

一种时空敏感的 QoS 预测方法

熊 伟¹⁾ 李 兵²⁾ 吴 钊¹⁾ 杭 波¹⁾ 谷 琼¹⁾

¹⁾(湖北文理学院 湖北 襄阳 441053)

²⁾(武汉大学计算机学院 武汉 430072)

摘 要 Web 服务作为自解释的应用程序,通过在网络上提供标准接口以实现互操作性.在过去的几年中,网络上的开放服务变得越来越丰富,而服务的广泛应用导致需要高效的方法以提升性能,尤其是在软件即服务(SaaS)的场景中.目前的服务计算研究已经提出了许多基于 QoS 的方法,包括服务组合、服务选择、服务容错等,准确的 QoS 值预测显得尤为关键.服务器端的 QoS 值通常由服务提供商发布,代表 Web 服务的共享特性,而在客户端测量的 QoS 值则不同于服务器端:不同位置的用户可能测量到不同的 QoS 值.其缘由是服务调用需要消耗资源,并受到服务的状态与网络环境的影响,而且实时测量通常会引入新负载,从而导致性能评估不准确.近年来,一些方法建议引入更多的因素来提高预测精度(比如,考虑用户的感受与环境因素,引入空间及时间因素等),但是这些研究仅仅局限于某一维度的性能提升,缺少多维因素对于预测结果影响的考虑.文中提出一种时空敏感的 QoS 预测方法,该方法通过深度学习方法挖掘多维因素的高层特征以提高预测精度.文中在真实数据集上进行了大量实验,实验结果验证了该方法的有效性.最后,对该方法的未来发展进行展望.

关键词 Web 服务; QoS 预测; 时空敏感; 隐藏特征

中图法分类号 TP311 **DOI 号** 10.11897/SP.J.1016.2019.00772

A Spatial-Temporal-Aware QoS Prediction Approach

XIONG Wei¹⁾ LI Bing²⁾ WU Zhao¹⁾ HANG Bo¹⁾ GU Qiong¹⁾

¹⁾(Hubei University of Arts and Science, Xiangyang, Hubei 441053)

²⁾(School of Computing, Wuhan University, Wuhan 430072)

Abstract Web services are self-described programmable applications conducted to achieve interoperability and accessibility over a network, and are implemented in standard interfaces and published through specific protocols. Open services on the Web become increasingly abundant in the past several years. Meanwhile, the wide-spread use of Web services requires the effective approaches, especially in Software-as-a-Service(SaaS). A lot of QoS-based approaches have been proposed for Web service composition, Web service selection, fault-tolerant Web services, etc. Accurate QoS values of Web services are desired to work well for these approaches. The QoS values of Web services can be measured both at the server-side and at the client-side. QoS values measured at the server-side are published by the service provider and represent the shared feature of Web services, which are consistent to all the users (e. g., price, popularity, etc). However, QoS values measured at the client-side are different. Users in different locations may experience different QoS performance of Web services. Therefore, it is necessary to obtain accurate and personalized client-side Web service QoS values or their estimates. Conducting the real-world

收稿日期:2017-12-26;在线出版日期:2018-07-10. 本课题得到国家重点研发计划(2017YFB1400602, 2016YFB0800400)、国家自然科学基金基金(61572371, 61832014)、湖北省自然科学基金(面上项目)(2016CFB406)、湖北文理学院教师科研能力培育基金(2016zk004, 2016zk013)资助. 熊 伟,男,1973年生,博士,副教授,中国计算机学会(CCF)会员,研究方向为云计算、服务计算和软件工程. E-mail: xwei9093@126.com. 李 兵(通信作者),男,1969年生,博士,教授,中国计算机学会(CCF)高级会员,研究领域为云计算、服务计算和软件工程. E-mail: bingli@whu.edu.cn. 吴 钊,男,1973年生,博士,教授,研究领域为服务计算和软件工程. 杭 波,男,1978年生,博士,教授,研究领域为多媒体. 谷 琼,女,1972年生,博士,教授,研究领域为数据挖掘、机器学习、物联网等.

Web service evaluation at the client-side, however, is critical challenge. Web service invocations have costs. They may be charged in terms of the resources consumed in the cyberspace or the time elapse of invocations, where the server status (e. g. , workload, number of clients, etc.) and the network environment (e. g. , congestions, etc.) may change by time. Real-time performance testing may introduce overloads, which may impact the user experience of systems. Moreover, with introduced transaction workloads, the performance evaluating may not be accurate. It is difficult for various QoS-based approaches to perform well with the lack of accurate QoS values. Collaborative filtering methods are widely adopted in commercial recommender systems. QoS prediction using collaborative filtering technique was conducted firstly by Shao et al. Furthermore, advances in mobile Internet technology have enabled the clients of Web services to adjust to context changes regarding time, location and other factors. Since the services consumed by a mobile client may be different along with context changes, a multi-dimensional context model is necessary for discovering hidden relations from multi-dimensional context. Most contemporary QoS prediction methods exploit the QoS characteristics for one specific dimension, e. g. , time or location, and do not exploit multidimensional factors of context. This paper proposes a learning approach to Quality-of-Service (QoS) prediction of Web services via Spatial-Temporal context derived from the past invocation history. Our approach exploits an unsupervised encoder-decoder framework such as LSTM (Long Short Term Memory) to generate a hidden feature, which embeds the inherent characteristics for each context tracklet. Then we can calculate similarity between two context tracklets, which serves to predict Web service QoS values more accurate. To validate our approach, large-scale experiments are conducted based on a real-world Web service dataset, WSDream. The results show that our proposed approach achieves higher prediction accuracy than other approaches. Moreover, we plan to conduct more studies to predict client-side QoS properties by utilizing more other factors, and apply our approach under fog paradigm environments. The changes of application scenarios usually bring new issues.

Keywords Web service; QoS prediction; spatial-temporal-aware; hidden feature

1 引 言

Web 服务作为面向开发者的标准接口,正在变得越来越丰富,而众多基于服务的应用则需要解决如何有效地运用服务的问题。

服务的非功能属性 QoS 主要包括可用性、响应时间、可靠性以及吞吐量等等。在服务计算中的各个子领域(比如服务组合、服务选择、服务的容错以及服务的排序等),获取准确的 QoS 值对于这些方法是关键。QoS 值可以分别从服务器端和客户端获取,其中服务器端获取的 QoS 值由服务供应商发布,其对于所有的客户来讲是一致的(比如价格和流行性等)。但是,在客户端度量的 QoS 值则不同于服务器端(受客观环境影响,不同客户的体验是不同的)。服务的调用是有代价的(包括资源消耗和调用本身的时间消耗),获取 QoS 值还受到服务的状态

(比如负载及用户的数量)与网络环境(比如拥塞)等因素的影响,而实时的获取 QoS 值会导致大量的负载,反过来又影响用户的体验,实际上获取的用户-服务矩阵是稀疏的,这就产生了对于矩阵中缺失的元素进行预测的问题。

更进一步,Web 服务所在的信息空间(Cyberspace)涵盖互联网、嵌入到多种设备中的计算机、移动无线互联、连网与非连网设备的交互等。终端用户能够在“全球信息空间生态”环境下从事各种活动。这时信息空间计算设备不止是局部连网,而是全局动态、移动连网,且包含瞬时不连网设备。信息空间中不仅信息空间及子空间中的信息是紧密联系与互相影响的,而且信息空间中的资源、服务是动态、实时、实地变化的。相对于互联网/局域网等网络系统,信息空间不仅包括多种不同的连网信息的融合,还包括某时刻非连网信息的特征,信息空间在规模、多样性、复杂性方面超过了之前的任何一种网络系统,

信息空间具有异质(heterogeneity)、异构、多空间、多时域的特点,目前服务在信息空间中的信息融合面临的巨大挑战是如何对信息空间中服务的时间与空间性质进行建模,多尺度、多维度、多层次精确描述信息空间中服务的行为,以及分析服务时空变化轨迹特征。

当前面向单一时间要素和空间要素的模型已经建立较多,将多种信息要素编织与组合起来是一件很有意义的事情,时空建模是研究的难点和重点。

目前时空建模存在的重要问题,包括:

(1) 缺乏揭示时空因素对 QoS 预测影响内在规律的研究。

(2) 缺乏对空间要素与时间要素间复杂交互作用的分析。现有的时空建模是简单将时间要素和空间要素进行正交,并不是时空建模合适的方法。

(3) 目前的时空建模大多对拓扑关系进行拓扑,缺乏引入距离和方向等要素的时空建模研究。

为此一些方法建议引入更多的因素来提高预测精度(比如:更多地考虑用户的感受与环境因素,引入空间与时间因素等)。但是这些方法往往仅仅关注某一个维度对于预测结果的影响,缺乏考虑多维(比如:时间与空间)因素影响的预测算法。本文聚焦于空间和时间维度,为挖掘多维因素的更高层特征,本文引入深度学习(Deep Learning)方法。

深度学习是一种特征学习方法,其通过多层隐层将原始信息通映射为更高层次的抽象特征表达,进而学习表达非常复杂的关系。其中,高层次的特征表达能够削弱不相关因素,删去具有相关关系的变量中多余信息。深度学习的一项最大优势是,不同于传统人工定义特征的方式,其通过学习数据获取隐层的特征。深度学习的发展还归因于两点:(1) 深度学习模型在大数据场景下能达到最好的精度,也是最适合大数据的机器学习方法;(2) 计算集群、GPU 的应用产生了匹配深度学习的相应计算能力。

本文的主要贡献包括:

(1) 通过深度学习方法挖掘多维因素的高层特征,在提高精度的同时给出了一种通用的时空敏感的 QoS 预测方法(STA)。

(2) 通过在真实数据集上的大量实验,验证了本文方法的有效性。

本文第 2 节回顾与本文相关的研究工作;第 3 节给出本文的研究场景;第 4 节给出研究问题的形式化定义;第 5 节给出时空敏感的 QoS 预测方法的

解决方案;第 6 节进行实验设计并评估本文方法;第 7 节总结本文的研究工作并给出未来方向。

2 相关研究

2.1 时空结合建模

时空结合建模包括逻辑代数模型和应用驱动模型。

逻辑代数模型可以分为两类模型(包括逻辑模型与代数模型)。逻辑模型是一种直观描述时空要素相互作用的方法,其主要通过融合时态逻辑和空间逻辑,从语义和语法的角度来展开。Wolter 等人^[1]从语义的角度提出一个多维时空结构:区域作为实体,时间结构采取线性、分支和区间的 3 种基本组织形式,融合时态逻辑和空间逻辑方法是通过迪卡尔乘积以及模态算子提出。Wolter 等人^[2]从语义的角度提出了一阶时空逻辑(FOST),得出在无限时间流的约束下,其可满足问题是不可判定的。为研究可满足问题,设计了命题时空语言(PST)。Gabelaia 等人^[3]则在此基础上研究了时空模型表达复杂关系的能力。Bennett 等人^[4]通过融合命题时态逻辑和空间模态逻辑,构建了“二维”时空逻辑。虽然其可判定性目前还不知道,但目前得到了一些可判定的子系统。代数模型方面的研究主要包括:Medak 等人^[5]提出了一种能表达时空演化关系的代数方法,其基于范畴论给出了用于演变的一组基本操作。Viqueira 等人^[6]基于量子(quantum)空间模型构造了时空代数模型。Claramunt 等人^[7]基于 Ladkin 代数,通过引入空间副词与时态副词来描述时空的演变以及分析复杂空间的轮廓。Claramunt 等人^[8]基于区间代数与 RCC 拓扑关系,构建了包含 56 种基本时空关系的“二维”时空逻辑。

应用驱动模型:根据实际应用场景的需要,学术界发展了一些有影响的时空模型。Erwig 等人^[9]基于时空抽象数据类型(ADT)构建时空谓词来描述实体间的时空演进。Egenhofer 等人^[10]则重点关注区域间的联系。为研究在机器人足球比赛中机器人实体间的时空关系,Andrea 等人^[11]提出了基于实体移动时段和空间关系时段的模型。Rodriguez 等人^[12]在机动车场景下给出了相关时空模型。

国内也开展了有关时空推理的研究。清华大学研究了多维动态空间数据处理,浙江大学以及中国科学院遥感研究所在基于拓扑的时空建模^[13]、基于状态变迁的统一时空建模^[14]等方面进行了研究。

2.2 QoS 预测模型

基于 QoS 约束的 Web 服务方法应当满足:必须已知或者易于获取客户端度量的 Web 服务的 QoS 值. 由于现实中获取的 User-Item 矩阵是稀疏的,其意味着矩阵中大多数项是空的,这影响了服务选择的精确性. 因此, QoS 预测是一个很重要的问题. 协同过滤方法被考虑运用在 QoS 预测上面,第一个运用协同过滤预测 QoS 值的方法是由 Deora 等人提出的^[15].

目前已有两种方法被广泛研究:基于邻域与基于模型的方法. 基于邻域方法除了传统的基于用户与基于物品的方法外还包括融合方法^[16]. 基于模型的方法通常会通过训练数据集训练出一个模型用于预测,其包括了聚类、面向方面和矩阵分解方法. 近来一些方法建议引入了更多的情境因素来提高 QoS 预测精度. 比如,浙江大学罗伟等人采用了一种两级选择机制,识别当地社区并将其集成到模型 LoNMF 中^[17];熊伟等人将数据分布不均衡性引入到协同过滤模型中,提出了 UDD 方法^[18];中山大学郑子彬等人将社交网络引入到模型中,提出了 NIMF 方法^[19];合肥工业大学丁帅等人将客户满意度评估引入到模型中,提出了 CSTrust 方法^[20];Lee 等人将偏好传播(Preference Propagation)引入到模型中,提出了 LMF-PP 方法^[21].

迄今为止,研究时间与空间因素对 QoS 预测的影响是一个新兴的研究领域. 目前考虑空间因素的 QoS 预测研究包括:北京航空航天大学陈曦等人根据用户空间位置建立了一个层状分类结构,并基于该结构做预测^[22];湖南科技大学唐明董等人提出的 LACF 方法也是基于层状分类结构的^[23]. 浙江大学罗伟等人考虑在 LFM 代价函数后面增加与空间位置相关的约束^[24];而尹建伟等人更进一步发展了它,提出了 WL-PMF 方法^[25]. 目前考虑时间因素的 QoS 预测研究包括:Zhang 等人提出了将时间因素引入用户相似度计算,并使用了指数衰减函数对 QoS 值进行动态预测^[26];中国科学院软件研究所胡堰等人在 WSRec 基础上增加了时间因素^[27],继而又提出运用卡尔曼滤波(Kalman filtering)模型来描述复杂时间序列^[28]. 北京邮电大学王尚广等人提出了运用张量(Tensor)将时空因素结合,将两种因素统一起来进行 QoS 预测的方法 HDOP^[29].

2.3 深度学习模型

传统的神经网络属于浅层学习模型(Shallow Learning),和深度学习的主要差异在于其只包含少许隐层. 其代表性的支撑向量机(SVM)、Boosting

可以看成具有一层隐层的节点,而逻辑回归 LR 则没有隐层节点. 其局限性在于:(1)对抽象特征的表达能力有限;(2)依赖人工设计抽象特征,受限于人的领域知识和经验. 而深度神经网络(Deep Neural Network, DNN),则通过增加隐层的层数,可以学习到更复杂的抽象特征,具有优异的学习能力. DNN 的每层都对应一次特征变换,可以把数据逐步变换到更高层抽象的表示,相关参数的确定是通过自主地从数据中学习获取的. 传统神经网络在隐层数增多时,存在局部最小值和梯度衰减以及计算量大的问题. 为此,通过引入“逐层初始化”的过程,使得模型的初始状态趋近最优,从而具备了提高后续学习性能的可能.

深度学习的常见模型包括有:(1)卷积神经网络^[30](Convolutional Neural Networks, CNN)是一种带有卷积结构的深度神经网络,具有 4 个关键的特性:局部感知、权值共享、池化以及多层结构,主要应用在图像识别领域;(2)循环神经网络^[31](Recurrent Neural Networks, RNN)适合于处理变长序列数据. RNN 中引入自反馈的神经元,加入了时序因素;(3)长短时记忆神经网络^[32](Long Short-Term Memory Neural Network, LSTM)则给 RNN 增加了记忆能力. RNN 及 LSTM 的应用领域包括语音识别、自然语言处理,以及金融行情预测等. 此外,还有自动编码器^[33]以及深信度网络^[34]等模型在实际场景中有大量应用.

2.4 小 结

目前大多数相关工作都集中在时间或空间单一因素上,缺乏统一的时空模型. 王尚广等人^[29]虽然运用经典的张量分解方法将值函数通过广义策略迭代(Generalized Policy Iteration)方法对最优值函数进行逼近,从而得到最优策略. 然而,在大规模维度空间问题中,张量表形式的值函数会产生“维度灾难”问题. 本文通过引入深度学习进行值函数估计(Value Function Approximation)作为解决维度灾难问题的手段,其主要思想是将值函数进行参数化,将值函数空间转化为参数空间,达到泛化(Generalization)的目的.

3 研究场景

在客户端度量的 QoS 值是不同的,实际上客户对于服务的调用是有限的,因此获取的用户-服务矩阵是稀疏的,进而产生了对于矩阵中缺失的元素的进行预测的问题. 图 1 说明了三维 QoS 预测的场景

(包括用户(或客户端)、服务和时间三个维度的因素):假定有 m 用户与 n 个 Web 服务. u_i 和 s_j 分别表示第 i -th 用户和 j -th 服务. t_k 表示第 k 次的时间. 如果一个服务在特定的时间段被用户调用,所产生的 QoS 值(例如,响应时间)在这个时间点上被记录下来. 图 1 中只说明了一个时间因素. 在实践中,在 QoS 数据集中往往包含多种因素,每个因素都可被建模为一个单独的维度. 比如,当请求服务的位置需要时考虑时,可将位置因素也作为另一个维度进行建模. 因此,所有这些 QoS 数据都可以形成一个 5 维空间,如图 2 所示(每一个立方体是图 1 的三维模型). 本文主要关注时空因素对 QoS 预测的影响.

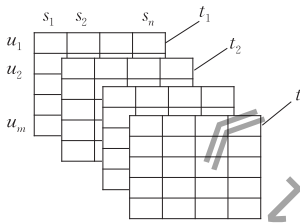


图 1 三维 QoS 预测的场景

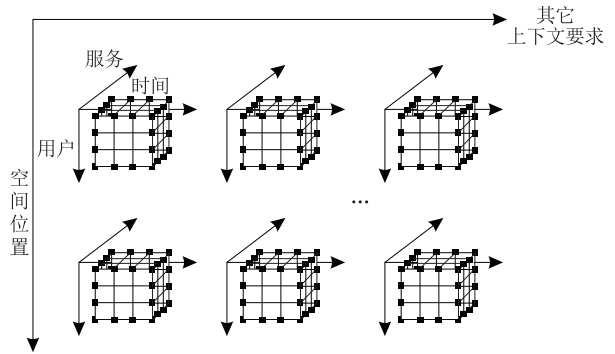


图 2 多维 QoS 预测的场景

4 问题定义

Web 服务的 QoS 预测是基于 QoS 的服务选择的基础,而 Web 服务的 QoS 值预测问题的本质是矩阵补缺问题(如图 3). 矩阵中 u_i 表示第 i 个用户, s_j 表示第 j 个服务. 矩阵元素 (x_{ijt}, QoS_{ijt}) 是一个时间序列,表明用户 u_i 在不同上下文 x_{ijt} (x_{ijt} 包括地理位置与时间信息)对于个服务 s_j 的调用获取的 QoS 的序列.

	s_1	s_2	s_3	s_4	s_5	s_6
u_1	x_{11t}, QoS_{11t}					
u_2			x_{23t}, QoS_{23t}		x_{25t}, QoS_{25t}	
u_3				x_{34t}, QoS_{34t}		
u_4		x_{42t}, QoS_{42t}				x_{46t}, QoS_{46t}
u_5	x_{51t}, QoS_{51t}					
u_6		x_{62t}, QoS_{62t}				

图 3 用户-服务调用矩阵

用户实际感知的 QoS 是通过监测而得到的,终端用户在不同地理位置不同时间对于同一个服务的调用其体验不同. 因此,Web 服务 QoS 的时空敏感性研究是一个重要问题.

问题的形式化描述如下:

假定用户-服务调用矩阵中元素表示为 (x_{ijt}, QoS_{ijt}) 时间序列. 其中 QoS_{ijt} 是用户 u_i 在上下文 x_{ijt} (包括调用时间与地理位置信息)下观察到的目标服务 s_j 的 QoS 值. 其中 $i \in \{1, \dots, m\}, j \in \{1, \dots, n\}$.

未知的 $\{QoS_{ijt} / (i, j, t) \in \Omega - \Delta\}$ 可以根据已知的 $\{(x_{ijt}, QoS_{ijt}) / (i, j, t) \in \Delta\}$ 来预测. Ω 是一个三元组集合 $\{(i, j, t)\}$, 其对应矩阵中所有元素的位置; Δ 是一个三元组集合 $\{(i, j, t)\}$, 其对应矩阵中所有已知元素的位置.

根据实验验证,我们发现 LSTM(Long Short-Term Memory Neural Network)模型是较好地选择:作为循环神经网络的一个变体,可以很好地刻画

多维数据特征并有效地解决简单循环神经网络的梯度爆炸或消失问题. 该模型能从多维情境纷繁复杂的细节中发现某些整体性的规律,帮助深刻认识情境各元素的属性,并能提炼出一些有价值的隐藏特征. 实验证明部分见第 6 节.

5 具体实现

5.1 位置信息的获取与处理

在已知用户 IP 地址的情况下获取其空间物理位置是比较容易的,现在有很多的在线服务或工具(比如在 Google maps 里包含的工具 IP Locator). 另外,由于 IP 地址分配的特殊性(根据前缀划分网络),IP 地址相近的两个用户在地理上可能并不相邻.

由此可以获取客户端的空间物理位置 (x, y) , 其中 x 是经度, y 是纬度,该信息包含在 x_{ijt} 之中.

5.2 利用 LSTM 模型学习隐变量

深度学习的本质是通过多层非线性变换,从数据中自动学习高层抽象特征,替代手工设计的特征.多层结构使其具有很好的表达能力,尤其体现在提取复杂特征与上下文信息方面.多个维度信息里各种隐含的因素往往以复杂的非线性方式关联在一起,而深度学习可以剔除掉相关因素中的多余信息以提高精度.

然而以 CNN 为代表的模型前馈神经网络受限于固定输入和输出的维数,不能适应变长序列的处理.而 RNN 固然适应变长序列的处理,且从结构上来讲更加符合生物自然特性,但是由于受限于梯度爆炸只能学习短周期的规律,缺乏学习长周期的能力.更进一步, LSTM 作为 RNN 的一个变体,通过加入时序因素以及增加记忆能力可以有效地学习长周期(如图 4). LSTM 模型的关键是在模型中加入了记忆单元,其可以有效地学习何时遗忘,何时更新.在时刻 t ,记忆单元 c_t 存储了时刻 t 前的历史信息,输入门 i_t 、遗忘门 f_t 和输出门 o_t 则具体实施了对记忆单元的控制.具体而言,该模型首先需要决定什么信息可以通过 Cell State.该决策由“Forget Gate”层通过 Sigmoid 层来控制,其根据前一时刻的输出 h_{t-1} 和当前输入 x_t 来产生一个在 $[0, 1]$ 之间的 f_t 值,来决定是否让前一时刻学到的信息 c_{t-1} 通过或

部分通过.

$$f_t = \sigma(\omega_{xf}x_t + \omega_{hf}h_{t-1} + \omega_{cf}c_{t-1} + b_f) \quad (1)$$

接下来我们需要更新的参数.该步包含两个步骤,首先是“Input Gate”层通过 Sigmoid 层来决定哪些参数需要更新,然后使用一个 tanh 层用来生成新的候选值 c_t ,它作为当前层产生的候选值可能会添加到 Cell State 中,最后将这两部分产生的值结合起来进行更新.

$$i_t = \sigma(\omega_{xi}x_t + \omega_{hi}h_{t-1} + \omega_{ci}c_{t-1} + b_i) \quad (2)$$

$$c_t = \tanh(\omega_{xc}x_t + \omega_{hc}h_{t-1} + b_c) \quad (3)$$

前面步骤结合起来就是丢掉不需要的信息,添加新信息的过程:

$$c_t = f_t \otimes c_{t-1} + i_t \otimes c_t \quad (4)$$

最后一步是决定模型的输出,首先是通过 Sigmoid 层来得到一个初始输出,然后使用 tanh 层将 c_t 值缩放到 $[-1, 1]$ 之间,再交由 Sigmoid 层处理,从而得到模型的输出

$$h_t = o_t \otimes \tanh(c_t) \quad (5)$$

$$o_t = \sigma(\omega_{xo}x_t + \omega_{ho}h_{t-1} + \omega_{co}c_{t-1} + b_o) \quad (6)$$

$$y_t = \text{Softmax}(\omega^* h_t) \quad (7)$$

其中, Sigmoid 函数的输出是不考虑先前时刻学到的信息的输出, Tanh 函数是对先前学到信息的压缩处理,起到稳定数值的作用,需要将两者结合起来.

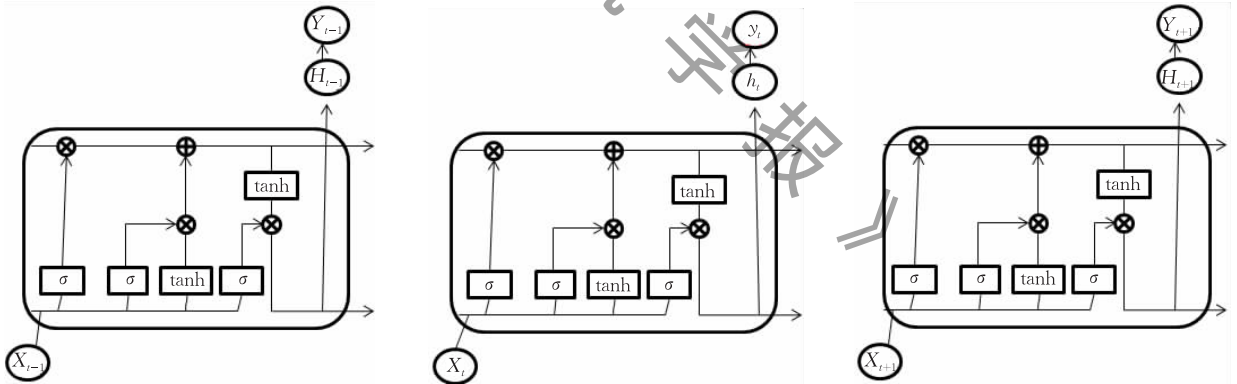


图 4 LSTM 模型

而模型的具体学习方法,则是后向传播误差学习权重的一个过程:期待使得损耗函数 L 达到最小, L 定义为

$$l_t = f(y_t, y_t^*) = \|y_t - y_t^*\| \quad (8)$$

$$L = \sum_1^T l_t \quad (9)$$

更进一步,需要计算参数迭代公式

$$\frac{dL}{d\omega} = \sum_{t=1}^T \sum_{j=1}^M \frac{dL}{dh_t(j)} \frac{dh_t(j)}{d\omega} \quad (10)$$

为方便计算,我们定义

$$L_t = \sum_{a=1}^T l_a \quad (11)$$

继而

$$\frac{dL}{dh_t(j)} = \sum_{a=1}^T \frac{dl_a}{dh_t(j)} = \sum_{a=t}^T \frac{dl_a}{dh_t(j)} = \frac{dL_t}{dh_t(j)} \quad (12)$$

将式(11)代入式(9),得到

$$\frac{dL}{d\omega} = \sum_{t=1}^T \sum_{j=1}^M \frac{dL_t}{dh_t(j)} \frac{dh_t(j)}{d\omega} \quad (13)$$

用已知的 $\{(x_{ijt}, QoS_{ijt}) / (i, j, t) \in \Lambda\}$ 时间序列来训练 LSTM 模型,其中 x_{ijt} 作为输入, QoS_{ijt} 作为输出.最终我们可以获得学习好的隐变量 $h_{ij}(u)$.

5.3 相似度计算

我们用学习好的隐变量 $h_{ij}(u)$ 计算相似度.本文计算两种相似度: $Sim(u, v)$ 表示用户间的相似

度,其实际上是对用户间距离正则化,公式如下:

$$Sim(u, v) = \frac{h_{ij}(u) \cdot h_{ij}(v)}{\|h_{ij}(u)\| \|h_{ij}(v)\|} \quad (14)$$

$Sim(i, j)$ 表示服务间的相似度,其实际上是对服务间距离正则化,公式如下:

$$Sim(i, j) = \frac{h_{ij}(i) \cdot h_{ij}(j)}{\|h_{ij}(i)\| \|h_{ij}(j)\|} \quad (15)$$

最终,我们将获取的用户间相似度与服务间的相似度应用到下面的方法中。

5.4 相似邻居的选取

为了提高预测精度,需要对相似邻居进行选取,我们用如下公式提取用户间与服务间的 top- k 个相似邻居:

$$S(u) = \{v | v \in T(u), sim(u, v) > 0, u \neq v\} \quad (16)$$

$$S(i) = \{j | j \in T(i), sim(i, j) > 0, i \neq j\} \quad (17)$$

其中, $S(u)$ 表示基于用户相似邻居的集合, $S(i)$ 表示基于服务相似邻居的集合, $T(u)$ 表示基于用户的 top- k 个代表性样本, $T(i)$ 表示通过基于服务的 top- k 个代表性样本。

5.5 QoS 值的预测

本文利用上文计算的 top- k 个相似邻居,分别基于经典的 UPCC(User-based collaborative filtering method using PCC)方法

$$\hat{r}_{u,t}^{current}(u, i) = \bar{r}(u) + \frac{\sum_{v \in S(u)} sim(u, v) (r_{i,t}^{current}(v, i) - \bar{r}(v))}{\sum_{v \in S(u)} sim(u, v)} \quad (18)$$

与 IPCC(Item-based collaborative filtering method using PCC)方法

$$\hat{r}_{i,t}^{current}(u, i) = \bar{r}(i) + \frac{\sum_{j \in S(i)} sim(i, j) (r_{i,t}^{current}(i, j) - \bar{r}(j))}{\sum_{j \in S(i)} sim(i, j)} \quad (19)$$

然后将两者综合起来,用参数 λ 进行平衡得到最终的预测公式:

$$\bar{r}_{i,t}^{current}(u, i) = \lambda \hat{r}_{u,t}^{current}(u, i) + (1 - \lambda) \hat{r}_{i,t}^{current}(u, i) \quad (20)$$

5.6 算法复杂度分析

这里我们讨论该预测方法(如算法 1)在最坏情况时的上界(假定用来训练的用户-服务矩阵矩阵是 m 行 n 列的)。

算法 1. 时空敏感的 QoS 预测方法。

输入: $\{(x_{ijt}, QoS_{ijt}) / (i, j, t) \in \Delta\}$

输出: $\{QoS_{ijt} / (i, j, t) \in \Omega - \Delta\}$

对数据做预处理,获取空间维度信息;

利用 LSTM 模型学习用户-服务调用矩阵中元素

(x_{ijt}, QoS_{ijt}) 序列的隐变量表示 $h_{ij}(u)$;

运用式(14)、(15)进行相似度计算;

运用式(16)、(17)进行相似邻居的选取;

运用式(18)、(19)、(20)进行 QoS 值的预测;

RETURN $\bar{r}_{i,t}^{current}(u, i)$

对于 LSTM 建模而言,假设一个时间序列长度为 n ,则需要进行 n 次操作,时间复杂度 $O(n)$ 。

由于用户 u 与 v 观察到的服务集合最多有 n 个,于是式(14)的复杂度是 $O(n)$ 。同样,服务 i 和 j 被调用的用户集合最多有 m 个,于是式(15)的复杂度是 $O(m)$ 。更进一步,由于 $T(u)$ 表示基于用户的 top- k 个代表性样本,而 $T(i)$ 表示基于服务的 top- k 个代表性样本,于是式(18)与(19)的复杂度分别是 $O(kn)$ 与 $O(km)$ 。最终混合两种方法的式(20)的复杂度是 $O(k(m+n+1))$,这使得其易于扩展,能够被应用在非常大的数据集上。

6 实验评估

本节设计实验将本文方法与当前的预测方法进行比较,实验将解决以下几个问题:

- (1) 与当前的方法相比本文方法表现如何?
- (2) 引入隐变量进行相似度计算的有效性如何?
- (3) 参数 λ 对于预测精度的影响如何?
- (4) 矩阵密度对于预测精度的影响如何?
- (5) 维度参数 k 对于预测精度的影响如何?
- (6) 与其它统一时空模型相比性能如何?

6.1 数据集描述

本文采用了来自真实世界的 Web 服务数据集 WSDream dataset。具体而言,我们主要使用了 WSDream-QoSDataSet3,该数据集包括了 142 个用户对于 4532 个 Web 服务 64 个时间间隔调用的记录,记录中包含有两种 QoS: 响应时间与吞吐率。另外,为获取客户端的 IP,我们同时使用了 WSDream-QoSDataSet2 中的 userlist。

由于数据集里存在响应时间不为 0 而吞吐量为 0 的不一致情况,直接删除异常值会造成样本量的不足,可能会改变原本数据的分布,造成分析结果的不准确,本文将其视为缺失值,用拉格朗日法进行了插补。

尽管在实验中我们仅使用了两种 QoS 属性,但是本文方法不需要做任何修改就可以很容易地运用到其它的 QoS 属性上面。

6.2 度量指标

本文采用平均绝对误差(MAE)与均方根误差(RMSE)作为预测精度的度量指标,其中平均绝对

误差定义为

$$MAE = \frac{\sum_{i,j} |R_{ij} - \hat{R}_{ij}|}{N} \quad (21)$$

其中, R_{ij} 是用户 i 观察到的服务 j 的 QoS 值, \hat{R}_{ij} 通过预测得到的 QoS 值, N 是预测值的总数. 而均方根误差定义为

$$RMSE = \sqrt{\frac{\sum_{i,j} |R_{ij} - \hat{R}_{ij}|^2}{N}} \quad (22)$$

其给的误差以相对较高的权重, 这意味着当较大的误差不期而至时, 其更有用.

6.3 比较方法

本文对比了以下 5 种不同 QoS 预测方法:

(1) 基于用户使用 PCC 的协同过滤方法 (UPCC)^[16]. 该方法首先使用 PCC 计算用户间相似度, 并将其作为权重作用于相对应的用户观察到 QoS 值, 最终产生的均值被用来预测 QoS 值.

(2) LACF (Location-Aware Collaborative Filtering for QoS-Based Service Recommendation)^[23]. 该方法基于层状分类结构来预测 QoS 值.

(3) 位置敏感的基于协同过滤的加权概率矩阵分解算法 WL-PMF (Weighted Location-Aware Probabilistic Matrix Factorization)^[25]. 该方法通过

在 PMF 方法中引入空间位置信息来预测 QoS 值.

(4) RTP (Real-Time Performance Prediction)^[26]. 该方法将时间因素引入预测方法, 设计了包含时间因素的衰减函数, 通过最小化衰减函数训练获得.

(5) TAW (Time-Aware WSRec with Hybrid Random Walk)^[27]. 该方法 WSRec 基础上增加了时间因素来预测 QoS 值, 并在计算相似度的时候考虑传递依赖.

6.4 实验结果分析

为使得实验贴近实际场景, 选择了数据密度分别为 10%, 20%, 30%, 将 k 定为 50, 参数 λ 设定为 0.4. 本文用 TensorFlow 工具来构建 LSTM 模型, 其参数设定为: 隐藏节点数 = 128, 隐藏层数 = 3, Softmax 采样样本数 = 256, 渐变限度 = 15, 学习率 = 0.01, 训练迭代数 = 50, 每段训练数据的大小 = 1024, 检查点的步骤数 = 2.

实验结果如表 1 所示, 从表中可知, 在与基准方法: 未考虑时间与空间因素的 UPCC, 考虑空间因素的 LACF 与 WL-PMF 方法, 考虑时间因素 RTP 与 TAW 方法相比较的过程中, 本文方法都获得了较高的精度. 这表明综合考虑时间与空间因素, 利用深度学习的方法分离出时间与空间因素间以复杂的非线性方式关联在一起的高层特征, 可以有效提高预测的精度.

表 1 不同方法的性能比较

QoS 属性	方法	数据密度 = 10%		数据密度 = 20%		数据密度 = 30%	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
响应时间	UPCC	0.9032	1.9428	0.8693	1.9341	0.8536	1.9272
	LACF	0.8039	1.9849	0.7471	1.8466	0.7033	1.7165
	WL-PMF	0.7933	1.8494	0.7671	1.8153	0.6928	1.7163
	RTP	0.6013	1.5487	0.5994	1.5238	0.4877	1.4856
	TAW	0.3825	0.9476	0.3076	0.7602	0.1237	0.3156
	STA	0.3776	0.9132	0.2538	0.5511	0.1188	0.2836
	吞吐量	UPCC	46.1432	120.9976	43.8712	113.5035	44.8455
LACF		45.1514	111.6165	40.6153	107.5957	38.1874	106.7037
WL-PMF		42.7335	109.8373	36.8648	104.7885	35.9163	104.1992
RTP		27.5612	76.2975	17.2365	52.2187	14.9870	47.8876
TAW		13.2578	42.7602	10.1276	36.5748	10.0037	35.4379
STA		13.1154	41.7344	10.0387	35.6948	10.0003	35.2279

另外, 在数据密度分别为 10%, 20%, 30% 的条件下, 我们观察到参与比较的算法精度一般情况下会随着数据密度的增大而获得提高, 这也符合直觉: 数据信息越丰富, 可利用的有效信息越多, 预测的精度也越高.

6.5 引入隐变量进行相似度计算的有效性

为比较不同相似度计算的差异, 由于皮尔逊相关系数 (PCC) 易于计算且有较高的精度, 本文选择 PCC 作为相似度计算的公式. 具体来讲, 本文采用

基于用户的协同过滤算法 UPCC 计算用户 u 与 v 间的相似度, 公式如下:

$$Sim(u, v) = \frac{\sum_{i \in I} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I} (r_{v,i} - \bar{r}_v)^2}} \quad (23)$$

其中, $I = I_u \cap I_v$ 是同时被用户 u 和 v 调用的服务的集合, $r_{u,i}$ 是用户 u 观察到的服务 i 的 QoS 属性, $r_{v,i}$ 是用户 v 观察到的服务 i 的 QoS 属性, \bar{r}_u 与 \bar{r}_v 分别地代表用户 u 与 v 观察到的服务的 QoS 的均值. 公

式 $Sim(u, v)$ 的值处于 $[-1, 1]$ 之间, 其值越大表明相似度越高. 特别地, 当用户 u 与 v 观察到的服务的交集为空时 ($I = \text{null}$), 此时 $Sim(u, v)$ 的值不能确定 ($Sim(u, v) = \text{null}$).

采用基于事物的协同过滤算法 IPCC 计算服务 i 与 j 间的相似度, 其公式如下:

$$Sim(i, j) = \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in U} (r_{u,i} - \bar{r}_i)^2} \sqrt{\sum_{u \in U} (r_{u,j} - \bar{r}_j)^2}} \quad (24)$$

其中, $U = U_i \cap U_j$ 是一个用户同时调用服务 i 和 j 的用户集合; $r_{u,i}$ 是用户 u 观察到的服务 i 的 QoS 属性;

$r_{u,i}$ 是用户 v 观察到的服务 i 的 QoS 属性; \bar{r}_i 与 \bar{r}_j 分别地代表服务 i 和 j 的不同用户观察到的服务的 QoS 的均值. 公式 $Sim(u, v)$ 的值同样处于 $[-1, 1]$ 之间, 其值越大表明相似度越高. 特别地, 当用一个用户同时调用服务 i 和 j 的用户的集合为空时 ($U = \text{null}$), 此时 $Sim(i, j)$ 的值不能确定 ($Sim(i, j) = \text{null}$).

实验结果如图 5, 从中可知, 在与基准方法: 采用 UPCC/IPCC 计算相似度方法相比较的过程中, 本文方法都获得了较高的精度. 这表明综合考虑时间与空间因素, 利用深度学习的方法学习高层特征, 可以有效提高预测的精度.

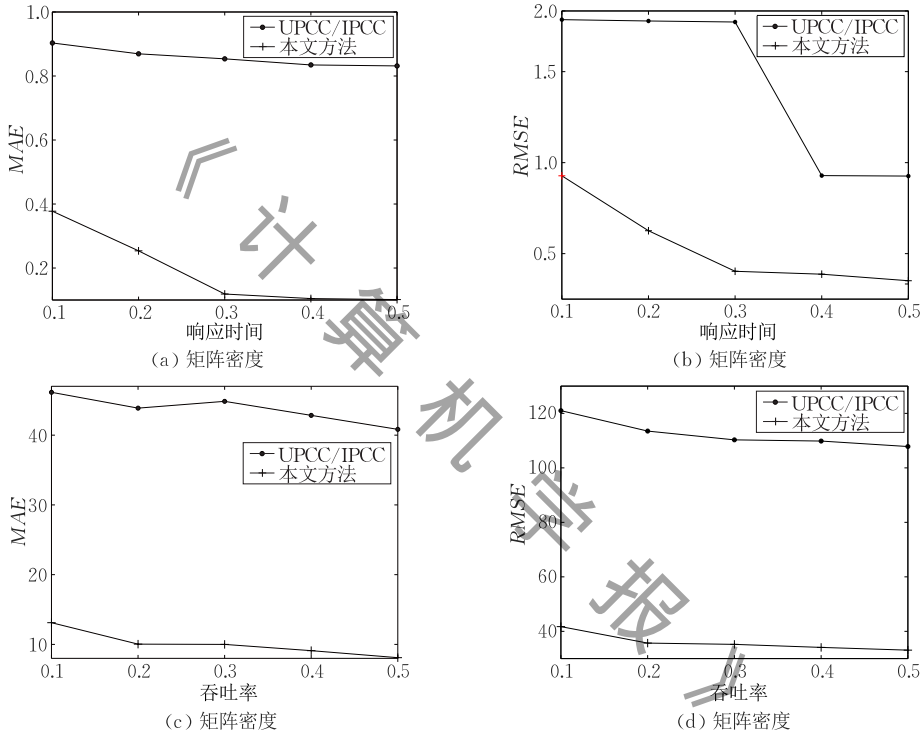


图 5 引入隐变量进行相似度计算的有效性

另外, 在数据密度从 10% 到 50% (每步为 10%) 的条件下, 我们观察到参与比较的算法精度一般情况下会随着数据密度的增大而获得提高.

6.6 参数 λ 对性能的影响

为获取参数 λ 对本文方法性能的影响, 我们需调整参数 λ . 本实验中, 我们逐步改变参数 λ 从 0.1 到 0.9, 每步为 0.1. 从图 6 可知, 在区间 $[0.1, 0.9]$ 存在着使得 MAE 与 RMSE 为最小的数据点, 在该位置预测的精度最好. 这表明适当的混合 UPCC 与 IPCC 方法可以得到更好的预测精度.

6.7 数据密度对性能的影响

为进一步明晰数据密度对本文方法性能的影响, 本实验中我们逐步改变数据密度从 10% 到 50%, 每步为 10%. 从图 7 可知, 随着数据密度的不断提高, 预测的精度越来越好.

6.8 参数 k 对性能的影响

为获取参数 k 对本文方法性能的影响, 我们需调整参数 k . 本实验中, 我们逐步改变参数 k 从 10 到 100, 每步为 10. 从图 8 可知, 在区间 $[10, 100]$ 存在着使得 MAE 与 RMSE 为最小的数据点, 在该位置预测的精度最好. 这表明在选择相似邻居时, 并不是 top- k 越大越好, 适当的选择 k 值可以得到更好的预测精度.

6.9 与其它统一时空模型的性能比较

王尚广等人运用经典的张量分解方法构建统一时空模型 HDOP^[29], 而张量表形式的值函数会产生“维度灾难”问题. 本文方法通过引入了深度学习进行值函数估计, 将值函数进行参数化, 将值函数空间转化为参数空间, 达到泛化 (Generalization) 的目的. 由实验结果 (图 9) 可知, 本文方法与 HDOP

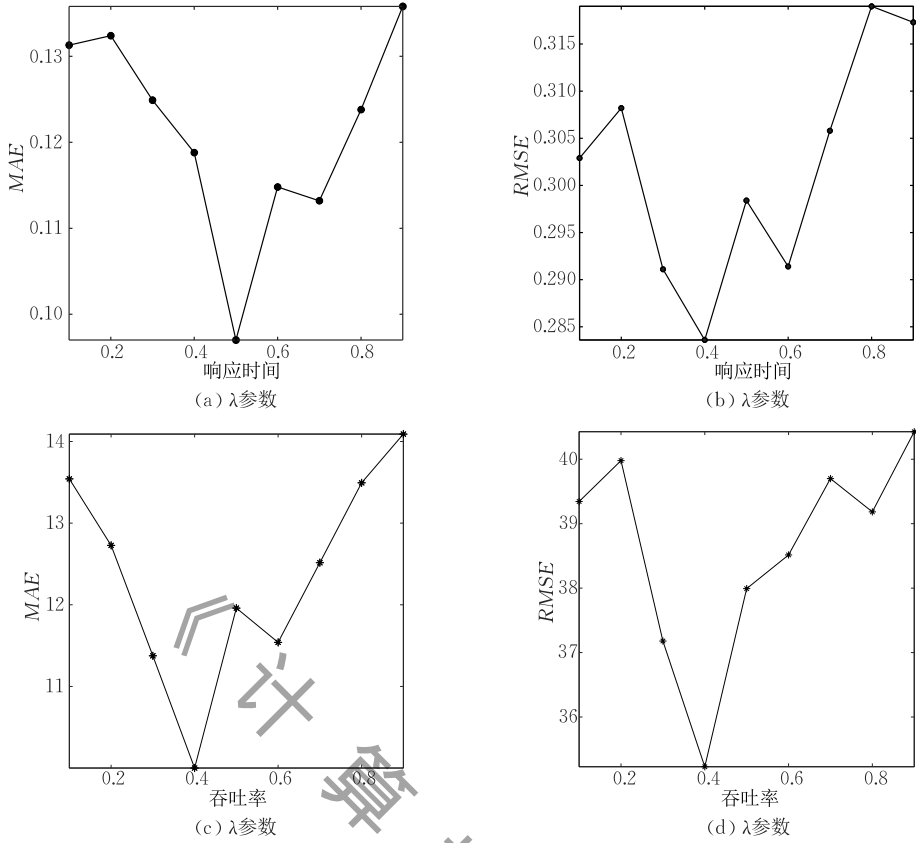


图 6 参数 λ 对性能的影响

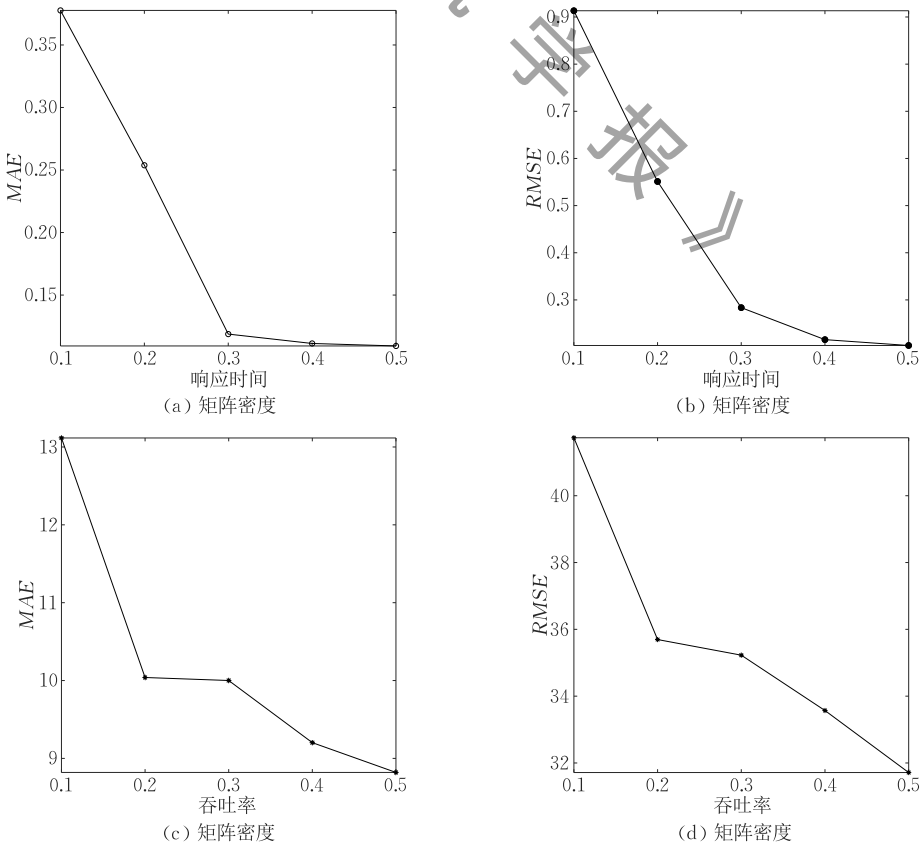


图 7 矩阵密度对性能的影响

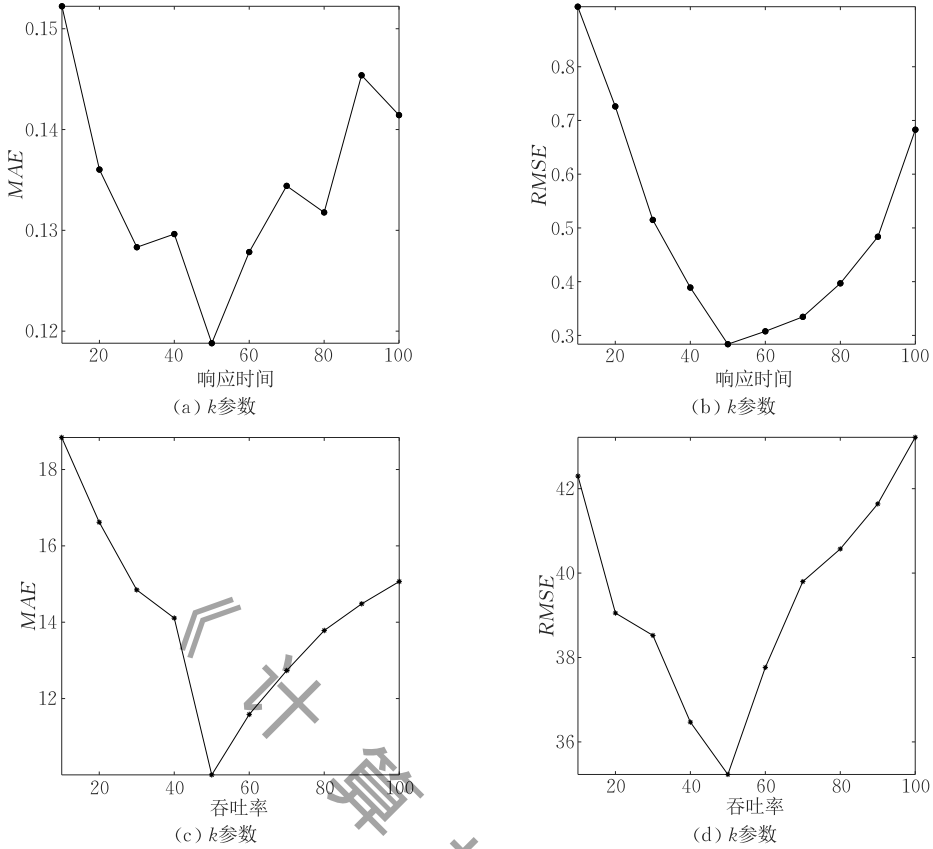


图 8 参数 k 对性能的影响

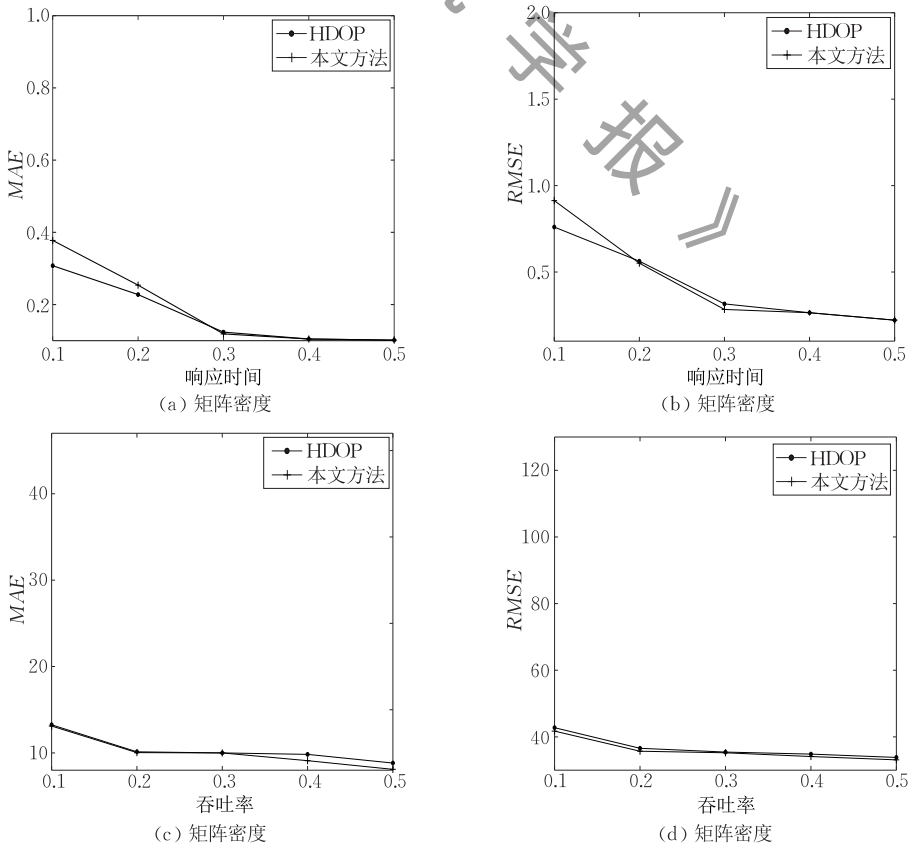


图 9 与其它统一时空数据模型的性能比较

相比,性能略好而且本文方法较大地减少了计算所消耗的时间,取得了良好的效果.

7 结 论

本文提出一种基于时空敏感的 QoS 预测方法. 该方法通过更多地考虑用户的感受与环境因素,引入空间与时间因素,并使用深度学习方法挖掘不同因素间的复杂关系,进而提高预测精度. 本文的实验结果验证了该方法的有效性.

未来我们计划在预测 QoS 值时引入更多因素,并探讨这些因素之间的复杂关系,并进一步将其应用到边缘计算环境中.

参 考 文 献

- [1] Wolter F, Zakharyashev M. Spatio-temporal representation and reasoning based on RCC-8//Proceedings of the 7th International Conference (KR2000). Breckenridge, USA, 2000: 3-14
- [2] Wolter F, Zakharyashev M. Qualitative spatio-temporal representation and reasoning: A computational perspective. *Missiology: An International Review*, 2012, 31(4): 140-140
- [3] Gabelaia D, Kontchakov R, Kurucz A, et al. Combining spatial and temporal logics: Expressiveness vs. complexity. *Journal of Artificial Intelligence Research*, 2005, 23(1): 167-243
- [4] Bennett B, Cohn A G, Wolter F, et al. Multi-dimensional modal logic as a framework for spatio-temporal reasoning. *Applied Intelligence*, 2002, 17(3): 239-251
- [5] Medak D I D. Lifestyles — A Paradigm for the Description of Spatiotemporal Databases [Ph. D. dissertation]. Technical University Vienna, Vienna, Austria, 1999
- [6] Viqueira J R R. Relational algebra for spatio-temporal data management//Proceedings of the 5th International Conference on Extending Database Technology (EDBT PhD 2008). Nantes, France, 2008: 1-4
- [7] Claramunt C. Extending Ladkin's algebra on non-convex intervals towards an algebra on union-of regions//Proceedings of the ACM International Symposium on Advances in Geographic Information Systems. Washington, USA, 2000: 9-14
- [8] Claramunt C, Jiang B. An integrated representation of spatial and temporal relationships between evolving regions. *Journal of Geographical Systems*, 2001, 3(4): 411-428
- [9] Erwig M, Schneider M. Spatio-temporal predicates. *IEEE Transactions on Knowledge & Data Engineering*, 2002, 14(4): 881-901
- [10] Egenhofer M J. The family of conceptual neighborhood graphs for region-region relations//Proceedings of the International Conference on Geographic Information Science. Zurich, Switzerland, 2010: 42-55
- [11] Bonarini A. The Body, the Mind or the Eye, First?. *RoboCup-99: Robot Soccer World Cup III*. London, UK, 2000, 1856: 39-60
- [12] Lamas-Rodriguez J, Arias J, Viqueira J R R, et al. OCURO: Estimation of Space Occupation and Vehicle Rotation in Controlled Parking Areas. Berlin, Germany: Springer, 2009
- [13] Zhang Xian-Feng, Cui Wei-Hong. Spatio-temporal analysis and modeling based on the integration of GIS and CA model. *Journal of Image & Graphics*, 2000, 5(12): 1012-1018(in Chinese)
(张显峰, 崔伟宏. 基于和模型的时空建模方法研究. *中国图象图形学报*, 2000, 5(12): 1012-1018)
- [14] Zheng Kou-Gen, Tan Shi-Yu, Pan Yun-He. A unified spatio-temporal data model based on state and change. *Journal of Software*, 2001, 12(9): 1360-1365(in Chinese)
(郑扣根, 谭石禹, 潘云鹤. 基于状态和变化的统一时空数据模型. *软件学报*, 2001, 12(9): 1360-1365)
- [15] Deora V, Shao J, Gray W A, et al. A quality of service management framework based on user expectations//Proceedings of the 2003 International Conference on Service-Oriented Computing(ICSOC). Trento, Italy, 2003: 104-114
- [16] Zheng Z, Ma H, Lyu M R, King I. QoS-aware Web service recommendation by collaborative filtering. *IEEE Transactions on Services Computing*, 2011, 4(2): 140-152
- [17] Lu 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
- [18] Xiong W, Li B, He L, et al. Collaborative Web service QoS prediction on unbalanced data distribution//Proceedings of the 2014 IEEE International Conference on Web Services (ICWS). Anchorage, USA, 2014: 377-384
- [19] Zheng Z, Ma H, Lyu M R, King I. Collaborative Web service QoS prediction via neighborhood integrated matrix factorization. *IEEE Transactions on Services Computing*, 2013, 6(3): 289-299
- [20] Ding S, Yang S, Zhang Y, et al. Combining QoS prediction and customer satisfaction estimation to solve cloud service trustworthiness evaluation problems. *Knowledge-Based Systems*, 2014, 56: 216-225
- [21] Lee K, Park J, Baik J. Location-based Web service QoS prediction via preference propagation for improving cold start problem//Proceedings of the 2015 IEEE International Conference on Web Services (ICWS). New York, USA, 2015: 177-184
- [22] Chen X, Liu X, Huang Z, Sun H. Regionknn: A scalable hybrid collaborative filtering algorithm for personalized Web service recommendation//Proceedings of the 2010 IEEE International Conference on Web Services (ICWS). Washington, USA, 2010: 9-16

- [23] Tang M, Jiang Y, Liu J, Liu X. Location-aware collaborative filtering for QoS-based service recommendation//Proceedings of the 2012 International Conference on Web Services (ICWS). Honolulu, USA, 2012; 202-209
- [24] Lo W, Yin J, Deng S, et al. Collaborative Web service QoS prediction with location-based regularization//Proceedings of the 2012 IEEE International Conference on Web Services (ICWS). Honolulu, USA, 2012; 464-471
- [25] Xu Y, Yin J, Lo W, Wu Z. Personalized location-aware QoS prediction for Web services using probabilistic matrix factorization //Proceedings of the 2013 International Conference on Web Information Systems Engineering (WISE 2013). Hangzhou, China, 2013; 229-242
- [26] Zhang Y, Zheng Z, Lyu M R. Real-time performance prediction for cloud components//Proceedings of the 2012 IEEE International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing Workshops (ISORCW). Shenzhen, China, 2012; 106-111
- [27] Hu Y, Peng Q, Hu X. A time-aware and data sparsity tolerant approach for Web service recommendation//Proceedings of the 2014 IEEE International Conference on Web Services (ICWS). Anchorage, USA, 2014; 33-40
- [28] Hu Y, Peng Q, Hu X, Yang R. Web service recommendation based on time series forecasting and collaborative filtering//Proceedings of the 2015 IEEE International Conference on Web Services (ICWS). New York, USA, 2015; 233-240
- [29] Wang S, Ma Y, Cheng B, et al. Multi-dimensional QoS prediction for service recommendations. IEEE Transactions on Services Computing, 2019, 12 (1): 47-57
- [30] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2012, 60(2): 2012
- [31] Graves A, Mohamed A R, Hinton G. Speech recognition with deep recurrent neural networks//Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, Canada, 2013; 6645-6649
- [32] Ma X, Tao Z, Wang Y, et al. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transportation Research Part C Emerging Technologies, 2015, 54; 187-197
- [33] Deng L, Seltzer M L, Yu D, et al. Binary coding of speech spectrograms using a deep auto-encoder//Proceedings of the Conference of the International Speech Communication Association (INTERSPEECH 2010). Chiba, Japan, 2010; 1692-1695
- [34] Mohamed A, Dahl G E, Hinton G. Acoustic modeling using deep belief networks. IEEE Transactions on Audio Speech & Language Processing, 2011, 20(1): 14-22



XIONG Wei, born in 1973, Ph. D., associate professor. His research interests include cloud computing, service computing and software engineering.

LI Bing, born in 1969, Ph. D., professor. His research interests include cloud computing, service computing and

software engineering.

WU Zhao, born in 1973, Ph. D., professor. His research interests include cloud computing, service computing.

HANG Bo, born in 1978, Ph. D., professor. His research interests focus on multimedia.

GU Qiong, born in 1972, Ph. D., professor. His research interests include data mining, machine learning and Internet of Things.

Background

Web services are self-described programmable applications conducted to achieve interoperability and accessibility over a network, and are implemented in standard interfaces and published through specific protocols. Open services on the Web become increasingly abundant in the past several years. Meanwhile, the wide-spread use of Web services requires the effective approaches, especially in Software-as-a-Service (SaaS).

Conducting the real-world Web service evaluation at the client-side, however, is critical challenge. Web service

invocations have costs. They may be charged in terms of the resources consumed in the cyberspace or the time elapse of invocations, where the server status (e.g., workload, number of clients, etc.) and the network environment (e.g., congestions, etc.) may change by time. Real-time performance testing may introduce overloads, which may impact the user experience of systems. Moreover, with introduced transaction workloads, the performance evaluating may not be accurate. It is difficult for various QoS-based approaches to perform well with the lack of accurate QoS values.

Furthermore, advances in mobile Internet technology have enabled the clients of Web services to adjust to context changes regarding time, location and other factors. Since the services consumed by a mobile client may be different along with context changes, a multi-dimensional context model is necessary for discovering hidden relations from multi-dimensional context. In order to enhance precision, historical multi-dimensional context data must be exploited as much as possible.

This paper proposes a learning approach to Quality-of-Service (QoS) prediction of Web services via Spatial-Temporal

context. The approach utilizes the history log of conducting real-world Web service evaluation at client-side to predict QoS values.

The work described in this study was fully supported by the National Key Research and Development Program of China (Nos. 2017YFB1400602 and 2016YFB0800401), the National Natural Science Foundation of China (No. 61572371 and No. 61702377), the Natural Science Foundation of Hubei Province in China (No. 2016CFB406), the Foundation of Teacher Research Ability Cultivation of Hubei University of Arts and Science (No.2016ZK004,2016ZK013).

《计算机学报》