

基于协同过滤的自适应 Web 服务 QoS 预测方法

庄 崧¹, 郭志川^{2,3}, 黄逍颖^{2,3}

(1. 江苏有线技术研究院有限公司, 江苏 南京 210001;

2. 中国科学院声学研究所国家网络新媒体工程技术研究中心, 北京 100190;

3. 中国科学院大学, 北京 100049)

摘 要:随着 Web 服务越来越多,服务质量 QoS 作为描述 Web 服务的非功能性属性变得越来越重要。通常,一种服务的 QoS 对用户来说是未知的,因此对于基于 Web 服务的应用,精确预测其未知的 QoS 对于成功部署该服务具有重要的价值。基于协同过滤的 WSRec 算法是一种高精度的 QoS 预测方法,为进一步提升 QoS 的预测精度,提出了一种协同过滤的自适应 Web 服务 QoS 预测方法。该方法通过客户端首先发出 QoS-Web 服务请求;服务端接到请求后,根据已有数据,计算两两用户或服务间的相似度;并根据相似性找到对于目标用户的 K 个最接近用户或服务,生成该 QoS 值预测值 A ;同时在计算相似性时,采用改进皮尔逊相关系数得到预测值 B ;最后将预测值 A 和 B 以权值相结合得到目标用户或服务的 QoS 值。该算法改进了单一的协同过滤在数据稀疏的情况下,对相似性给予过高估计的不足,使得 QoS 预测值精度得以提高,取得了更好的实验结果。实验表明该方法预测精度优于 WSRec 算法。

关键词:协同过滤;Web 服务;QoS 预测;自适应;客户端

中图分类号:TP317

文献标识码:A

文章编号:1673-629X(2020)03-0093-05

doi:10.3969/j.issn.1673-629X.2020.03.018

An Adaptive Web Service QoS Prediction Method Based on Collaborative Filtering

ZHUANG Yin¹, GUO Zhi-chuan^{2,3}, HUANG Xiao-ying^{2,3}

(1. Jiangsu Cable Technology Research Institute Co., Ltd., Nanjing 210001, China;

2. National Network New Media Engineering Research Center, Institute of Acoustics,

Chinese Academy of Sciences, Beijing 100190, China;

3. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: With more and more Web services, quality-of-service (QoS) is becoming more and more important for as a non-functional attribute describing Web services. Generally, the QoS values of a service are unknown to its users, so the accurate prediction of unknown QoS values is significant for the successful deployment of Web service-based applications. WSRec algorithm based on collaborative filtering is a highly accurate method for predicting QoS. In order to improve the accuracy of QoS prediction further, an adaptive Web service QoS prediction method based on collaborative filtering is proposed. This method firstly sends QoS-based Web request to the server through the client. After receiving the request, the server calculates the similarity between each of the two users or between each of the two services based on the QoS data. At the same time, according to these similarities, the K closest users or services to the target user are found, and the predicted value A of QoS is generated. When calculating the similarity, the predicted Pearson correlation coefficient is used to obtain the predicted value B . Finally, the predicted values A and B are given to obtain the QoS value by changing the weight between the two values. The algorithm improves the accuracy of QoS prediction by improving the shortcomings of a single collaborative filtering to overestimate the similarity in the case of sparsely populated data, and obtains a better experimental result. The experiment shows that the proposed method achieves better prediction accuracy than WSRec algorithm.

Key words: collaborative filtering; Web service; QoS prediction; adaptive; client

收稿日期:2019-03-26

修回日期:2019-07-29

网络出版时间:2019-12-05

基金项目:中国科学院声学研究所“率先计划”项目(Y654101601);核高基项目(2014ZX01039101)

作者简介:庄 崧(1984-),男,硕士,工程师,研究方向为计算机软件与理论;郭志川,博士,研究员,研究方向为计算机软件与理论、网络多媒体。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20191205.1113.022.html>

0 引言

随着互联网技术的发展,Web 服务的便捷性导致用户对它的需求逐渐增大。如今大量的 Web 服务充斥互联网,在增加用户选择的同时,也提出了更高的要求:面对相同种类和条件下的服务,除了满足功能性需求之外,用户希望得到更高的服务质量(QoS)。目前,服务 QoS 预测已经成为 Web 服务领域的一个热点问题^[1-6]。这需要根据一些指标来对已有的服务进行衡量,在这些指标的对比之下,根据每一个用户的具体需求来对用户进行个性化的服务推荐,缩小用户的选择范围,从而更高效地享受相关服务。

QoS 度量有多种方法^[7-12],服务器端度量的 QoS 值(价格、流行度等等),服务提供商通常会在广告中展示内容,这些对于不同的用户来说都是相同的,但是在客户端度量的 QoS 值(响应时间、吞吐量、服务可用性等)在不同的用户之间却会因受到不可预测的网络连接和完全不同的用户环境的影响而大相径庭。需要对不同的用户获得精确的个人用户端 Web 服务的 QoS 值,而 QoS 值的准确测量需要对用户端得到的 Web 服务进行评估,因此选择一种合适的评估方法是非常必要的。但是在实际操作中,从用户角度去评测 Web 服务是很困难的。它具有如下缺点:(1)为了得

到 QoS 信息,用户需要执行相关的服务调用,服务提供者可能会收取调用费用,同时会增加资源消耗;(2)随着网络上的 Web 服务数量的日益增多,测评所有的 Web 服务将很耗时;(3)为了连续监视 Web 服务的 QoS 性能,服务用户需要定期调用评估服务;(4)为了深度评测 Web 服务,服务用户还需要做更多的工作。因此,为了更好地对用户端 QoS 进行评价从而做到推荐,目前最流行的做法是对用户端的 QoS 以预测的方式进行,从而使用户不需要再对每个 Web 服务进行测评,并且可节约时间。

采用协同过滤的手段已被证明是当前预测精度最高的方法^[13-16]。WSRec 算法具有很好的效果^[16],但是目前单一采用协同过滤的预测精度不高,尤其是在矩阵较为稀疏的情形下。基于此,文中将多种协同过滤方法进行动态组合,给出了一种基于协同过滤的自适应 QoS 服务预测方法,以提升用户体验。

1 算法描述

提出一种基于协同过滤的自适应 QoS 服务预测方法,解决现有的 Web 服务 QoS 预测方法当中精度较低的问题。方法流程如图 1 所示。

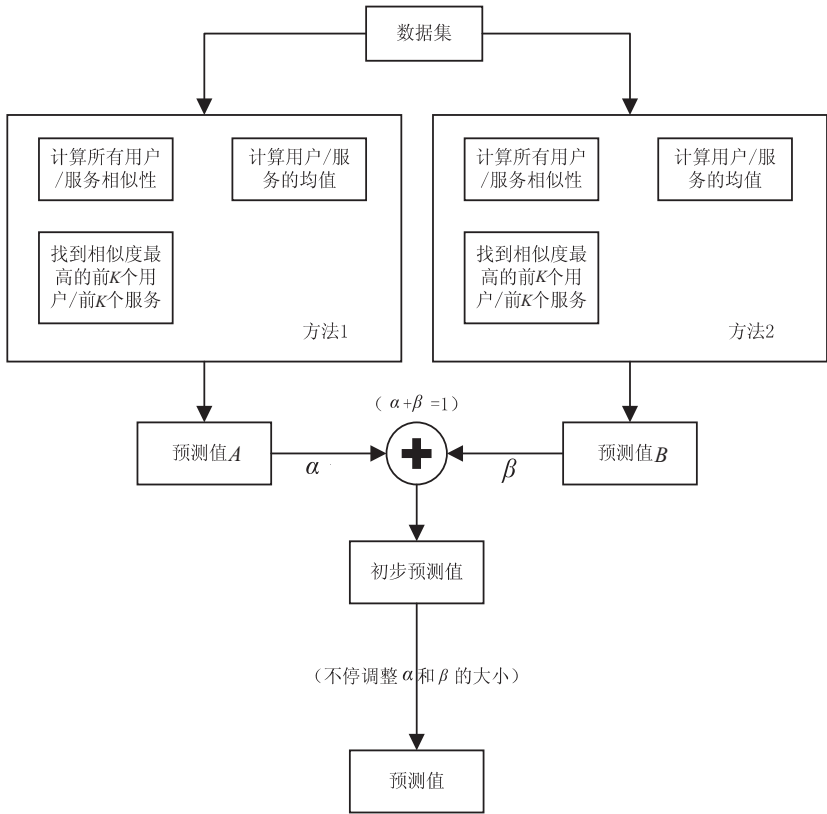


图 1 基于协同过滤的自适应 Web 服务 QoS 预测方法的流程

具体的算法描述如下:
(1)用户端向服务端提出基于 QoS 的 Web 服务请求。

(2)服务端根据用户端提出的 Web 服务请求,以及已有的用户端的其他 Web 服务 QoS 数据,以及与该用户端相类似的其他客户端的 Web 服务 QoS 数据,生

成用户-服务矩阵,其中每一列代表一种 Web 服务,每一行对应一个用户。

(3)在该用户-服务矩阵中,计算两两用户以及两两服务之间的相似度,然后对相似结果进行聚类分析;相似度的计算采用的是皮尔逊相关系数:

$$\text{Sim}(u_i, u_j) = \frac{\sum_{s \in S} (r(u_i, s) - \bar{r}_i) \cdot (r(u_j, s) - \bar{r}_j)}{\sqrt{\sum_{s \in S} (r(u_i, s) - \bar{r}_i)^2} \cdot \sqrt{\sum_{s \in S} (r(u_j, s) - \bar{r}_j)^2}} \quad (1)$$

$$\text{Sim}(s_i, s_j) = \frac{\sum_{u \in U} (r(s_i, u) - \bar{r}_{s_i}) \cdot (r(s_j, u) - \bar{r}_{s_j})}{\sqrt{\sum_{u \in U} (r(s_i, u) - \bar{r}_{s_i})^2} \cdot \sqrt{\sum_{u \in U} (r(s_j, u) - \bar{r}_{s_j})^2}} \quad (2)$$

其中, u 表示用户, i 表示服务, u_i 表示第 i 个用户, s_i 表示第 i 个服务。 $\text{Sim}(u_i, u_j)$ 表示用户 u_i 和 u_j 的相似度, $\text{Sim}(s_i, s_j)$ 表示服务 s_i 和 s_j 的相似度。 $r(u_i, s)$ 表示用户 u_i 对每个服务的 QoS 值, $r(s_i, u)$ 表示每个用户对服务 s_i 的 QoS 值。

(4)根据相似性找到对于目标用户的 K 个最接近用户,或者找到对于目标服务的 K 个最接近服务。记录下它们对应的相似度值。

(5)把步骤4记录的相似度值与 K 个用户(或者 K 个服务)的均值相结合,利用如下公式得到对该 QoS 值所做出的预测值 A :

$$\hat{r}(u, s) = \bar{u} + \frac{\sum_{i=1}^k \text{Sim}(u_i, u) \cdot (r(u_i, s) - \bar{u})}{\sum_{i=1}^k |\text{Sim}(u_i, u)|} \quad (3)$$

$$\hat{r}(u, s) = \bar{s} + \frac{\sum_{i=1}^k \text{Sim}(s_i, s) \cdot (r(u, s_i) - \bar{s})}{\sum_{i=1}^k |\text{Sim}(s_i, s)|} \quad (4)$$

其中, $\hat{r}(u, s)$ 表示用户 u 对于服务 s 的预测值, \bar{u} 表示用户 u 在所有服务的 QoS 均值, \bar{s} 表示所有用户在服务 s 上的 QoS 均值。

基于混合协同过滤的 QoS 服务预测方法,特征在于首先计算两两用户以及两两服务之间的相似度,然后对相似结果进行聚类分析,找到对于目标用户的 K 个最接近用户以及对于所需 Web 服务最接近的 K 个项目,并把计算出的相似度与 K 个用户(或者 K 个服务)的均值相结合,形成对该 QoS 值所做出的预测值 A 。

(6)同步步骤1~步骤5,不同之处是在计算两两之间的相似性的时候,使用一种改进的 UPCC 和 IPCC 相关系数:

$$\text{Sim}'(a, u) = \frac{2 \times |I_a \cap I_u|}{|I_a + I_u|} \text{Sim}(a, u) \quad (5)$$

$$\text{Sim}'(i, j) = \frac{2 \times |U_a \cap U_u|}{|U_i| + |U_j|} \text{Sim}(i, j) \quad (6)$$

从而得到预测值 B 。

其中, $|I_a \cap I_u|$ 是用户 a 和用户 u 共同调用的服务, I_a 是用户 a 调用的服务, I_u 是用户 u 调用的服务。 $|U_a \cap U_u|$ 是共同调用服务 i 和 j 的用户, U_i 是调用服务 i 的用户, U_j 是调用服务 j 的用户。

(7)把预测值 A 和 B 以一定的权值相结合,得到目标用户(或服务)的 QoS 值:

$$\hat{r} = \alpha A + \beta B \quad (7)$$

$$\alpha + \beta = 1 \quad (8)$$

通过自适应调整权值 α 与 β 的大小,得到较好的 QoS 预测值。

以上所述的基于混合协同过滤的 QoS 服务预测方法,通过预测值 A 和 B 以一定的权值相结合得到目标用户(或服务)的 QoS 值。在不同的数据集里,通过调整权值系数的值,得到该条件下最优的 QoS 预测方案。

2 实验结果

基于 WSDream(339 * 5 825) 数据集来进行相应的实验, WSDream 是一个公开的 QoS 方面的数据集。将数据集文件分别以 8 : 2 的比例,划分为训练数据集和测试数据集。然后以训练数据集为输入,输出测试数据集中各用户-项目的预测评分。最后对预测评分进行协同过滤的评价指标评估:

平均绝对偏差(MAE)计算公式如下:

$$\text{MAE} = \frac{\sum_{i,j} |r_{i,j} - \hat{r}_{i,j}|}{N} \quad (9)$$

均方根误差(RMSE)计算公式如下:

$$\text{RMSE} = \sqrt{\frac{\sum_{i,j} (r_{i,j} - \hat{r}_{i,j})^2}{N}} \quad (10)$$

这两个指标主要是通过计算预测值与实际值之间的偏差来评估目标方法的预测精度。

表2~表4分别是采样10%、20%、30%作为测试集之后得到的结果。结合3张表发现,文中方法在 MAE、RMSE 指标上优于单独使用 WSRec(一种改进的协同过滤)方法,如表1所示。结合每张表上 α 的取值,在 $\alpha = 0.65$ 附近时, MAE 和 RMSE 可以达到最低值。

表2~表4显示了文中方法中 α 的取值对于预测精度(MAE/RMSE)的影响(分别为抽取10%/20%/30%的数据集作为测试集)。结合每张表上 α 的取值,在 $\alpha = 0.65$ 时, MAE 和 RMSE 可以达到最优值。

表 1 改进方法与 WSRec 方法的比较

指标	抽样 10% 做测试集		抽样 20% 做测试集		抽样 30% 做测试集	
	WSRec	改进方法	WSRec	改进方法	WSRec	改进方法
MAE	0.856 965	0.851 608	0.776 001	0.771 63	0.732 832	0.727 988
RMSE	2.017 53	2.015 75	1.872 38	1.866 49	1.794 85	1.787 73

表 2 α 取值与预测精度的关系 (抽取 10% 作为测试集, α 的最优值: $\alpha = 0.65$)

指标	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90
MAE	0.856	0.855	0.855	0.854	0.854	0.853	0.852	0.852	0.852	0.851	0.851	0.851	0.85	0.85	0.85	0.85	0.85	0.85
RMSE	2.017	2.017	2.017	2.016	2.016	2.016	2.016	2.016	2.016	2.016	2.016	2.016	2.016	2.017	2.017	2.017	2.017	2.018

表 3 α 取值与预测精度的关系 (抽取 20% 作为测试集, α 的最优取值范围: $0.4 < \alpha < 1$)

指标	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90
MAE	0.775	0.775	0.774	0.774	0.773	0.773	0.773	0.772	0.772	0.772	0.772	0.772	0.772	0.772	0.772	0.772	0.772	0.772
RMSE	1.872	1.871	1.87	1.87	1.869	1.868	1.868	1.868	1.867	1.867	1.867	1.867	1.867	1.867	1.867	1.867	1.867	1.867

表 4 α 取值与预测精度的关系 (抽取 30% 作为测试集, α 的最优取值范围: $0.5 < \alpha < 1$)

指标	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90
MAE	0.732	0.732	0.731	0.731	0.73	0.73	0.73	0.729	0.729	0.728	0.728	0.728	0.728	0.728	0.728	0.728	0.728	0.728
RMSE	1.794	1.793	1.792	1.792	1.791	1.79	1.79	1.79	1.79	1.789	1.788	1.788	1.788	1.788	1.788	1.788	1.788	1.788

针对不同的抽样率,MAE 与 α 取值的曲线关系如 图 2 ~ 图 4 所示。

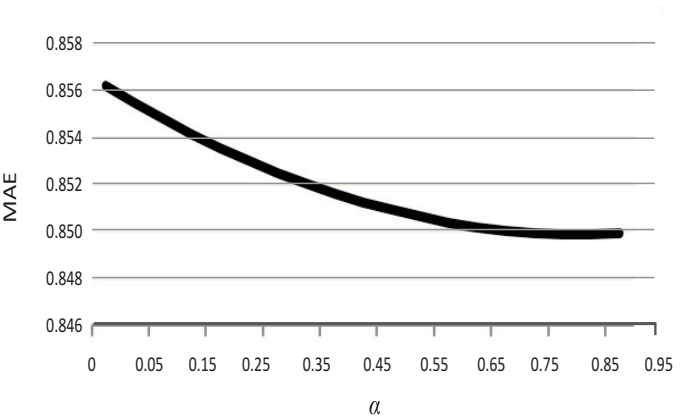


图 2 10% 抽样关系

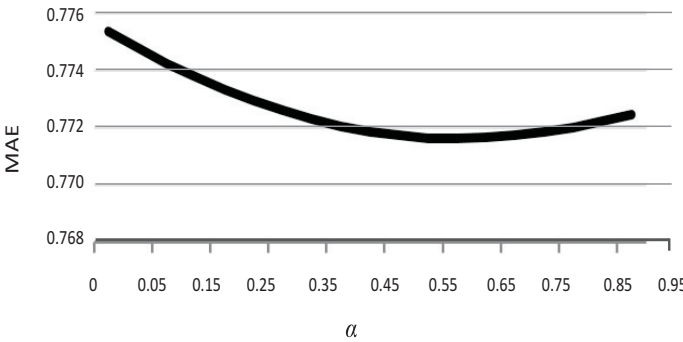


图 3 20% 抽样关系

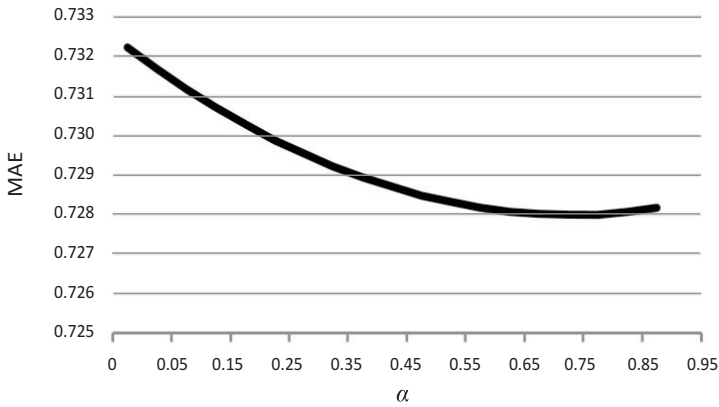


图4 30% 抽样关系

3 结束语

采用协同过滤的手段是当前预测精度较好的方法,但是目前单一采用协同过滤的预测精度不高,尤其是在矩阵较为稀疏的情形下。该文有效改进了这一问题,使得 QoS 预测精度得到提高。提出的采用两种协同过滤进行混合的算法,可以通过自适应调整权值的大小,提高协同过滤算法的精度。该算法改进了皮尔逊相关系数法可能在数据稀疏的情况下对相似性给予过高估计的不足,从而提高了最终的 QoS 预测值的精度,同时指标上也优于 WSR_{ec} 方法。

参考文献:

[1] 张龙昌,杨艳红. 动态 QoS 数据驱动的可靠 Web 服务选择[J]. 电子与信息学报,2016,38(6):1368-1376.

[2] 汪 浩,肖建茂,龙 浩,等. 基于 Bootstrap 技术的 Web 服务 QoS 值的置信区间估计和预测[J]. 电子学报,2018,46(3):665-671.

[3] 王文彬,孙其博,赵新超,等. 基于非均衡变异离散粒子群算法的 QoS 全局最优 Web 服务选择方法[J]. 电子学报,2010,38(12):2774-2779.

[4] 刘 枫,雷振明,刘 芳. 用户感知的 Web 业务 QoS 指标与测量[J]. 北京邮电大学学报,2009,32(2):97-100.

[5] 吴江霞,杨放春. 支持事务机制的 Web 服务组合 QoS 属性预测方法[J]. 电子与信息学报,2008,30(3):703-706.

[6] JIANG W, WU T, HU S L, et al. QoS-aware automatic service composition: a graph view[J]. Journal of Computer Science and Technology, 2011, 26(5): 837-853.

[7] DAI Y, YANG L, ZHANG B. QoS-driven self-healing web service composition based on performance prediction[J]. Journal of Computer Science and Technology, 2009, 24(2):

250-261.

[8] ZHANG M W, ZHANG B, LIU Y, et al. Web service composition based on QoS rules[J]. Journal of Computer Science and Technology, 2010, 25(6): 1143-1156.

[9] 李 玲,刘 敏,成国庆. 一种基于 FAHP 的多维 QoS 局部最优服务选择模型[J]. 计算机学报,2015,38(10):1997-2010.

[10] 庄 媛,张鹏程,李雯睿,等. 一种环境因素敏感的 Web-Service QoS 监控方法[J]. 软件学报,2016,27(8):1978-1992.

[11] 何志鹏,张鹏程,江 艳,等. 一种时效感知的动态加权 Web 服务 QoS 监控方法[J]. 软件学报,2018,29(12):3716-3732.

[12] 马 友,王尚广,孙其博,等. 一种综合考虑主客观权重的 Web 服务 QoS 度量算法[J]. 软件学报,2014,25(11):2473-2485.

[13] MA Y, WANG S, HUNG P C K, et al. A highly accurate prediction algorithm for unknown web service QoS values[J]. IEEE Transactions on Services Computing, 2016, 9(4): 511-523.

[14] LUO X, ZHOU M, XIA Y, et al. Predicting web service QoS via matrix-factorization-based collaborative filtering under non-negativity constraint[C]//2014 23rd wireless and optical communication conference (WOCC). Newark, NJ: IEEE, 2014: 1-6.

[15] WU C, QIU W, WANG X, et al. Time-aware and sparsity-tolerant QoS prediction based on collaborative filtering[C]//2016 IEEE international conference on web services. San Francisco, CA: IEEE, 2016: 637-640.

[16] ZHENG Z, 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.