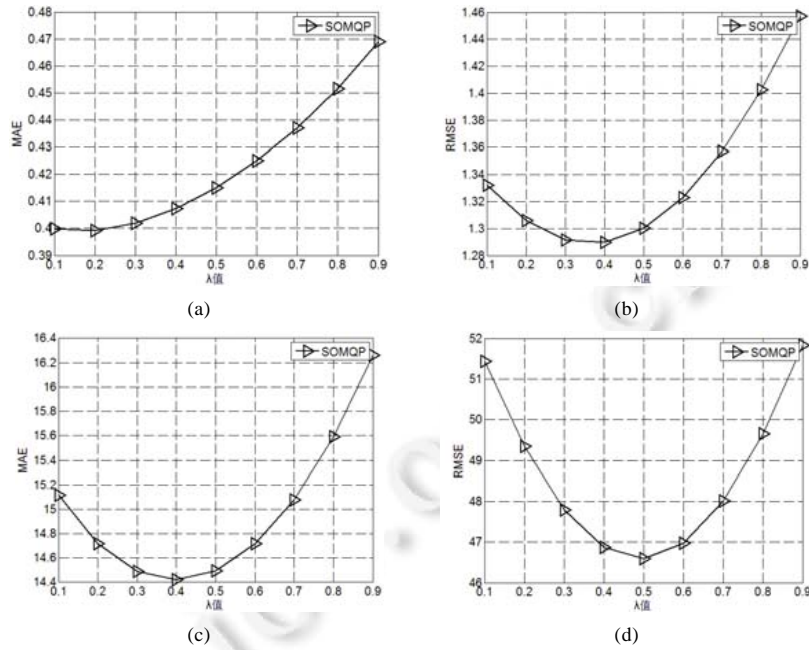


Fig.6 Impact of  $\lambda$  on prediction accuracy

图 6  $\lambda$  对预测精度的影响

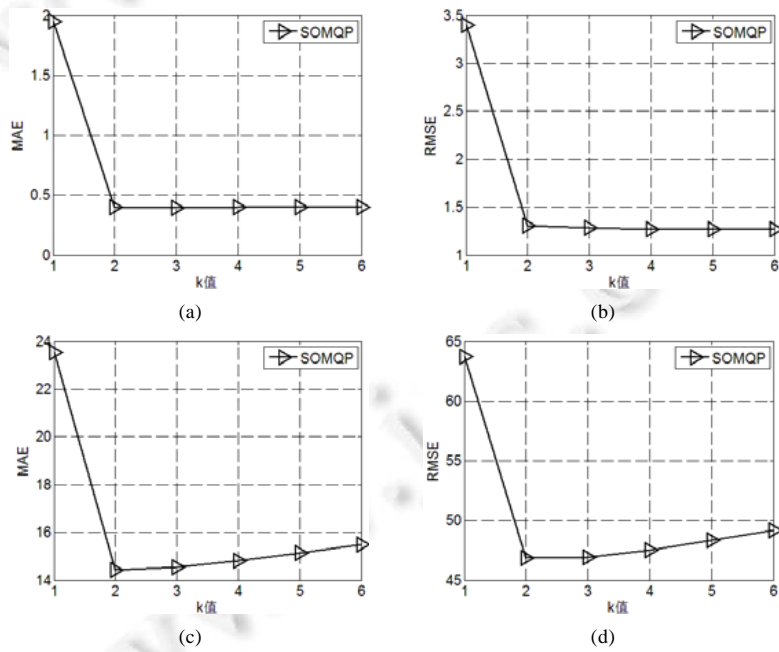


Fig.7 Impact of Top- $k$  on prediction accuracy

图 7 Top- $k$  对预测精度的影响

其中,图 7(a)、图 7(b)是在响应时间的实验结果,图 7(c)、图 7(d)是在吞吐量上的实验结果.从图 7(a)、图 7(c)可以看出,随着  $k$  值的增加,预测精度先变大然后变小,并在  $k$  为 2 时达到最大.这说明相似用户或相似服务数越多,并不一定有利于提高预测精度.相似用户数和相似服务数过少或过多,都会降低预测精度:如果相似用户数太少,不能全面考虑用户之间的关系;如果相似用户数太多,则不能排除那些对服务进行恶意评价的用户,这些恶意评价的用户会对预测精度有很大的影响.同样地,相似服务数越多,那些被用户恶意评价的服务信息会大大降低预测精度.

### (3) 矩阵密度的影响

矩阵密度表示从用户-服务矩阵中移出作为训练集的数据比例.图 8 是在 WS-Dream 数据集上,分别对响应时间和吞吐量两个 QoS 属性研究矩阵密度的影响.设置矩阵密度从 0.05 到 0.25,以 0.05 的间隔增加, $\lambda=0.4$ ,相似用户数和相似服务数为 2.

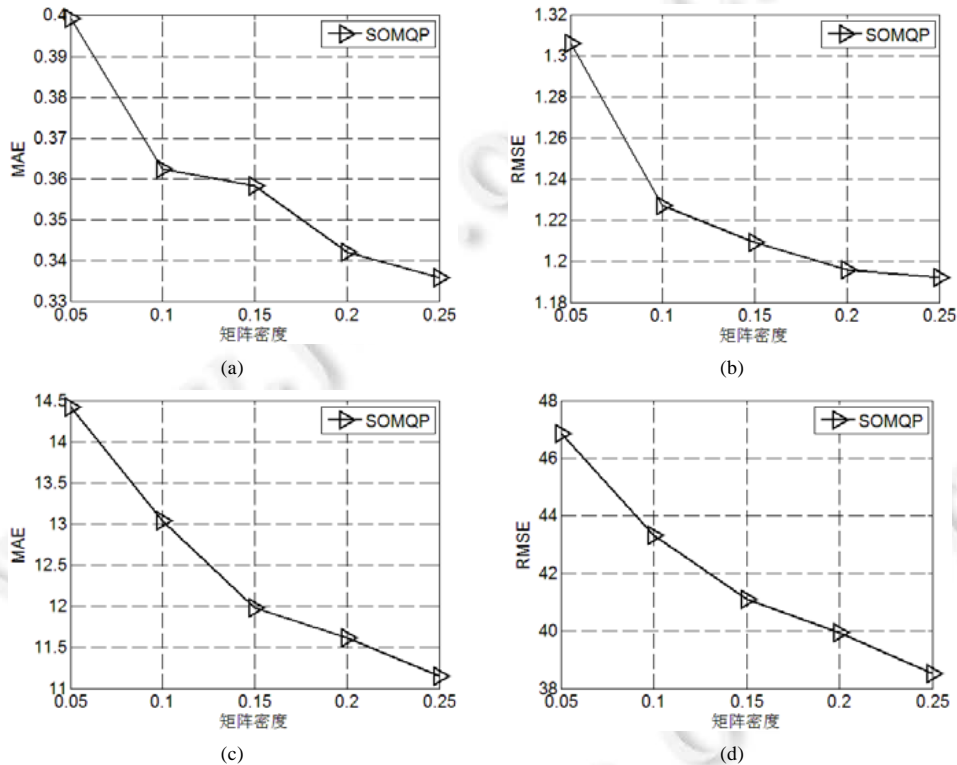


Fig.8 Impact of matrix density on prediction accuracy

图 8 矩阵密度对预测精度的影响

从图 8 可以看出,随着矩阵密度的增加,SOMQP 方法的服务质量预测精度也逐步提高.这是因为矩阵密度越大,说明更多的用户调用过更多的服务,所产生的数据集中包含更多的用户信息和服务信息,可以从这些信息中挖掘更多有用信息来提高预测精度.

## 4 结束语

本文提出一种基于 SOM 神经网络的服务质量预测方法 SOMQP.首先根据历史 QoS 数据,用 SOM 分别对用户和服务进行聚类,得到用户关系矩阵和服务关系矩阵;然后,基于一种新的 Top- $k$  选择机制,获得目标用户和目标服务的相似用户集和相似服务集;最后,结合基于相似用户的和基于相似服务的方法,提出一个混合策略对

缺失 QoS 值进行预测.在真实数据集上的实验结果表明,与经典的 CF 算法、基于改进的 PCC 预测算法和 K-means 聚类算法相比,我们方法的质量预测精度分别提高了 34.9%、29.5%和 4.3%.随着服务规模的增大,用户-服务信息矩阵也变得越来越大,在接下来的研究中,将在 Spark 平台使用并行聚类技术来验证本文算法的有效性和稳定性.

### References:

- [1] Yin J, Tang Y, Lo W, Wu Z. From big data to great services. In: Proc. of the 2016 IEEE Int'l Congress on Big Data (BigData Congress). IEEE, 2016. 165–172.
- [2] Yu Y, Chen J, Lin S, Wang Y. A dynamic QoS-aware logistics service composition algorithm based on social network. IEEE Trans. on Emerging Topics in Computing, 2014,2(4):399–410.
- [3] Chattopadhyay S, Banerjee A. QSCAS: QoS aware Web service composition algorithms with stochastic parameters. In: Proc. of the 2016 IEEE Int'l Conf. on Web Services (ICWS). IEEE, 2016. 388–395.
- [4] He Q, Han J, Chen F, Wang Y, Vasa R, Yang Y, Jin H. QoS-Aware service selection for customisable multi-tenant service-based systems: Maturity and approaches. In: Proc. of the 2015 IEEE 8th Int'l Conf. on Cloud Computing. IEEE, 2015. 237–244.
- [5] Dahan F, El Hindi K, Ghoneim A. Enhanced artificial bee colony algorithm for QoS-aware Web service selection problem. Computing, 2017,99(5):507–517.
- [6] Liu SD, Meng XB. Approach to network services recommendation based on mobile users' location. Ruan Jian Xue Bao/Journal of Software, 2014,25(11):2556–2574 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4561.htm> [doi: 10.13328/j.cnki.jos.004561]
- [7] Xu Y, Yin J, Deng S, Xiong N, Huang J. Context-Aware QoS prediction for Web service recommendation and selection. Expert Systems with Applications, 2016,53:75–86.
- [8] Su K, Xiao B, Liu B, Zhang H, Zhang Z. TAP: A personalized trust-aware QoS prediction approach for Web service recommendation. Knowledge-Based Systems, 2017,115:55–65.
- [9] Shao LS, Zhou L, Zhao JF, Xie B, Mei H. Web service QoS prediction approach. Ruan Jian Xue Bao/Journal of Software, 2009, 20(8):2062–2073 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3375.htm> [doi: 10.3724/SP.J.1001.2009.03375]
- [10] Wang Y, He Q, Yang Y. QoS-Aware service recommendation for multi-tenant SaaS on the cloud. In: Proc. of the 2015 IEEE Int'l Conf. on Services Computing (SCC). IEEE, 2015. 178–185.
- [11] Hu Y, Peng Q, Hu X, Yang R. Time aware and data sparsity tolerant Web service recommendation based on improved collaborative filtering. IEEE Trans. on Services Computing, 2015,8(5):782–794.
- [12] Wang X, Zhu J, Zheng Z, Song W, Shen Y, Lyu MR. A spatial-temporal QoS prediction approach for time-aware Web service recommendation. ACM Trans. on the Web, 2016,10(1):No.7.
- [13] Zhang X, Wang Z, Lv X, Qi R. A clustering-based QoS prediction approach for Web service selection. In: Proc. of the 2013 Int'l Conf. on Information Science and Cloud Computing Companion (ISCC-C). IEEE, 2013. 201–206.
- [14] Wu C, Qiu W, Zheng Z, Wang X, Yang X. QoS prediction of Web services based on two-phase K-means clustering. In: Proc. of the 2015 IEEE Int'l Conf. on Web Services (ICWS). IEEE, 2015. 161–168.
- [15] Yu C, Huang L. CluCF: A clustering CF algorithm to address data sparsity problem. In: Proc. of the Service Oriented Computing and Applications. 2016. 1–13.
- [16] Sarwar B, Karypis G, Konstan J, Riedl J. Item-Based collaborative filtering recommendation algorithms. In: Proc. of the 10th Int'l Conf. on World Wide Web. ACM Press, 2001. 285–295.
- [17] Zheng Z, Ma H, Lyu M R, King I. Wsrec: A collaborative filtering based Web service recommender system. In: Proc. of the 2009 IEEE Int'l Conf. on Web Services (ICWS 2009). IEEE, 2009. 437–444.
- [18] Zhao Y, Pi Q, Luo C, Yan D. CAPred: A prediction model for timely QoS. In: Proc. of the 2015 IEEE Int'l Conf. on Web Services (ICWS). IEEE, 2015. 599–606.
- [19] Chen L, Yang J, Zhang L. Time based QoS modeling and prediction for Web services. In: Proc. of the Int'l Conf. on Service-Oriented Computing. Berlin, Heidelberg: Springer-Verlag, 2011. 532–540.

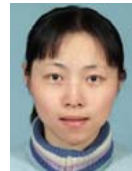
- [20] Lo W, Yin J, Li Y, Wu Z. Efficient Web service QoS prediction using local neighborhood matrix factorization. *Engineering Applications of Artificial Intelligence*, 2015,38:14–23.
- [21] Zhu J, Kang Y, Zheng Z, Lyu MR. A clustering-based QoS prediction approach for Web service recommendation. In: *Proc. of the 2012 15th IEEE Int'l Symp. on Object/Component/Service-Oriented Real-Time Distributed Computing Workshops*. IEEE, 2012. 93–98.
- [22] Chen F, Yuan S, Mu B. User-QoS-Based Web service clustering for QoS prediction. In: *Proc. of the 2015 IEEE Int'l Conf. on Web Services (ICWS)*. IEEE, 2015. 583–590.
- [23] Zhang M, Liu X, Zhang R, Sun H. A Web service recommendation approach based on QoS prediction using fuzzy clustering. In: *Proc. of the 2012 IEEE Ninth Int'l Conf. on Services Computing (SCC)*. IEEE, 2012. 138–145.
- [24] Kohonen T. The self-organizing map. *Proc. of the IEEE*, 1990,78(9):1464–1480.
- [25] Cottrell M, Olteanu M, Rossi F, Villa-Vialaneix N. Theoretical and applied aspects of the self-organizing maps. In: *Proc. of the Advances in Self-Organizing Maps and Learning Vector Quantization*. Springer Int'l Publishing, 2016. 3–26.
- [26] Zhang Y, Zheng Z, Lyu MR. WSPred: A time-aware personalized QoS prediction framework for Web services. In: *Proc. of the 2011 IEEE 22nd Int'l Symp. on Software Reliability Engineering (ISSRE)*. IEEE, 2011. 210–219.

#### 附中文参考文献:

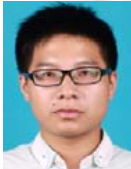
- [6] 刘树栋,孟祥武.一种基于移动用户位置的网络服务推荐方法.软件学报,2014(11):2556–2574. <http://www.jos.org.cn/1000-9825/4561.htm> [doi: 10.13328/j.cnki.jos.004561]
- [9] 邵凌霄,周立,赵俊峰,谢冰,梅宏.一种 Web Service 的服务质量预测方法.软件学报,2009,20(8):2062–2073. <http://www.jos.org.cn/1000-9825/3375.htm> [doi: 10.3724/SP.J.1001.2009.03375]



张以文(1976—),男,安徽马鞍山人,博士,副教授,CCF 专业会员,主要研究领域为服务计算,云计算,大数据.



贾兆红(1976—),女,博士,副教授,CCF 专业会员,主要研究领域为多目标优化算法,组合优化.



项涛(1989—),男,硕士,主要研究领域为服务推荐.



何强(1982—),男,博士,高级讲师,博士生导师,主要研究领域为面向服务的软件工程,大数据.



郭星(1983—),男,博士,讲师,CCF 专业会员,主要研究领域为服务计算,智能算法.