

基于辅助特征的 Web 服务 QoS 预测^①

黄坤^{1,2}, 王丹丹¹, 崔强^{1,2}, 郝春亮^{1,2}, 王青¹

¹(中国科学院软件研究所 基础软件国家工程研究中心, 北京 100190)

²(中国科学院大学, 北京 100190)

摘要: 随着互联网以及 Web 服务技术的快速发展, 相同功能的 Web 服务数量越来越多. 在构建面向服务的应用时, 服务质量(QoS)作为 Web 服务的非功能特性开始被越来越多的用户所重视. 为了向用户推荐高质量的服务, 首先我们需要对服务质量进行预测. 现今有很多关于 Web 服务 QoS 预测的工作, 这些研究大都关注在建模方法的优化上, 忽视了辅助特征对于 QoS 预测的影响. 着重分析辅助特征对于 QoS 预测的影响, 例如服务类别和用户地理位置. 为了实现此目标, 基于因子分解机(Factorization Machines)设计并构建了一个统一的 QoS 预测架构, 该架构可以灵活、方便地考虑进多个辅助特征. 结合服务类别和用户地理位置这两类辅助特征, 提出了一种 QoS 预测方法, 并通过在真实数据上的实验证明了我们的方法的优越性.

关键词: Web 服务; QoS 预测; 辅助特征; 基于因子分解机

Web Service QoS Prediction Based on Auxiliary Features

HUANG Kun^{1,2}, WANG Dan-Dan¹, CUI Qiang^{1,2}, HAO Chun-Liang^{1,2}, WANG Qing¹

¹(Institute of Software, Chinese Academy of Sciences, Beijing 100190, China)

²(University of Chinese Academy of Sciences, Beijing 100190, China)

Abstract: With the rapid growth of Internet and Web service technology, the number of Web services having the same function is getting more and more. When constructing service-oriented applications, Quality of Service(QoS) as the non-functional properties of services is attracting more and more attention from users. To recommend high quality services to users, first we need to predict the quality of services. Now there are many research works on Web service QoS prediction, but most of these works focus on the optimization of modeling methods, which ignore the impact of auxiliary features. This paper emphatically analyzes auxiliary features' impact on QoS prediction such as service category and user location. To achieve this goal, in this paper designs and builds up a unified QoS prediction framework based on Factorization Machines, which can incorporate multiple auxiliary features easily and conveniently. Combined with service category and user location, this paper develops a QoS prediction method, which is proved to be advantageous via experiments on real-life data sets.

Key words: Web services; QoS prediction; auxiliary feature; factorization machines

1 引言

Web 服务是一种松耦合的能够通过标准通信协议来支持网络中异构平台之间互操作的软件系统^[1]. 得益于其跨语言、跨平台的特点, Web 服务被企业和个人开发者广泛用于构建面向服务的应用. 当开发者(用户)需要开发面向服务的应用时, 首先, 他们需要根据需求

设计出业务流程, 然后只需要通过 Web 服务搜索引擎寻找到符合功能需求的服务进行复用, 以此来构建整个业务流程, 这样大大缩短了开发周期且节省了资源. 然而, 当开发者面对大量具有相同或者相似功能的 Web 服务时, 如何选择最佳的 Web 服务用于构建 Web 应用便成了一个难题.

^① 基金项目: 国家自然科学基金(91218302, 91318301)

收稿时间: 2016-01-26; 收到修改稿时间: 2016-03-08 [doi:10.15888/j.cnki.csa.005364]

服务质量(Quality of service, QoS)是一系列用户可以感知的属性集合,包括响应时间、可用性、吞吐率等,它被广泛用来表示 Web 服务的非功能特性。许多研究者提出 QoS 属性应该作为衡量面向服务的应用是否构建成功的关键因素之一^[2,3]。近些年,许多基于 QoS 的方法不断被研究者们提出并且成功应用于 Web 服务选择、发现、组合、推荐、路由等多个研究领域^[4-8]。这些方法的成功应用都建立在一个共同前提的基础之上,即它们都需要准确的 QoS 属性值作为基础。然而,现实世界中这个前提却经常受到挑战:

1) 真实的网络环境中让终端用户探索所有的 Web 服务是非常耗时的,同时也耗费了 Web 服务提供商的大量资源。

2) 大多数的 Web 服务提供商都是商业公司,让用户通过调用 Web 服务来获得 QoS 的相关信息,这种方式对于终端用户来说是非常耗费金钱的。

3) 由于真实的网络环境往往是不稳定的并且 Web 服务的服务质量和调用时间密切相关,因此想要一直收集 QoS 属性记录是不切实际的。

因此,为了满足上述面向服务研究领域的基本需求,如何准确地预测 QoS 值是一个亟待解决的问题。

协同过滤(Collaborative Filtering, CF)技术是推荐系统领域中常用的一种推荐方法,其核心思想是通过用户的历史偏好记录发掘出相似的用户或者物品(邻居),然后通过相似邻居的偏好得分的加权平均作为目标用户对物品的评分。研究者们已经提出了多种协同过滤算法,一般来讲可以将协同过滤算法分为两种类别:基于内存的协同过滤和基于模型的协同过滤。基于内存的协同过滤算法也被称为基于邻域的方法,即通过相似度较高的一部分邻居用户或者物品来帮助预测偏好值,它又可以细分为基于用户的方法^[9]和基于物品的方法^[10]。对于基于用户的方法是向活跃用户推荐和他相似用户喜欢的物品,对于基于物品的方法是向活跃用户推荐和他喜欢物品相似的其他物品。基于模型的协同过滤算法一般利用机器学习或者数据挖掘方法训练得到一个预测模型,然后用这个预测模型来预测用户对目标物品的偏好值,常用的基于模型的协同过滤算法包括贝叶斯模型、潜在因子模型、主题模型等。

受推荐系统中广泛采用的协同过滤技术的启发,很多研究者也采用协同过滤的方法来进行 QoS 属性值

的预测,并且取得了不错的效果^[11-13]。然而,先前的这些方法模型都存在着一个局限:它们并不能方便、有效地利用到其它特征对于 QoS 预测所起到的作用。这是因为经典的基于内存的协同过滤或者基于矩阵分解的协同过滤,其中只有两种类型的特征,一种特征是 User,一种特征是 Service,并不能处理多特征的情况,而合适的辅助信息(比如用户所在的地理位置、Web 服务的类别)对于 QoS 的预测是能够起到帮助作用的。虽然有一些针对用户地理位置信息或者时间信息进行利用的模型^[14-18],但是这些模型都是针对特定情形的应用而开发的,对于其他挖掘到的特征并不能够利用上,拓展性和通用性较差。

Factorization Machines(FM)^[24]是推荐系统领域里近几年比较受欢迎的一种通用预测模型,它能够充分利用辅助特征对于预测任务的帮助,并且具有很好的特征拓展性。在本文中,我们借助 FM 来解决 QoS 预测问题,并且在实验中通过加入有效的辅助特征(用户所在的地理位置、Web 服务的类别)来改善预测效果。

总的来说,本文的贡献如下:

1) 本文基于 Factorization Machines 模型设计并实现了一个统一的 Web 服务 QoS 预测系统架构,它能够灵活、方便地考虑进多个特征,充分地利用辅助特征在 QoS 预测当中所起到的作用。

2) 通过真实的实验数据分析发现用户的地理位置(L)和服务的类别(C)会影响到 Web 服务质量,我们将这两个维度的信息作为辅助特征加入 FM 模型当中提出了 FM-L,FM-C,FM-LC 方法,能够改善 QoS 属性预测效果。

2 相关工作

随着 QoS 属性作为 Web 服务的非功能特性越来越得到重视,近些年,许多不同的基于 QoS 预测的 Web 服务推荐方法相继被研究者提出,比如基于语义的,基于链接预测的,和基于协同过滤的,其中协同过滤技术由于其简单有效性获得了最多的关注。Shao 等提出通过基于用户的协同过滤算法来解决 QoS 预测问题^[19]。Zheng 等在此基础上提出了基于用户和基于物品的混合协同过滤方案^[20],进一步改善了 QoS 预测的准确度。随后,Zheng 等认为按照皮尔逊相关系数计算相似度并不是一种合适的方法,因为皮尔逊相关系数并不能区分开拥有相同相似度但是物品交集个数不

同的情况,因此提出了一种带有权重的皮尔逊相关系数方法^[18].与以上基于内存的协同过滤方法不同,Wei等使用基于模型的协同过滤中的SVD矩阵分解方法来进行QoS预测^[13],并且加入了用户的相似邻居和服务的相似邻居作为正则项,通过相似邻居在矩阵分解中的约束作用改善预测效果.随后,Zheng等提出了另外一种集成用户邻居到矩阵分解当中的方法^[21],其具体做法是在将用户服务对应的QoS记录矩阵分解为用户潜在因子矩阵和服务潜在因子矩阵乘积的基础之上,加上用户相似邻居对应的矩阵分解值,并采用调和参数的方式将二者融合.不过,之前的这些方法都存在着一个问题:不论是基于内存的协同过滤还是基于矩阵分解的协同过滤,它们只能利用到User和Service这两种特征,对于其他的特征并不能直接利用上,虽然有人提出了一些面向特定应用情形的方法^[16,17,22,23],能够间接利用上其他特征来帮助QoS预测,但是这些方法要么是拓展性不好,对于新挖掘到的特征并不适用,要么是计算时间复杂度随特征数目呈指数增长而不可取.

Factorization Machines模型曾被同时用在KDD CUP 2012中完成推荐和点击率预估任务,并且都取得了非常优异的成绩,是一个灵活、通用、拓展性很强的分类预测模型^[24].当前一些比较成功的其他协同过滤算法比如MF,SVD++,PITF,FPMC,这些模型算法实际上可以看做是FM的子集^[24].FM模型能够支持多维特征,在线性模型的基础上增加了特征之间的组合交互关系,并且对于特征之间的组合交互关系采用矩阵分解的方法来求解,这样既发挥了特征在分类预测问题中起到的作用,又通过矩阵分解的方法利用到了相似邻居的群体智慧.正是由于这种优势,本文借助FM模型来解决Web服务QoS预测问题,通过加入辅助特征来改善QoS预测效果.

3 辅助特征对Web服务质量的影响

Web服务质量并不是由单一因素决定的,它可能会受到诸多因素的影响,比如网络状况、网络运营商、Web服务提供商、用户所在地理位置、Web服务类别等,都可能会影响到用户调用Web服务时的服务质量.在本文中,由于我们获取到的数据中的辅助特征只有有Web服务类别和用户所在地理位置,因此我们考虑Web服务类别和用户所在地理位置这两个维度

的辅助特征对于QoS属性的影响.如图1所示,是本文所用的真实的QoS记录中不同类别的服务对应的平均响应时间,可以看到不同类别的服务之间平均的响应时间差异(方差)是比较大的.如图2所示,是本文所用数据集QoS记录中不同地理位置(以城市为粒度)的用户调用Web服务时对应的平均响应时间,可以看到不同地理位置的用户发送服务请求时对应的平均响应时间差异(方差)也非常大.这些统计数据形象地说明了辅助特征会影响到Web服务质量,因此,这就有必要设计出一个能够方便、灵活地考虑进辅助特征的Web服务QoS预测系统.

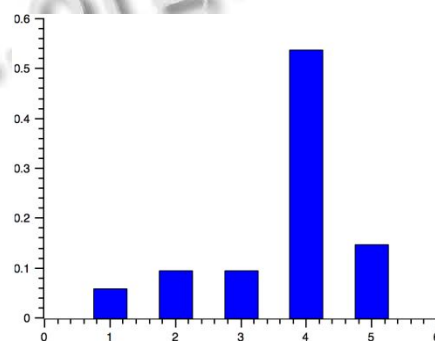


图1 不同类别服务的平均响应时间(单位:秒)

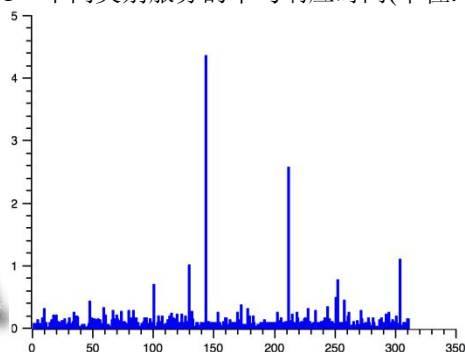


图2 不同城市的服务请求平均响应时间(单位:秒)

4 基于辅助特征的Web服务QoS预测

在本文中我们利用辅助特征在QoS预测当中所起到的帮助作用,并结合历史QoS记录进行QoS预测.假设 n 个用户 $\{u_1, u_2, u_3, \dots, u_n\}$ 组成的用户集为 U , m 个Web服务 $\{s_1, s_2, s_3, \dots, s_m\}$ 组成的服务集为 S ,则构成了一个 $n \times m$ 的用户-服务QoS矩阵 R (如表1所示),矩阵中的元素 R_{ij} 表示的是Web服务 j 被用户 i 调用时的QoS属性值.注意到在矩阵 R 中会有很多缺失的QoS属性值(表1中灰色部分),我们的目标就是通过构建出一个统一的架构,充分发挥辅助特征的作用,更加准确地对这些缺失的属性值进行预测.与文献[16]一样,

我们以 Web 服务响应时间作为具体的 QoS 属性例子，但是我们设计的 Web 服务 QoS 预测系统同样也可以预测其他 QoS 属性，比如吞吐量、可用性等等。

表 1 用户-服务 QoS 矩阵

	s ₁	s ₂	...	s _m
u ₁	100		200	300
u ₂		400	500	150
...	420			
u _n	30		85	600

4.1 系统架构

与以往的 QoS 预测系统只关注历史的用户服务 QoS 记录不同，本系统还将用户相关以及 Web 服务相关的其他属性作为辅助特征考虑进来，整体的系统架构如图 3 所示。

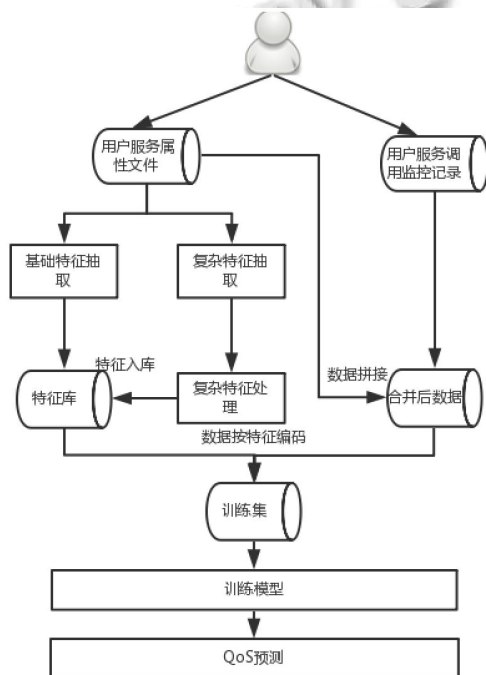


图 3 基于辅助特征的 Web 服务 QoS 预测系统架构图

首先，用户服务调用监控记录存放的是用户每次调用一个 Web 服务时对应的 QoS 属性值(比如响应时间、吞吐量、可用性等等)情况，用户服务属性文件存放的是与用户相关的属性数据以及与服务相关的属性数据。对于用户服务属性文件，我们可以从中抽取发掘出有利于 QoS 预测的特征属性，具体而言根据特征的情况可分为基础特征抽取和复杂特征抽取。基础特征抽取主要是针对一些比较简单、不用经过复杂处理

就能够用来协助 QoS 预测的特征，比如服务的类别可以直接作为特征使用。复杂特征抽取针对的是一些并不能直接利用上的特征，这些特征需要进行进一步的特征处理才能发挥作用，比如用户的 IP 信息，直接作为特征对于 QoS 预测并不能起到帮助作用，相反可能会起到副作用，但是如果将用户的 IP 信息映射成用户的地理位置信息获取到用户所在的国家或者城市就能够起到帮助作用，因为网络服务的质量显然是会受到用户所在位置的影响的。之后，将用户服务属性文件与用户服务调用监控记录进行合并，按照特征编码生成训练数据集，根据选定的模型进行训练并用于 QoS 预测，在本系统中，采用了 Factorization Machines 模型作为预测模型，关于 FM 的详细介绍将在 4.2 节中阐述。当缺失的 QoS 属性值预测完毕之后，便可以用于 Web 服务推荐。

4.2 预测模型

Factorization Machines 是一个通用的预测模型^[24]，它能够处理任意实值特征向量，并且能够方便、灵活地加入辅助特征。FM 的另外一个特点是即便是在数据极度稀疏的情况下，也能够以线性时间学习到各个特征之间两两组合时对应的交互关系，这也是为什么 FM 在大规模稀疏性数据上依然能够表现出色的原因。与经典的协同过滤模型中用户和物品的交互关系是通过一个评分矩阵来表示的不同，在 FM 中用户和物品之间的交互关系是通过一个特征向量来表示的，用户对物品的评分可以看作是特征向量对应的得分。具体地，对于 Web 服务 QoS 预测问题中的数据，可以表示成元组(x,y)组成的集合 S，其中 $\mathbf{x}=(x_1, \dots, x_n) \in \mathbb{R}^n$ 是一个 n 维特征向量，y 是特征向量 x 对应的 QoS 属性值。当不考虑辅助特征时，特征向量 \mathbf{x} 就是一个用户-服务对，其表达方式如下：

$$\mathbf{x} = (\underbrace{0, \dots, 1, \dots, 0, 0, \dots, 1, \dots, 0}_{|U|}, \underbrace{0, \dots, 1, \dots, 0}_{|I|})$$

其中，只有第 i 个元素和第 j 个元素为 1，表示活跃用户 i 调用了服务 j。当需要将辅助特征考虑进来时，可以将上述特征向量 \mathbf{x} 的表达形式进行拓展：

$$\mathbf{x} = (\underbrace{0, \dots, 1, \dots, 0, 0, \dots, 1, \dots, 0}_{|U|}, \underbrace{0, \dots, 1, \dots, 0}_{|I|}, Z(\mathbf{u}), Z(\mathbf{i}))$$

其中 Z(u)是用户相关的属性，Z(s)是服务相关的属性，在本文中，我们使用到的用户相关属性是通过用户的 IP 信息获取到的用户地理位置信息(城市编号)，使用

到的服务相关属性是服务的类别信息(类别编号). 通过加入合适的辅助特征, FM 不仅能够发挥该特征对于 QoS 预测的作用, 而且能够捕获到各个特征之间的组合交互关系.

本文所采用的 FM 模型的组合特征维度 $d=2$, 即只考虑两两特征之间组合时的交互关系. 这样, FM 模型就可以表示成以下形式:

$$\hat{y}(x) = w_0 + \sum_{j=1}^n w_j x_j + \sum_{j=1}^n \sum_{j'=j+1}^n w_{j,j'} x_j x_{j'} \quad (1)$$

其中, $w_0, w_j \in \mathbb{R}$ 是模型参数, w_0 表示全局偏差, w_j 表示特征 x_j 的权重, $w_{i,j} \in \mathbb{R}$ 是特征 x_j 与 $x_{j'}$ 之间的交互关系参数, 定义 $w_{i,j} = \langle \mathbf{v}_j, \mathbf{v}_{j'} \rangle = \sum_{f=1}^k v_{j,f} v_{j',f}$, \mathbf{v}_j 是特征 x_j 的 k 维因式化向量, 将特征之间的交互关系 $w_{i,j}$ 通过因式分解的方式计算(矩阵分解), 这样使得 FM 即使在数据非常稀疏的情况下依然能够学习到特征之间的交互关系参数. 对于一个含有 n 维特征并且因式化向量为 k 维的 FM 来说, 需要学习的模型参数是 $\Theta = \{w_0, w_1, \dots, w_n, v_{11}, v_{12}, \dots, v_{1k}, \dots, v_{nk}\}$. FM 的模型参数学习方法主要由三种: 随机梯度下降(SGD)、交替最小平方(ALS)和蒙克卡罗采样(MCMC).

在文献[24]中已经证明, FM(方程(1))能够在线性时间 $O(k*n)$ 内完成特征之间组合时的交互关系计算, 因为方程(1)可以转化成以下的等价形式(k 是常数):

$$\hat{y}(x) := w_0 + \sum_{j=1}^n w_j x_j + \frac{1}{2} \sum_{j=1}^k \left(\left(\sum_{j=1}^n v_{j,f} x_j \right)^2 - \sum_{j=1}^n v_{j,f}^2 x_j^2 \right)$$

在上一节中, 我们通过真实的 QoS 数据的统计分析知道用户的地理位置和服务的类别会影响到用户调用服务时的响应时间, 而 FM 模型能够将这类上下文信息作为辅助特征考虑进目标函数的构造当中, 充分利用辅助特征对于预测任务的帮助. 受此启发, 我们基于 FM 模型提出了以下 QoS 预测方法:

FM-L: 将用户的地理位置信息考虑进来的 FM 模型.

FM-C: 将服务类别信息考虑进来的 FM 模型.

FM-LC: 同时考虑用户的地理位置信息和服务类别信息的 FM 模型.

具体地, 对于 FM-L 方法, 是在 FM 模型的基础上加入了和用户相关的辅助特征 $Z(u)$, $Z(u)$ 是通过分析用户的 IP 地址得到的用户地理位置信息, 这里取用户所在地的城市编号. FM-C 方法是在 FM 模型的基础上

加入了和服务相关的辅助特征 $Z(i)$, $Z(i)$ 通过服务的类别编号表示. FM-LC 方法则是同时考虑了用户的地理位置 $Z(u)$ 和服务类别 $Z(i)$ 的 FM 模型.

5 实验

5.1 数据说明

我们使用的实验数据来自于百度 ApiStore(apistore.baidu.com), 百度 ApiStore 是一个为开发者提供 API 服务资源的服务网站. 我们获取到的数据集中包含了 5 个类别的 143 个 API 服务, 10363 个用户(以 IP 标识), 一共 4391770 条服务调用记录. 对于同一用户对同一 API 服务的多次调用, 我们计算用户调用 API 时的平均响应时间, 最终得到了 11239 条用户服务平均响应时间记录. 对于以 IP 标识的用户, 我们采用百度 LBS 开放服务“IP 定位 API”将 IP 信息转换成地理位置信息并获得用户所在的城市编号, 10363 个用户用户分别分布在 310 个不同的城市. 具体的实验数据细节如表 2 所示.

表 2 实验数据细节描述

用户数	10363
服务数	143
用户-服务平均响应时间记录数	11239
平均响应时间	204.72ms
最短响应时间	9ms
最长响应时间	10248ms
用户城市分布数	310
服务类别数	5
数据稀疏度	99.24%

5.2 评估方法

我们使用流行的平均绝对误差(MAE)和均方根误差(RMSE)作为预测准确性的衡量标准. MAE 的定义形式如下:

我们使用流行的平均绝对误差(MAE)和均方根误差(RMSE)作为预测准确性的衡量标准. MAE 的定义形式如下:

$$MAE = \frac{1}{N} \sum_{i,j} |R_{ij} - \hat{R}_{ij}|$$

其中, R_{ij} 是用户 i 调用 API 服务 j 时的平均响应时间, \hat{R}_{ij} 是预测的响应时间, N 是预测值的个数. MAE 对于每个个体的偏差都给与相同的权重. RMSE 的定义形式

如下:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i,j} (R_{ij} - \hat{R}_{ij})^2}$$

RMSE 会放大相对较大的偏差的权重,因此它对于较大的偏差比较敏感.

5.3 效果对比

为了证明我们的方法的效果,我们将我们的方法和以下众所周知的方法进行了比较.包括:

UPCC^[9]: 基于用户的协同过滤算法.

IPCC^[10]: 基于物品的协同过滤算法.

Funk-SVD^[25]: 基本的 SVD 矩阵分解方法.

Bias-SVD^[25]: 同时融合了偏置项和 Funk-SVD 的方法.

以及本文提出的方法:

FM-C: 包含服务类别信息的 FM 方法.

FM-L: 包含用户地理位置信息的 FM 方法.

FM-LC: 同时包含服务类别和用户地理位置信息的 FM 方法.

需要指出的是 FM 模型在不添加辅助特征时是与 Funk-SVD 和 Biased-SVD 是等价的.

实验采用 5 折交叉验证,将数据集中 80% 的数据作为训练集,剩下 20% 的数据作为测试集^[26],实验的对比效果如表 3 所示.

表 3 实验效果对比

实验方法	MAE	RMSE
UPCC	97.25	330.23
IPCC	99.24	343.30
Funk-SVD	80.52	275.49
Biased-SVD	79.04	271.55
FM-C	72.66	256.37
FM-L	68.48	240.03
FM-LC	65.43	229.21

从表 3 中的实验效果对比,可以发现,基于模型的协同过滤方法普遍要比基于内存的协同过滤方法效果好.对于基于矩阵分解的协同过滤方法 Funk-SVD 和 Biased-SVD,它们的实验效果差不多,这是因为真实环境的实验数据非常稀疏,并不能够学习到很好的偏置项参数.通过实验可以发现,不论是加入 Web 服务的类别信息还是用户的地理位置信息都能够改善 QoS 的预测效果,尤其是用户的地理位置信息,通过

加入用户的地理位置信息的 FM-L 方法相比于矩阵分解方法(Funk-SVD、Biased-SVD)能够使得 MAE 降低 14% 左右,使得 RMSE 降低 12% 左右.当同时考虑进用户的地理位置信息和服务类别信息时,FM-LC 方法相比于矩阵分解方法能够在 MAE 上降低 18% 左右,在 RMSE 上降低 16% 左右,这也得益于 FM 模型能够充分学习到各个特征之间的组合关系.这种效果的提升是能够解释得通的,因为现实生活中的用户是分布在不同地理位置的,用户与目标服务器之间的距离越近,当然网络信号的传输时间就越短,在我们的系统中我们对这种距离的度量是通过用户和 Web 服务之间的交互关系参数来表示的,同样地,不同类别的服务的平均响应时间也存在着较大差异,像 OCR 这类图像识别服务本身就需要大量的计算处理时间,而天气查询类服务所需要的计算时间开销则较小从而使得服务响应时间较短.通过上述实验对比,我们充分说明了可以通过引入合适的辅助特征来改善 QoS 预测效果.

6 结论和未来工作

在本文中,我们首先讨论了以往的一些 Web 服务 QoS 预测方法的局限,即它们不能够很好的利用到辅助特征进行 QoS 预测,即使存在一些利用辅助信息的方法,但是这些方法都是针对特定情形的应用而开发的,其拓展性较差且不能够利用到新发掘的特征.所以,本文设计并实现了一个统一的 Web 服务 QoS 预测系统架构,对于辅助特性能够灵活、方便地考虑进来并加以利用.并且通过数据统计分析,我们发现用户的地理位置和服务的类别这两种辅助特征对 QoS 属性的影响较大,因此将这两种辅助特征考虑进来.在具体的问题求解时,采用了能够支持辅助特征的 Factorization Machines 作为预测模型,将用户的地理位置和服务的类别作为辅助特征融合进来,提出了新的预测方法 FM-L,FM-C,FM-LC,并和常用的 QoS 预测方法进行了实验效果对比,最终我们提出的方法取得了更好的预测效果.同时,我们也需要指出,并不是加入任何辅助特征都能够起到改善 QoS 预测效果的作用,这需要根据特征选择的方法^[27]筛选出有效的特征,比如本文中我们发现 QoS 响应时间在服务类别和用户地理位置这两个维度上的方差很大,这两个维度的特征是能够起到增益作用的.尽管如此,本文所提出的方法为我们利用辅助特征进行 QoS 属性预测提供了可

能和便利. 本文所设计的系统的计算复杂度主要来源于 FM 模型对于特征参数以及组合特征参数的学习, 其所需的时间复杂度为 $O(k*n)^{[24]}$, k 是因式化向量的维度(常数), n 是引入的特征的个数, 整体的计算时间与特征的个数呈线性关系, 具有较低的计算复杂度. 未来, 我们将会探索更多对于 QoS 预测有帮助的辅助特征, 比如用户对于 Web 服务的评论反馈信息以及服务提供商相关信息等, 具体而言, 可以针对用户对服务的响应时间的正向和负向评论给出一个综合评分作为评论反馈特征, 对于服务提供商可以将其认证的级别(比如个人、企业)作为辅助特征. 除此之外, 我们还会探索将我们的方法用于其他 QoS 属性预测的实验效果, 比如吞吐量、可用性等.

参考文献

- 1 Zhang LJ, Zhang J, Hong C. Services Computing. Springer & Tsinghua University Press, 2007: 32–35.
- 2 Moser O, Rosenberg F, Dustdar S. Non-intrusive monitoring and service adaptation for WS-BPEL. Proc. of the 17th International Conference on World Wide Web(WWW'08). ACM. 2008. 815–824.
- 3 Papazoglou MP, Georgakopoulos D. Introduction: Service-oriented computing. Communications of the ACM, 2003, 46(46): 24–28.
- 4 Sun SX, Zhao J. A decomposition-based approach for service composition with global QoS guarantees. Information Sciences, 2012, 199(15): 138–153.
- 5 Zheng Z, Ma H, Lyu MR, et al. QoS-aware web service recommendation by collaborative filtering. IEEE Trans. on Services Computing, 2011, 4(2): 140–152.
- 6 Xiong P, Fan Y, Zhou MC. A petri net approach to analysis and composition of web services. IEEE Trans. on Systems Man & Cybernetics Part A Systems & Humans, 2010, 40(2): 376–387.
- 7 Lu Y, Wang W, Bhargava B, et al. Trust-based privacy preservation for peer-to-peer data sharing. IEEE Trans. on Systems Man & Cybernetics Part A Systems & Humans, 2006, 36(3): 498–502.
- 8 Skogsrud H, Benatallah B, Casati F. Model-driven trust negotiation for Web services. IEEE Internet Computing, 2003, 7(6): 45–52.
- 9 Breese JS, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. Fourteenth Conference on Uncertainty in Artificial Intelligence. 1998. 43–52.
- 10 Resnick P, Iacovou N, Suchak M, et al. GroupLens: An open architecture for collaborative filtering of netnews. Proc. of the 1994 ACM Conference on Computer Supported Cooperative Work. ACM. 1994. 175–186.
- 11 Zheng Z, Lyu MR. Collaborative reliability prediction of service-oriented systems. Proc. of the 32nd ACM/IEEE International Conference on Software Engineering. ACM. 2010, 1. 35–44.
- 12 Jiang Y, Liu J, Tang M, et al. An effective web service recommendation method based on personalized collaborative filtering. 2013 IEEE 20th International Conference on Web Services. IEEE. 2011. 211–218.
- 13 Wei L, Yin J, Deng S, et al. An extended matrix factorization approach for QoS prediction in service selection. 2012 IEEE Ninth International Conference on Services Computing (SCC). IEEE. 2012. 162–169.
- 14 Zhang Y, Zheng Z, Lyu MR. WSPred: A time-aware personalized QoS prediction framework for Web services. 2011 IEEE 22nd International Symposium on Software Reliability Engineering (ISSRE). IEEE. 2011. 210–219.
- 15 Chen X, Zheng Z, Yu Q, et al. Web service recommendation via exploiting location and QoS information. IEEE Trans. on Parallel & Distributed Systems, 2014, 25(7): 1913–1924.
- 16 Chen X, Zheng Z, Liu X, et al. Personalized QoS-aware web service recommendation and visualization. IEEE Trans. on Services Computing, 2013, 6(1): 35–47.
- 17 Liu J, Tang M, Zheng Z, et al. Location-aware and personalized collaborative filtering for web service recommendation. IEEE Trans. on Services Computing, 2015: 1–1.
- 18 Chen X, Liu X, Huang Z, et al. RegionKNN: A scalable hybrid collaborative filtering algorithm for personalized web service recommendation. 2010 IEEE International Conference on Web Services (ICWS). IEEE. 2010. 9–16.
- 19 Shao L, Zhang J, Wei Y, et al. Personalized QoS prediction for web services via collaborative filtering. IEEE Computer Society, 2007: 439–446.

- 20 Zheng Z, Ma H, Lyu MR, et al. WSRec: A collaborative filtering based web service recommender system. IEEE International Conference on Web Services, 2009. ICWS 2009. IEEE. 2009. 437–444.
- 21 Zheng Z, Ma H, Lyu MR, et al. Collaborative web service QoS prediction via neighborhood integrated matrix factorization. IEEE Trans. on Services Computing, 2013, 6(3): 289–299.
- 22 Chen X, Zheng Z, Yu Q, et al. Web service recommendation via exploiting location and QoS information. IEEE Trans. on Parallel & Distributed Systems, 2014, 25(7): 1913–1924.
- 23 Zhang Y, Zheng Z, Lyu MR. WSPred: A time-aware personalized QoS prediction framework for Web services. 2011 IEEE 22nd International Symposium on Software Reliability Engineering (ISSRE). IEEE. 2011. 210–219.
- 24 Rendle S. Factorization machines. 2010 IEEE 10th International Conference on Data Mining (ICDM). IEEE. 2010. 995–1000.
- 25 Koren Y. Factorization meets the neighborhood: A multifaceted collaborative filtering model. Proc. of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM. 2008. 426–434.
- 26 Ricci F, Rokach L, Shapira B, et al. Recommender systems handbook. Springer, 2011.
- 27 Guyon I, André E. An introduction to variable and feature selection. Journal of Machine Learning Research, 2002, 3: 1157–1182.