丝路文化虚拟体验中的多视角立体重建技术研究

李兆歆¹⁾ 蒋 浩¹⁾ 刘衍青²⁾ 王兆其¹⁾

¹⁾(中国科学院计算技术研究所 北京 100190) ²⁾(宁夏师范学院 宁夏 固原 756000)

摘 要 丝路文化是联系一带一路战略的重要纽带,其传承意义重大,但是由于历史地理原因,丝路文化中代表性的历史遗产分散或损坏,难以有效地呈现,因此,本文面向丝路文化的虚拟展示与数字化,提出并实现了基于虚拟现实技术的丝路文化传承平台,通过历史遗迹复原以及基于图像的三维重建,还原了丝路文化中重要节点宁 夏固原有关的历史遗迹、文物和事件.特别地,本文提出一种面向高清图像的多视角立体三维重建算法,包括采用 normal-aware PatchMatch stereo 复原高质量的法线图,反映文物表面精细结构,以及提出一种基于 GPU 的增量 式的深度融合方法,以较小的显存处理大规模的数据.在公共数据集和本文收集的室内外文物数据上的实验表明, 本文提出的三维重建方法可恢复物体表面的精细结构,同时还对大规模数据具有良好的可扩展性.重建的模型可导入到虚拟互动系统中,对丝路文化的传播起到了积极的作用.

关键词 多视角立体;深度融合;丝绸之路;文化遗产;虚拟现实系统 中图法分类号 TP393 **DOI**号 10.11897/SP.J.1016.2022.00500

Research on Multi-View Stereo 3D Reconstruction in Virtual Reality System of Silk Road Cultural Inheritance

LI Zhao-Xin¹⁾ JIANG Hao¹⁾ LIU Yan-Qing²⁾ WANG Zhao-Qi¹⁾

¹⁾(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190) ²⁾(Ningxia Normal University, Ningxia, Guyuan 756000)

Abstract Silk Road culture is an important link in the Belt and Road strategy. Its heritage is of great significance. However, due to historical and geographical reasons, the representative historical heritage in the Silk Road culture is scattered or damaged, and it is difficult to present the historical heritage effectively. Therefore, in this work we propose and implement a virtual reality platform for the Silk Road Cultural Heritage. Through historical restoration and image-based 3D reconstruction, we effectively restored the historical sites, cultural relics and events of Guyuan of Ningxia Province in China, one of the important nodes in Silk Road Culture. For outdoor historical sites, we use a DJI Mavic Pro to capture 4K video clips of the giant Buddha of Xumi Mountain in a sunny day. For indoor cultural relics, we use a turntable with digital single lens reflex (DSLR) camera and multiple light sources to capture high-resolution images in 180 degrees. Based on these image data, we propose a simple and efficient multi-view stereo 3D reconstruction method for high-resolution images, which consists of a normal-aware PatchMatch stereo for the high-quality normal recovery to represent the detailed surface of the cultural relics, and a GPU-friendly incremental depth map fusion method which can fuse a large amount of depth maps by leveraging a small size of GPU memory. The high-resolution input images are essential for representing the geometric details

收稿日期:2020-10-14;在线发布日期:2021-07-28. 本课题得到国家自然科学基金(61702482和61532002)以及北京市自然基金(L172049)的资助. **李兆**歆,博士,助理研究员,中国计算机学会(CCF)会员(70190M),主要研究领域为三维重建和三维计算机视觉. E-mail: cszli@hotmail.com. **蒋**浩(通信作者),博士,副研究员,中国计算机学会(CCF)高级会员(185538),主要研究领域为群体仿真, E-mail: jianghao@ict.ac.cn. 刘衍青,博士,教授,主要研究领域为固原历史文化. **王兆其**,博士,研究员,主要研究领域为虚拟现实.

in historical sites and cultural relics. However, the state-of-the-art depth map fusion method needs to import all depth maps and normal maps into the GPU memory, and then globally fuse the depth points into the 3D point clouds for each reference image. Nevertheless, the space complexity almost linearly increases when the amount of data and image resolution increase. For instance, doubling image size will result in a fourfold increase in GPU memory. Due to limitation of GPU memory, this kind of global fusion strategy cannot address high-resolution input image data. The proposed incremental depth map fusion method in this paper

the amount of data and image resolution increase. For instance, doubling image size will result in a fourfold increase in GPU memory. Due to limitation of GPU memory, this kind of global fusion strategy cannot address high-resolution input image data. The proposed incremental depth map fusion method in this paper mainly consists of three steps: (1) We first set a reference view and a counter map for cross-view consistency check; (2) Then, we import α neighboring images of the reference view into GPU memory each time, and perform the cross-view consistency check for the depth points in reference view. And then, depth points are accumulated, and counter map is also updated. We then release the memory of these α images and import another α images into GPU and repeat the above operations; (3) When all neighboring views are processed, we can fuse the depth points whose values in counter map are larger than a threshold. The quantitative and qualitative experiment results on the public multi-view stereo benchmark as well as our captured datasets clearly highlight that the proposed method can recover the detailed surfaces while keeping a good scalability for the large-scale image data. The reconstructed high-quality 3D models of historical sites and cultural relics by our method can effectively support immersive virtual reality applications, playing a positive role in the dissemination of Silk Road culture.

Keywords multi-view stereo; depth map fusion; silk road; cultural heritage; VR system

1 引 言

丝绸之路在东西方文化交往中起到了十分重要 的作用,丝路上留下的文化遗产是丝路历史的实物 见证,具有极高的历史文化价值.宁夏固原作为古 丝绸之路的重要节点,留下了丰富的丝路文化遗产. 对这些历史文化遗产的保护和利用,对服务当地社 会文化的发展具有重要意义,一方面可以方便地方 历史文化的研究,另外一方面也可以作为文化传承 的纽带帮助学生、公众了解当地的历史文化,把更 多的人培养成为丝路历史文化的传播者和宣传者.

目前,针对历史文化遗产的传承与弘扬的传统 方式主要有两类:一种主要采用文字和多媒体为主 的传承与传播方式,这类方法比较灵活,却难以展 现完整动态的历史风貌,缺乏用户交互,在自主体 验方面有较大的局限性;另一种是利用文化古迹、 文物等有形实物的方式,虽然这种方式可以方便人 们直观地接触和体会历史文化风貌,但易受自然和 人为因素的影响而遭受损坏,而且对丝路文化体验 来说,古丝绸之路西北走廊多处于偏远山区,文物 遗址分散于丝路沿线,地处偏远、交通不便,在便 利性和全面性、以及传播发展和保存保护方面难以 平衡.

针对传统方法的不足,基于虚拟现实技术的历

史文化虚拟展示与数字化受到广泛关注,在国内外 被应用于不同种类的历史文化传承与传播,取得了 巨大的社会经济效益. 陈颖借助虚拟现实技术, 在 "天坛神乐署中和韶乐"应用实践中建立演练场所 的虚拟现实系统,提升了非物质文化遗产的动态展 示效果[1]. 针对大型的文化遗址,研究人员开展了相 应的数字化技术研究,潘云鹤等人提出了一整套数 字化壁画保护修复的技术,用于智能化壁画临摹辅 助和石窟壁画文物保护修复辅助^[2],刘箴利用虚拟 现实技术开发河姆渡遗址博物馆资源,制作了河姆 渡遗址博物馆的三维场景,并在微机上采用 VRML 语言实现了一个三维漫游系统^[3]. 王圣华等人针对 中国传统皮影的文物性保护与场景再现进行了研 究,建立了皮影可视化数据模型和渲染方法,实现 了逼真的中国皮影戏可视化表达[4]. 刘世光等人基 于流体力学和向量图表达技术,提出一种石纹纸染 艺术图案生成方法,用于这种古老文化艺术形式的 保护、记录与传承,可以绘制得到具有花型或特殊 特征等具有艺术效果的石纹纸染图案^[5].为了弘扬 中国剪纸艺术,张显全等人利用纹样组合,利用计 算机辅助生成剪纸形象,可得到具有民族风格的剪 纸图案^[6],涂传朋等人则从剪纸作品里面的流水动 画效果着手,构建流水波纹模型以及它们的动态控 制方法,使得手工剪纸中的流水通过用户少量交互

就可以生成动画,用该方法生成的剪纸风格流水动 画在视觉效果上自然、流畅,对推广剪纸艺术很有 价值^[7]. 在历史文物的虚拟展示方面, 刘晓等人对表 面腐蚀或表面残缺不全的文物表面图案进行建模^[8], 用计算机修复虚拟文物,这样可以重现古文物的原 来面貌,这对文物资源的立体展示以及文物资源考 古研究等将起到重要的推动作用,但这种方法构建 的文物三维模型只能重现原始文物,无法改变文物 器型展示效果,而胡晏秋等人面向青铜器建模与绘 制,通过区域几何特征对三维网格模型进行分割和 参数化处理,实现青铜器纹的凹凸绘制以及表面锈 蚀效果绘制,在三维展示、艺术设计等领域有较好 的应用前景^[9].此外,增强现实技术也被广泛应用于 文化遗产的数字化保护与传承^[10],研究人员针对不 同类型文化遗产的特点,开展了有针对性的应用研 究(例如,数字圆明园增强现实系统^[11]).

文物的数字化需要借助三维重建技术.相比基 于激光扫描的方式^[12],基于多视角立体(Multi-view stereo, MVS)的方式具有低成本和对室内外场景的 广泛适应性的优势. 特别地, 由于捕获高清图像数 据越来越便利,多视角重建的应用越来越广泛.随 着多个评估数据库的提出[13-15],多视角重建方法的 精度和完整度不断提升. 当前主流方法主要是基于 深度图的 GPU 友好的 MVS 方法^[16-19],这类方法首 先重建每个视角的深度图,然后将不同视角的深度 图导入到 GPU, 通过跨视角一致性检验和一致性深 度的平均,进行深度图融合,形成统一的三维点云[17]. 然而,由于噪声和视差范围的影响,重建的深度图 往往不能很好地恢复物体表面的精细结构. 通过增 加正则化项^[20,21]虽然可以减弱噪声带来的影响,但 会增加计算成本,并破坏物体表面的细节.相比低 分辨率图像, 高分辨图像数据蕴含了物体表面可辨 识的细节信息,可增加视角间的匹配精度,但是由 于重建的高分辨率图像的深度图包含大量的数据, 融合深度图将需要较大的显存资源,限制了方法在 高清图像数据集的应用.

本文主要面向丝路重镇宁夏固原的代表性历史 遗迹和文物等丝路文化遗产(如图1),基于虚拟现 实技术,综合建模技术和交互技术,通过高清图像 数据采集与多视角三维重建,将丝绸之路中具代表 性的文物古迹进行数字建模与虚拟化,提供一个高 真实感、沉浸感和多种交互方式的虚拟现实互动环 境,通过互动体验的方式,在宁夏、西部乃至世界 范围内让更多的人能够更直观、全面地了解固原丝 路文化. 本文的主要贡献包括:

(1)提出一种针对古迹和文物高清图像数据集的简单高效的三维重建算法,包括一种具有较低显存占用率的增量式深度融合算法,一方面显著提升了古迹和文物三维几何细节复原精度,另外一方面对图像分辨率和图像数据规模具有较好的可扩展性.

(2)通过多视角三维重建技术将丝路文化历史 遗产进行数字化与虚拟化,提供一个高真实感的虚 拟现实互动体验环境,提出的整个系统框架,特别 是三维重建方法,可潜在应用于其它面向文化保护 和传播的虚拟现实系统中.



图 1 固原境内的丝路历史遗迹和文物:(a)须弥山大佛 造像;(b)固原市博物馆馆藏文物.

2 相关工作

本文面向室外大型古迹和室内馆藏文物提出一 种有效的多视角立体重建算法.多视角立体重建旨 在从场景的一组图像集中复原场景三维结构,是图 像生成的逆过程.基于三维表面的表达形式,多视角 立体重建方法可分为基于体素^[20,21]、三角面片^[22,23] 和深度图的方法^[16-19],近年来基于深度学习的端到 端重建方法也逐渐获得广泛关注^[24,25].关于多视角 方法的详细综述可参见^[26],本节主要讨论和本文工 作密切相关的方法,即面向大尺度数据的基于深度 图的多视角立体重建方法.

基于深度图的重建方法尝试估计输入图像的深 度信息,取决于深度点的稠密度,可分为面向稠密 深度图的重建方法和面向半稠密深度图的重建方 法.半稠密深度图的重建方法主要通过提取图像中 的显著特征点进行多视图间的匹配,也称为基于特 征点的方法,典型工作如 PMVS^[27].这类方法对强 纹理表面非常有效,然而容易在弱纹理区域处产生 不完整的重建结果.稠密深度图的重建方法尝试估 计每一参考图像的稠密深度信息,其主要步骤包括了 深度图估计和深度图融合两个子问题.由于这类方 法可以重建相对完整的场景表面并可借助 GPU 加速 提升重建效率,是当前多视角重建的主流方法^[16-19]. Galliani 等^[17]首次将 Black-Red 网格传播方法应用于 PatchMatch Stereo, 方法通过 GPU 加速交替传播黑 格和红格像素区域的深度和法线,显著提升了重建 每一视角深度图的速度. Schönberger 等在文献[17] 的基础上将遮挡关系作为约束项进行联合优化^[16], 获得了更高质量的重建结果. 文献[28]提出一种非 对称的网格传播方法,同时采用由粗到细(coarseto-fine)的策略提升方法在弱纹理区域的表现. 基于 深度学习的端到端的深度估计方法尝试使用神经网 络刻画从图像到深度图的过程^[24, 25],对弱纹理较为 鲁棒. 文献[24]提出一种面向稠密图像采样的小场 景深度图预测网络,其网络需要固定输入的视点数 量,在处理更大场景时无法有效的处理图像间可见 性信息. 文献[25]提出一种面向大场景非结构化数 据的网络架构,但是由于网络结构需要较大的显存, 只能应用于低分辨率的深度图恢复,且无法有效重 建物体表面的精细结构,限制了方法在高清图像数 据和大尺度场景下的应用. 而在深度图融合问题方 面,为了高效的融合大尺度数据,基于 GPU 的逐像 素滤波和融合[17]已经在各类主流方法上获得广泛应 用^[19, 24, 28]. 然而此类方法需要将目标场景的深度 图、法线和原始图像数据全部导入显存,随着图像 规模和分辨率的增大,其对显存的需求量迅速增加,

限制了方法在大场景和高清图像数据集上的应用.

针对古迹和文物往往包含精细三维结构这一特 点,本文提出一种针对高清图像数据集的简单高效 的三维重建算法,一方面显著提升了古迹和文物三 维细节复原精度,另外一方面对图像分辨率和数据 规模具有较好的可扩展性.

3 总体设计

虚拟现实技术综合运用图形学、计算机视觉、 多媒体等技术,将现实场景在计算机上真实呈现, 日益成为一个研究热点.虚拟现实数字化互动体验 系统将传统的丝路文化遗产展示从二维转向三维, 能更逼真、更准确地呈现丝路文化遗产资源,使用 户可以利用虚拟现实设备实现丝路文化遗产数字资 源的 3D 浏览,有如身临其境的感觉.突破了时空的 限制,让人们从能"走进"这一场景,真正做到方 寸之间就能直接通过计算机全方位和清晰地体验丝 路文化遗产数字资源.

丝路文化虚拟现实互动体验系统由 2 个底层基 础模块构成,分别为数字资源采集,多视角立体重建 与模型植入模块.在此基础上建立数字资源库,协助 虚拟现实互动体验平台的建立,实现人机交互以及场 景的真实感展示.提出的系统平台示意图如图 2 所示.



图 2 系统框架图

系统使用的硬件资源主要分为数据采集、三维 重建和虚拟展示三部分.为了在虚拟现实体验系统 中能同步浏览丝路文化代表性历史遗产,需要采集 历史遗迹或文物的三维数字资源,并根据采集对象 的特点选用不同的采集方式.在本文中,主要考虑 较低部署成本的、以图像采集为主的方案,包括无 人机和自动控制的环绕采集系统.无人机比较灵 活,适合采集较大范围的图像数据,比如人们难以 获得完整数据的山体、大型像等.环绕采集系统可 以配置好之后由软件自动控制,适合近距离采集物 品清晰的图像序列,主要针对文物等体积较小的需 要精细展示的代表性物品.此外,采集并进行标定 后的数据,通过本文提出的三维重建流程生成三维 数字资源,再通过 VR 设备进行展示与体验,为获 得更好的用户体验,本文主要使用 VR 头盔与 VR 自由行动平台相结合的方式,用户在使用头盔进行 观察的同时,还可以在三维场景内自由行走,并通 过手柄与系统进行互动,并辅助与背景资料、语音 解说等相关说明材料进行关联.本文中的虚拟现实 互动体验系统选择 Unity3D 作为软件开发平台.通 过融合三维重建的虚拟场景,在系统中实时渲染沉 浸画面和实时反馈用户交互信息,可实现高质量的 虚拟体验.用户通过佩戴 VR 头盔,处在一个完全 被包裹的环境中,大大提高用户的体验感和沉浸感. 作为 VR 系统的核心模块,本文系统中的三维 重建模块可以处理高清图像的输入,并生成高精度 的三维数字模型.接下来将对本文采用的三维重建 方法的具体流程进行详细描述.

4 文化遗产多视角立体三维建模

文化遗产的三维数字化可用于文化的宣传与传承,三维展示使得大众通过交互应用或网络就能获得对文化瑰宝的身临其境的感受.文化遗产不仅包含了小尺度的文物藏品还包含室外大型建筑古迹.虽然基于结构光扫描的方式可以实现小尺度文物的精细重建,但此类方法成本较高,且易受光线干扰,不适用于大尺度的古迹重建.本文提出一种基于高清图像为输入的多视角立体(Multi-view stereo, MVS)重建方法,实现文物古迹的低成本高质量重建.数字化的三维模型可导入虚拟现实场景,使用

户获得身临其境的体验效果.给定目标场景下的多视角图像数据,多视角立体方法尝试寻找图像间的 对应点并推断场景的三维形状^[13-15].与时间飞行法 和结构光法相比,多视角立体方法具有更好的可扩 展性,能够应用于大尺度场景的重建任务,并且只 需使用相机对场景进行拍照,具有较低的应用成本. 本文针对固原历史文化遗产中的两类代表性文物进 行三维数字化:须弥山景区石窟遗址和固原馆藏文 物(包括固原市博物馆、彭阳博物馆和农耕博物馆).

为了真实地数字化文物数据,本文捕获高分辨 率的图像数据以真实反映物体表面的精细结构.本 文分别设计了基于 normal-aware PatchMatch stereo 和 增量点云融合的方法来处理高分辨率图像数据.方 法示意图如图 3 所示.下面将按照数据获取、相机 标定和多视角立体重建的步骤,对本文方法进行详 细介绍.



图 3 三维重建方案示意图

4.1 数据获取

多视角立体重建方法的输入是被测物的多角度 RGB 图像数据. 针对须弥山石窟和固原馆藏文物的 不同特点,本文采用不同的数据采集方案.

须弥山石窟是固原的代表性文化遗址,由于此 遗址为大尺度场景,本文使用无人机航拍进行数据 的采集,如图 4 (a).针对日照强度和方向以及风 力对拍摄的影响,航拍时间选择晴朗无风的上午. 本文使用 Mavic Pro 无人机,并在无人机移动过程 中拍摄 4K 分辨率视频(分辨率 3840×2160),无人 机每次飞行的有效航拍时间约为 20 分钟,为了尽可 能让拍摄的图片覆盖大佛的不同区域,本文进行多 次航拍.对拍摄的视频数据进行关键帧抽取,提取 一组相互交叠且不同角度的视频帧,并手动剔除那 些模糊的视频帧.

固原各博物馆收藏了大量与丝绸之路文化密切

相关的珍贵文物.本文面向文物的拍摄系统包括可 自动控制转动步长的电动转台和高清相机,以及若 干漫反射光源. 如图 4(b) 所示, 被拍摄的物体放 置于转盘中心,当转盘匀速缓慢转动时,自动触发 相机进行等时间间隔的拍摄,共拍摄以物体为中心 的不同角度的 60 张高清图像(平均每6度拍摄一张 图像,图像分辨率: 5184×3456),为了清晰捕获文 物表面的精细纹理避免散焦导致的模糊,相机在拍 摄过程中采用自动变焦模式. 值得指出的是, 在上 述两类不同特点场景的数据获取过程中,本文均是 捕获高清图像. 通过捕获高清图像可以让物体表面 的纹理特征更具辨识度,一方面减少由于弱纹理和 重复纹理导致的匹配二义性, 增加了后续相机标定 和三维重建的精度,另一方面也可以提升物体表面 的小尺度几何结构的复原效果, 而这些小尺度的精 细几何结构往往是文物高超艺术性的体现.



图 4 数据获取: (a) 室外无人机航拍; (b) 室内自动拍摄

4.2 相机标定

给定一组输入 RGB 图像 { $I_1...I_i...I_n$ },下一步 将是对每一图像所对应的相机参数进行求解.考虑 到对场景的适应性,本文采用 structure from motion (SFM)方法求解相机的内外参数.对于一幅观察图 像 I_i ,相机*i*的参数包括了相机的内部参数和相机的 位姿.当假设场景符合针孔模型(pinhole model)时, 其内参数矩阵可表达为 3×3 矩阵 (公式 (1)):

$$\boldsymbol{K} = \begin{bmatrix} f_x & 0 & c_x \\ & f_y & c_y \\ & & 1 \end{bmatrix}$$
(1)

其中 f_x 与 f_y 为相机焦距在像平面X方向与Y方向的 度量, c_x 与 c_y 为相机主点位置,即相机光心在像平 面上的投影位置,单位均为像素.相机的外参数包 含了相机相对于参考世界坐标系的旋转和平移,分 别用 3×3 旋转矩阵R和 3×1 平移向量t表示.

使用 SFM 实现相机标定的步骤主要包含 3 步: 特征点提取,多视角模型估计和捆绑调整(bundle adjustment).SFM 效果很大程度依赖特征点的提取 和匹配精度,通过尺度不变特征点检测与描述算法 SIFT,可以在每一高清图像上提取一些辨识度较高 的特征点,并建立视角间的特征点对应.由于本文 采用了高分辨率的图像作为输入,保证了特征点提 取的精度,减少了噪声的影响.多视角模型基于特 征点对应估计局部相机组的内外参数,同时剔除特 征点对应中的误匹配(outliers).经由上述两步, 本文获得特征点对应、初始的相机内外参数和一组 稀疏的三维点集 *A*.为了进一步最小化计算误差, 采用捆绑调整对相机参数进行进一步优化.

假设在世界坐标系三维点 $X_k \in \Lambda$ 在其可见相 机 集 合 $i \in N(X_k)$ 的 二 维 重 投 影 p'_i 为 $\phi_z(K_i$ $(R_iX_k + t_i))$,其中 $\phi_z(\mathbf{a})$ 为深度归一化操作,即 \mathbf{a}/z . 图像 I_i 中与 X 对应的特征点位置为 p_i ,捆绑调整可 表述为(公式(2)):

$$\varepsilon(\boldsymbol{X}, \boldsymbol{K}, \boldsymbol{R}, \boldsymbol{t}) = \sum_{\boldsymbol{X}_k \in \Lambda} \sum_{i \in N(\boldsymbol{X}_k)} \rho(\boldsymbol{p}_i - \phi_z(\boldsymbol{K}_i(\boldsymbol{R}_i \boldsymbol{X}_k + \boldsymbol{t}_i))) \ (2)$$

其中 ρ 为鲁棒惩罚函数用于度量重投影误差. 公式(2)可通过迭代优化方法 Levenberg-Marquardt 方法求解.

SFM 的重建具有尺度二义性,即无法确定场景的真实物理尺寸.对于基于转盘的拍摄系统,本文在拍摄时在文物周围放置若干标定板(图5(a)和图5(b)分别为两个被测物体和放置的标定板),从而增加了 SFM 的精度和成功率,同时由于可以测量标志物的真实物理尺寸(毫米),从而可以获得 SFM 场景与真实物理场景之间的尺度比例因子*s*,帮助实现真实物理尺寸的三维数字化.

在研究当中,本文综合比较了几种主流 SFM 方法^[29-31],发现文献[31]在多数情况下能够获得最好的效果.文献[31]采用一种局部增量捆绑调整策略,逐步减少匹配的误差,增强了算法的鲁棒性,因此在实验中本文基于文献[31]完成相机内外参数的标定.



图 5 围绕物体放置的标定物

4.3 深度估计

给定一组多角度的观察图像 $\{I_1...I_i...I_n\}$ 和对 应的相机投影矩阵 $\{P_1 \dots P_i \dots P_n\}$, 多视角立体重建 (Multi-view stereo)的目标是复原被测物体的三维 表面 $S \subset \mathbb{R}^3$,可以看作是图像生成的逆问题. 与体 素方法^[20, 21]相比,点云方法数据冗余较少,对内存 的需求低,而与基于三角网格的方法^[22, 23]相比,点 云方法能够灵活的处理拓扑结构变化. 基于点云的 多视角重建方法可分为基于深度图的方法(depth map-based MVS)^[16-19]和基于特征点的方法(featurebased MVS)^[27]. 基于特征点的方法依赖场景包含丰 富的纹理,在缺乏纹理的区域容易产生重建缺失, 而基于深度图的方法在弱纹理表面有更好的效果, 能够重建更完整稠密的表面,并且可以恢复物体表 面的高频细节,增加物体表面的几何细节辨识度. 文化遗址和文物的重建需要能够清晰复原几何细 节,因为这些细节是这些杰出艺术品的精湛艺术性 的体现,因此本文采用基于深度图的方法,以便重 建模型完整且包含精细几何结构.为了有效处理大 尺度图像数据集,本文基于改进的 Gipuma 算法^[17] 进行三维点云的重建. 算法^[17]包含两个基本步骤: 基于 GPU 加速的 PatchMatch stereo 深度图生成, 以 及基于 GPU 的深度图的融合.

4.3.1 基于 Normal-aware PatchMatch 的深度图估计

在 多 视 角 深 度 图 估 计 (Multi-view depth estimation) 阶段,目标是估计各个输入图像的对应 深度图.在文献[17]中,除了可以估计深度图,还可 以估计法线图,实际是为每一个像素估计一个 3D 平面.具体地,以任意图像 I_i 为参考图像, $I_j \in V(i)$ 为参考图像 I_i 的邻域图像集中的任意图像.对于像 素 $p \in I_i$,设它当前的深度假设值为 d_p 与法线假设 值为 n_p ,经由深度 d_p 和法线 n_p 所构成的平面 П 和 相机参数 P_i, P_j ,可计算 p在邻域图像 I_i 的对应点 q_j .通过参考图像中以p为中心的 $r \times r$ 的图像块 R(p)与 邻域图像 $I_j \in V(i)$ 的对应匹配图像块 $R(q_j)$ 计算匹 配代价选择最优的深度和法线假设.匹配相似性函 数定义为: $\rho_j = \rho(R(p), R(q_j))$.方法[17]采用了 intensity+gradient 的成本函数(公式(3)和公式(4)):

$$m_{j}(\boldsymbol{s},\boldsymbol{t}) = (1-\alpha) \cdot \min\left(\left\|\boldsymbol{I}_{i}(\boldsymbol{s}) - \boldsymbol{I}_{j}(\boldsymbol{t})\right\|, \tau_{c}\right) + \alpha \cdot \min\left(\left\|\nabla\boldsymbol{I}_{i}(\boldsymbol{s}) - \nabla\boldsymbol{I}_{j}(\boldsymbol{t})\right\|, \tau_{g}\right)$$
(3)

$$\rho_j = \sum_{s \in R_p, t \in R_{q_i}} w_p^s m_j(s, t)$$
(4)

其中 $s \in R_p$ 和 $t \in R_{q_j}$ 是参考图像和邻域图像的对应 点,参数 α 调节图像块的灰度值差异和梯度值差异 的权重, τ_c 和 τ_g 为两个控制最大差异的常数.仿射 权重函数定义为 $w_p^s = e^{\frac{\|I(p)-I(s)\|}{\gamma}}$,其中 γ 是参数,而 $\|I(p)-I(s)\|_1$ 计算I(p)和 $I(s)之间的 L_1$ -距离,其中 s为像素位置p的邻域.仿射权重w减少远离中心像 素的像素影响.

另外一种常用的匹配成本为自适应的零均值交 叉相关 ANCC^[32],其定义如公式(5):

$$\rho_{j} = 1 - \frac{\sum_{s \in R_{p}, t \in R_{q_{j}}} w_{p}^{s} w_{q_{j}}^{t} \left(I_{i}(s) - \overline{A}_{p}\right) \left(I_{j}(t) - \overline{A}_{q_{j}}\right)}{\sqrt{\sum_{s \in R_{p}} \left|w_{p}^{s} \left(I_{i}(s) - \overline{A}_{p}\right)\right|^{2}} \sqrt{\sum_{t \in R_{q_{j}}} \left|w_{q_{j}}^{t} \left(I_{j}(t) - \overline{A}_{q_{j}}\right)\right|^{2}}}$$
(5)

其中, $\overline{A}_p = \sum_{s \in R_p} w_p^s I(s)$ 和 $\overline{A}_{q_j} = \sum_{t \in R_{q_j}} w_{q_j}^t J(t)$ 是 窗口

 R_p and R_{q_i} 内图像灰度的加权均值.

对于采用转台和固定光源拍摄的室内文物场 景,本文观察到采用 intensity+gradient 成本可产生 更平滑精细的重建效果,而对于室外场景,为了增 加对于场景光照变化的鲁棒性,本文采用 ANCC 作 为匹配成本. 为了避免遮挡的影响,选择匹配成本最小的 K 个邻域图像子集 $V^*(i)$,则关于像素 p 的深度 d_p 和法 线 n_p 的多视角累积成本如公式 (6):

$$g^{*}(d_{p}, \boldsymbol{n}_{p}) = \frac{1}{\left| V^{*}(i) \right|} \sum_{I_{j} \in V^{*}(i)} \rho_{j}$$
(6)

对于参考图像 *I*, 邻域图像集合 *V* 的选择需要 保证场景内容尽可能交叠的同时,使得图像之间的 基线尽可能的大.对于采用转台拍摄的小体积文物 场景,由于图像围绕物体成圆形分布,本文采用参考 图像的视点方向 v_i 与邻域图像的视点方向 v_i 的角度 差异来选择邻域图像: $V(i) \leftarrow \{v_j | v_i \cdot v_j > \cos(\tau)\}$, 当角度差异的阈值 τ 设置为 45 度时,选择的邻域图 像的数量设置为 7.而对于无人机航拍的大佛造像 场景,由于包含了不同距离下多次拍摄的数据,本 文采用对尺度鲁棒的视点选择方法^[18]来进行邻域视 点的选择,邻域图像的最大数量设置为 9.

PatchMatch stereo 首先随机初始化深度和法线 值,方法通过交替执行邻域传播(propagation)和精化 (refinement)步骤来不断的优化深度和法线假设. 在文献[17]所提出的方法中,采用一种基于 GPU 的 红黑棋盘格传播算法,可以在整幅图像上并行的传 播邻域的深度与法线假设.其中精化操作采用二分 法,不断地在更小的区间内随机寻找更优的深度和 法线. 然而相比深度,由于法线包含 3 个自由度, 通过在三个分量上同时搜索不容易找到最优值.同 时低精度的法线将导致重建的结果无法恢复文物表 面的精细结构.为了提升法线的精度,本文根据前 期的工作,采用一种 normal-aware 的 PatchMatch stereo 方法^[19],基于当前估计的深度图计算出新的 法线假设,提升法线估计的准确度.

同时对于室内拍摄场景,在所有像素上进行深度精化是不必要的,因为部分场景实际对应的是背景区域,无需对其进行精细的估计.为此,本文首先检测背景区域,并在此区域上取消精化操作.检测背景采用三组阈值实现,如公式(7):

M(p) =

$$\Lambda(I^{r}(\boldsymbol{p}), \epsilon_{r}) \cdot \Lambda(I^{g}(\boldsymbol{p}), \epsilon_{g}) \cdot \Lambda(I^{b}(\boldsymbol{p}), \epsilon_{b})$$

(7)

其中 $I^{r}(p)$, $I^{g}(p)$ 和 $I^{b}(p)$ 分别表示像素 I(p) 的 R, G, B 分量. $\Lambda(x, t)$ 是一个二值函数, 当 $x \ge t$ 时值 为 1, 否则值为 0.

4.4 面向大尺度图像数据的增量深度图融合

当获得了一组输入图像的深度图 {*D*₁...*D_n*} 后,下一步的工作将是对深度图进行噪声的剔除和融合,生成三维点云.文献[17]提出一种全局融合方

法, 对于任意的深度值 d_p 和法线 n_p , 检测其与邻 域深度图的一致性, 如公式 (8):

$$f_j(d_p, \boldsymbol{n}_p) = \Lambda \left(\left\| \boldsymbol{p} - \boldsymbol{p}'_j \right\| < \epsilon_r \cap \left| \boldsymbol{n}_{q_j} \cdot \boldsymbol{n}_p \right| < \epsilon_n \right) \quad (8)$$

其中 $p'_{j} \neq p$ 经由邻域深度图 D_{j} 的重投影位置,其 可通过将 p 在邻域图像 I_{i} 的对应点 q 经由邻域深度 投影回 D_{i} 而得到. n_{ai} 是在像素位置 q_{i} 处的法线值. 参考图像上的深度和法线估计值将被认为是正确, 仅当满足公式 (9):

$$\sum_{i\in\Omega(i)} f_j(d_p, \boldsymbol{n}_p) \ge \beta$$
(9)

其中 $\Omega(i)$ 为除了参考视角 *i* 之外的所有邻域视角. β 为预定义的阈值,反映视角一致性.假如深度值 d_p 和法线 \mathbf{n}_p 通过上述一致性检查过程,并且一致 性邻域图像集为 $\Omega'(i) \subset \Omega(i)$,则其对应的三维点 X_p 与所有与该三维点一致的视点 $j \in \Omega'(i)$ 中的三维点 X_{ai} 平均,如公式(10)所示:

$$\overline{X}_{p} = \frac{\left(X_{p} + \sum_{j \in \Omega'(i)} X_{qj}\right)}{1 + |\Omega'(i)|}$$
(10)

由于上述融合方法的精确性和高效性,该方法 已经被广泛应用到最近的三维重建主流工作中.然 而在上述实现过程中,文献[17]方法需要将所有的 深度图和法线图全部调入到显存中,然后逐个参考 图像执行上述融合操作.其空间复杂度随着图像分 辨率和图像集规模近似线性增长.高清图像的深度 图和法线图的处理需要占用大量的显存,随着图像 尺寸增大一倍,所需要显存将增大4倍,随着输入 图像数量的增加,显存占用率也将迅速增加.然而, 硬件资源,特别是显存资源是非常有限的,这限制 了上述方法的应用范围.为此本文提出一种增量式 融合策略以减少对于显存的占用:

(1)设视角 *i* 为参考视角,令 *C* 为与深度图同 大小的一致性计数图像,初始值设置为 0.

(2) 对于图像集中的除视角 *i* 之外所有其它图 像集 Ω,每次增量地导入 α 张图像进入 GPU 的显 存,执行跨视角一致性检查.若参考视点的深度 *d_p* 和法线 *n_p*在与这 α 张图像比较时,可通过公式(8) 与公式(9)所确立的一致性判断,则:

(a) *p* 位置的一致性计数 *C*(*p*) = *C*(*p*) +1;

(b) 对 应 的 点 云 *X*(*p*) 执 行 增 量 累 加 操 作 *X_v* = *X_v* + *X_{oi}*.

当执行完 α 幅图像的视角一致性检查后,释放 其显存,继续导入后续的 α 幅图像,直至处理完图 像集Ω中所有图像.

(3) 若已处理完 Ω 中的所有图像,则对于所有 $C(p) \ge \tau$ 的像素位置,获得其融合三维点坐标 $X = X_p / C(p)$.

4.5 表面重建

给定一组三维点云,表面重建算法用于从点云 重建三角网格表面(mesh).本文基于 Screened Poisson surface reconstruction (SPSR)^[33]进行表面重 建,并剪枝掉三角形面积过大的区域.经由 SPSR 获取的模型包含数十万到数百万的三角面片,如此 稠密的模型将会限制在虚拟环境下的渲染效率.为 此需要在不损失模型质量的情况下尽可能减少面片 的数量.本文对 SPSR 输出的三维网格表面,使用 Qslim 算法^[34]进行必要的网格精简,从而保证简化 的三维表面包含的三角面片数量在十万面以下,同 时保留物体表面的几何结构.

5 实验结果

本文实验平台为联想 P500 工作站,搭载 CPU E5-1620v3 (3.5Hz),64 GB 内存,Nvidia GTX1080 显卡.固原须弥山石窟包括大量与丝绸之路文化相 关的珍贵遗址,如须弥山大佛(第五窟)、石门关和 相国寺等,其中须弥山大佛是该景区的代表建筑, 始凿于北魏孝文帝太和年间(公元 477 至 499 年), 高 20.6 米.针对须弥山大佛采集的部分航拍图像如 图 6 所示.经过 SFM 标定相机参数后,重建的稀疏 点云和相机位姿如图 7 所示.经由本文多视角立体 重建方法实现的须弥山大佛重建结果如图 8 所示. 可以看出本文方法基于高清航拍数据,在保证三维 形状完整重建的同时可复原表面精细几何结构,如 大佛的脸部和衣服的褶皱等.

固原地区馆藏文物的部分多角度采集图像如图 9 所示.基于这些图像数据,实现的三维重建结果如 图 10 所示.可以看出重建模型忠实复原文物的几何



图 6 须弥山大佛部分航拍图像



图 7 经由 SFM 重建的须弥山大佛稀疏点云和航拍相机 参数



图 8 须弥山大佛重建效果图(从左至右,依次是全局重 建效果展示和局部几何细节展示)



图 9 固原馆藏文物部分多视角图像



形状,模型中几何细节清晰可见.同时本文验证 normal-aware 的 PatchMatch stereo 方法在法线恢复方 面的能力,如图 11 所示,相比基准方法,normal-aware

图 10

PatchMatch stereo 可以更精准的恢复表面的法线信息.

本文进一步在公共数据集 ETH3D 上评测本文 提出的重建算法的有效性. ETH3D^[15]包含了室内外 场景的高清图像数据集,原始图像大小为 6,048× 4.032, 这些高分辨率图像能够较好地验证本文提出 的面向高分辨率图像三维重建算法的有效性. ETH3D 的真值数据使用专业的激光扫描仪获得,经 由 MVS 方法重建的点云可与真值点云数据进行量 化的比对,评价方法的有效性.本文在其13个训练 集上进行量化实验.



(a) 基准方法的法线图 (b) Normal-aware PathMatch 估计 的法线图 图 11 Normal-aware 重建效果展示

首先本文验证提出的增量融合的有效性,由于 在高分辨率图像的数据集上,全局方法将在一些图 片数量较多的数据上导致显存不足的问题,因此本 文将图像降采样到 1600×1064, 在相同的深度图计 算方法的基础上,分别采用全局融合和提出的增量 融合进行深度图的融合. ETH3D采用 F1度量重建的 综合质量,数值越高,质量越好.在实验中发现, 全局方法随着图像集规模的增加,显存占用增加, 最小显存占用为 1828.5 MB(14 张图像), 最大显存 的占用率为 4429.0 MB (76 张图像). 可以预见, 如果将输入图像大小增加到 3200×2128, 则显存将 增加到 16GB 以上,由此限制了方法对于大尺度数 据的应用. 而本文提出的增量方法可将最大显存占 用率控制在 1821MB 以内 (α=10),因此可以在有限 的显存下处理更多的数据. 通过表 1 的量化质量评 价可以看到,提出的增量方法的重建质量基本和全

表 1 在 ETH3D 上评测不同融合方法的显卡显存消耗 (MB) 与重建质量 F_1 (评估阈值为 0.02)

指标	全局 融合	増量融合 (α=20)	増量融合 (α=15)	增量融合 (α=10)
重建质量 F_1 (%)	72.88	72.75	72.78	72.63
最大显存 (MB)	4429	2400	2018	1821

局方法一致. 以上定量实验证明了增量融合方法的 有效性. 虽然通过增加显卡的数量可增强全局融合 方法的适应性, 但是全局方法的显存瓶颈将很快限 制它在更大规模数据的应用. 例如, 对于大佛数据, 由于包含大量的大尺寸深度图, 全局方法在两块显 卡上也将导致显存不足, 无法完成重建, 而本文提 出的增量融合方法可以在单个 8 GB 显存的显卡上 产生理想的融合结果.

为了验证本文提出的重建算法的整体有效性, 本文在 ETH3D 数据集上与当前主流方法进行了对 比,包括 Gipuma^[17],PMVS^[27],COLMAP^[16], CMPMVS^[35]和 ACMH^[28].需要指出的是,这些方法 和本文一样都是不包含正则化项的方法.一些包含 正则化项的方法虽然可提升重建在弱纹理区域的完 整度,但是也会导致物体表面细节被抹去,并且显 著增加计算时间.文物和古迹数据的纹理比较丰 富,且细节是其艺术价值的体现,因此非正则化的 方法更能适应这一特定应用.在表 2 中本文给出了 量化评价的结果, ETH3D 高分辨数据集鼓励高清图 像进行重建,如 ACMH^[28]和 COLMAP^[16]将图像缩 放为 3200×2128 大小进行重建. 对于本文提出的方 法,本文分别给出了低分辨率版本(1600×1064)和 高分辨率版本(3200×2128)的重建结果.从表2可 以看出,本文的高分辨率版本可以产生最好的重建 结果, 而低分辨率版本也在多数指标上优于主流算 法,由此验证了方法的有效性.一般来说,较高的 图像分辨率对计算机显卡的性能有更高的要求,如 ACMH 采用了 2 块 GTX Titan X, 总共 24 GB 显存, 而本文的计算机仅包含一个 GTX 1080 显卡(8 GB 显存).图12给出了本文方法(高分辨率版本)和 对比方法在 ETH3D 数据集上的定性(主观)评价结 果,可以看出本文方法明显提升了重建的完整度. 值得指出的是,表2的结果还说明了,通过在高分 率图像上执行重建可以提升重建质量,验证了本文 提出的面向高分辨率图像三维重建方法的潜在应用 价值.



图 12 ETH3D 数据集的定性评估结果(相比对比方法,可以看出本文方法明显提升了重建的完整度)

表 2 在 ETH3D 对比本文方法和主流 MVS 重建算法(其 中本文方法分别给出了低分辨率和高分辨率的重 建结果.其中加粗为最好的结果,下划线为第二好 的结果)

方法/阈值	0.01	0.02	0.05	0.1	0.2
Gipuma	25.87	36.38	49.17	58.52	68.35
PMVS	33.81	46.06	55.81	61.27	67.27
COLMAP	51.99	67.66	80.50	87.61	93.27
CMPMVS	49.24	62.49	73.84	79.86	84.76
ACMH	58.24	70.71	81.86	87.31	91.51
Ours(low-res)	56.97	72.78	<u>84.16</u>	<u>89.27</u>	92.90
Ours(high-res)	64.86	75.43	84.79	89.52	<u>93.02</u>

本文的重建算法主要包含 2 部分,深度图估计 和增量深度融合.以 ETH3D 的 *relief* 数据集为例, 其包含 31 张图像,本文采用 1,600×1,064 分辨率为 输入,计算每一视角图像的深度图的平均计算时间 为11.6秒,采用全局深度融合的时间为58.4秒,采 用本文增量融合(a=15)的时间为68.2秒,采用本 文增量融合(a=10)的时间为69.6秒.以 ETH3D 的electro为例,其包含45张多视角图像,计算每 一图像深度的平均时间为:11.2秒,采用全局深度 融合的时间为80.6秒,采用本文增量融合(a=15) 的时间为99.6秒,本文增量融合(a=10)的时间为 113.4秒.本文增量融合相比全局融合算法时间开销 略多,分析原因为数据从CPU内存到GPU内存的 传输带来了额外的时间开销.整体来说,本文提出 的重建算法具有较高的计算效率.相比全局融合算 法,本文提出的增量融合算法在减少显存的同时, 并没有显著增加计算时间,保证了算法的执行效率 和实用性.

最后,本文在文物数据"甲骑具装俑"上对比

了本文方法和著名的 PMVS 算法^[27]的重建效果,如 图 13 所示.可以看出,本文方法不仅产生更平滑的 表面,同时还有效复原了文物表面的精细结构,验 证了方法的优越性.重建的模型经过表面重建,可 导入到虚拟互动体验平台,进行内容展示和人机互 动体验,如图 14 所示.



(a) 本文方法算法

(b) PMVS 算法

图 13 重建效果对比图(可以看出,本文方法不仅可产 生更平滑的表面,同时保留表面的细节,更真实 地复原了文物的三维模型)



图 14 重建模型集成到沉浸式交互虚拟漫游系统

6 结 论

丝路文化传承等文化科技领域对虚拟现实互动体验系统的需求正在快速增长,本文基于虚拟现实 技术,围绕固原丝路文化遗产,通过高清图像数据 采集与面向大尺度图像数据的三维重建,提供了一 个高真实感、沉浸感和多种交互方式的虚拟现实互 动环境,可以为公众提供全新的交互体验方式,打 造动静相宜、内容丰富的互动漫游体验效果,使人 们能够身临其境般地感受中华传统文化的博大内 涵,提升丝路文化遗产的推广、普及和数字化服务 能力.

此外提出的整个系统框架,特别是三维重建方法,还可潜在应用于其它面向文化保护和传播的 VR 漫游系统中,如虚拟博物馆等.

致 谢 感谢国家自然科学基金(No. 61702482 和 61532002),以及北京市自然基金(No. L172049) 对本研究的资助.

参考文献

- Chen Ying. The application of virtual reality technology in the Inheritance and dissemination of non heritage culture. Study on Natural and Cultural Heritage, 2017, 2(05): 146-148 (in Chinese) (陈颖. 虚拟现实技术在非遗文化传承与传播中的应用. 遗产 与保护研究, 2017, 2(05): 146-148)
- Pan Yun-He, Lu Dong-Ming. Digital protection and restoration of dunhuang mural. Journal of System Simulation, 2003, 15(3): 310-314 (in Chinese)

(潘云鹤, 鲁东明. 古代敦煌壁画的数字化保护与修复. 系统 仿真学报, 2003, 15(3): 310-314)

- [3] Liu Jian. Research on virtual exhibition system for hemudu site museum. Journal of System Simulation, 2009, 21(7): 1945-1449 (in Chinese)
 (刘箴. 河姆渡遗址博物馆虚拟展示系统的研究. 系统仿真学
- 报, 2009, 21(7): 1945-1949)
 [4] Wang Sheng-Hua, Tan Jian. Chinese shadow play simulation based on volume visualization. Journal of System Simulation, 2015, 27(09): 2126-2134 (in Chinese)
 (王圣华, 谭剑. 中国皮影体可视化仿真方法研究. 系统仿真 学报, 2015, 27(09): 2126-2134.)
- [5] Liu Shi-Guang, Chen Di. Simulation technique for marbled paper patterns. Journal of Software, 2012, 23(Suppl.(2)): 1-7 (in Chinese) (刘世光,陈迪. 一种石纹纸染艺术图案仿真技术. 软件学报, 2012, 23(Suppl.(2)): 1-7.)
- [6] Zhang Xian-Quan, Yu Jin-Hui, Jiang, Ling-Lin, Tao Xiao-Mei. Computer assisted generation of paper cut-out images. Journal of Computer-Aided Design & Computer Graphics, 2005, 17(6): 1378-1382 (in Chinese)
 (张显全,于金辉,蒋凌琳,陶小梅. 计算机辅助生成剪纸形 象. 计算机辅助设计与图形学学报, 2005, 17(6): 1378-1382.)
- [7] Tu Chuan-Peng, Peng Ren, Chen Hai-Ying. Computer generation of water animation with the style of paper-cuts. Journal of Computer-Aided Design & Computer Graphics, 2009, 21(7): 949-953 (in Chinese)
 (涂传朋,彭韧,陈海英,于金辉. 计算机生成剪纸风格流水 动画. 计算机辅助设计与图形学学报, 2009, 21(7): 949-953.)
- [8] Liu Xiao, Wu Xun-Wei. Computer-rebuilding damaged cultural relics in three dimensions. Journal of Electronics & Information Technology, 2001, 23(7): 650-656 (in Chinese) (刘晓, 吴训威. 破损古文物的计算机三维重构. 电子与信息, 学报, 2001, 23(7): 650-656.)
- [9] Hu Yan-Qiu, Yu Jin-Hui, Jiang Wei, Peng Ren. Modeling and rendering of bronze articles. Journal of Computer-Aided Design & Computer Graphics, 2008, 20(9): 1140-1145 (in Chinese) (胡晏秋, 千金辉, 姜威, 彭韧. 面向青铜器的建模与绘制. 计 算机辅助设计与图形学学报, 2008, 20(9): 1140-1145.)
- [10] Kong Li-Ming, Rong Xiao-Man, The review of application of augmented Reality in culture presentation. China Cultural Heritage, 2017(02): 62-69 (in Chinese)
 (孔黎明, 荣晓曼.增强现实技术在文化遗产展示中应用综评. 中国文化遗产, 2017(02): 62-69.)
- [11] Shi-Guo-Wei, Wang Yong-Tian, Liu Yue, Zheng Wei. Digital conservation of cultural heritage using augmented reality. Journal of System Simulation, 2009, 21(7): 2090-2093. (in Chinese) (师国伟, 王涌天, 刘越, 郑伟. 增强现实技术在文化遗产数

字化保护中的应用. 系统仿真学报, 2009, 21(7): 2090-2093.)

- [12] Levoy M, Pulli K, Curless B, et al. The Digital michelangelo project: 3D scanning of large statues//Proceedings of the Siggraph. New Orleans, USA, 2000: 1-14
- [13] Seitz S, Curless B, Diebel J, Scharstein D, Szeliski R. A comparison and evaluation of multi-view stereo reconstruction algorithms//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York, USA, 2006: 519-526
- [14] Strecha C, Hansen W. von, Van Gool L, Fua P, Thoennessen U. On benchmarking camera calibration and multiview stereo for high resolution imagery//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA , 2008: 1-8
- [15] Schöps T, L. Schönberger J, Galliani S, Sattler T, Schindler K, Pollefeys M, Geiger A. A multi-view stereo benchmark with high-resolution images and multi-camera videos// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 1-8
- [16] Schönberger J. L, Zheng E, Pollefeys M, Frahm J. M. Pixelwise view selection for unstructured multi-view stereo//Proceedings of the European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 1-17
- [17] Galliani S, Lasinger K, Schindler K. Massively Parallel Multiview Stereopsis by Surface Normal Diffusion//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 873-881
- [18] Bailer C, Finckh M, Lensch H P A. Scale robust multiview stereo//Proceedings of the European Conference on Computer Vision. Florence, Italy, 2012: 1-14
- [19] Li, Z, Zuo W, Wang Z, Zhang L. Confidence-based large-scale dense multi-view stereo. IEEE Transactions on Image Processing, 2020, 29: 7176-7191
- [20] Kostrikov I, Horbert E, Leibe B. Probabilistic labeling cost for high-accuracy multi-view reconstruction//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 1-8
- [21] Hane C, Zach C, Cohen A, Pollefeys M. Dense semantic 3D reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(9): 1730-1743
- [22] Delaunoy A, Prados E. Gradient flows for optimizing triangular mesh-based surfaces: applications to 3D reconstruction problems dealing with visibility. International Journal of Computer Vision,



LI Zhao-Xin, Ph.D., assistant professor. His research interests include 3D reconstruction and 3D computer vision.

JIANG Hao, Ph.D., associate professor. His research interest is crowd

2011, 95(2): 100-123

- [23] Li Z, Wang K, Zuo W, Meng D, Zhang L. Detail-preserving and content-aware variational multi-view stereo Reconstruction. IEEE Transactions on Image Processing, 2016, 25(2): 864-877
- [24] Yao, Y, Luo Z, Li S, Shen T, Fang T, Quan L. Recurrent mvsnet for high-resolution multi-view stereo depth inference //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 5520-5529
- [25] Im S, Jeon H, Lin S, Kweon I. Dpsnet: End-to-end deep plane sweep stereo//Proceedings of the International Conference on Learning Representations. New Orleans, USA, 2019: 1-12
- [26] Furukawa Y, Hernndez C. Multi-view stereo: A tutorial. Foundations and Trends in Computer Graphics and Vision, 2015, 9(6): 1-148
- [27] Furukawa Y, Ponce J. Accurate, dense, and robust multiview stereopsis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 32(8): 1362-1376
- [28] Xu Q, Tao W. Multi-scale geometric consistency guided multiview stereo//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 5483-5492
- [29] Snavely N, Seitz S. M, Szeliski R. Photo Tourism: Exploring image collections in 3D. ACM Transactions on Graphics, 2006 (1): 835-846
- [30] Wu C. Towards Linear-time incremental structure from motion// Proceedings of the 3D Computer Vision. Seattle, USA, 2013: 127-134
- [31] Johannes Lutz S, Jan-Michael F. Structure-from-motion revisited// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 4104- 4113
- [32] Heo Y, Lee K, Lee S. Robust stereo matching using adaptive normalized cross-correlation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(4): 807-822
- [33] Kazhdan M, Hoppe H. Screened poisson surface reconstruction. ACM Transactions on Graphics, 2013, 32(3): 1-13
- [34] Garland M, Heckbert P. S. Surface simplification using quadric error metrics//Proceedings of the ACM SIGGRAPH. New York, USA, 1997: 209-216
- [35] Jancosek M, Pajdla T. Multi-view reconstruction preserving weakly-supported surfaces// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Colorado, USA: Springs, 2011: 3121-3128

simulation.

LIU Yan-Qing, Ph.D., professor. Her research interest is Guyuan historic culture.

WANG Zhao-Qi, Ph.D., professor. His research interest is virtual reality.

Background

This research focuses on 3D reconstruction and visualization of the historical sites and cultural relics. It is overlapping between cultural heritage protection and computer science. Silk Road culture is an important link in the Belt and Road strategy. Its heritage is of great significance. However, due to historical and geographical reasons, the representative historical heritage in the Silk Road culture is scattered or damaged, and it is difficult to present it effectively. Through historical restoration and image-based 3D reconstruction, we restored the historical sites, cultural relics and events of Guyuan, one of the important nodes in Silk Road Culture. Especially, we first reconstruct the historical sites and cultural relics, and then visualize them in the virtual reality application. The 3D reconstruction of the historical sites and cultural relics is an important research area in computer vision, architecture and archaeology. The traditional 3D reconstruction methods for the historical sites and cultural relics rely on laser scanner, which is bulky and expensive. In this paper, we propose an effective and low-cost 3D reconstruction method based multi-view stereo. The proposed method fully explores multi-view highresolution images and can generate the high-quality 3D surface with detailed geometric features for both outdoor large-scale historical sites and indoor small-scale cultural relics. We also propose and implement a virtual reality platform for the Silk Road Cultural Heritage.

The proposed MVS 3D reconstruction method for high-resolution images, which consists of a normal-aware PatchMatch for the high-quality normal recovery to represent the detailed surface of the cultural relics, and a GPU-friendly incremental depth fusion method which can fuse a large amount of depth maps by a small size of GPU memory. The experiment results on the public MVS benchmark and our captured datasets highlight that the proposed method can recover the detailed surfaces while keeping a good scalability for the large-scale data. The reconstructed models can effectively support VR application, playing a positive role in the dissemination of Silk Road culture.

This work was supported by National Natural Science Foundation of China under Grants No. 61702482. This NSFC project mainly focus on the complete and detailed 3D reconstruction based multi-view stereo. The research team has worked on 3D reconstruction research for more than 10 years and has already built a complete pipeline for the entire reconstruction procedure, including depth map estimation, volumetric fusion and mesh-based refinement. The related publications include IEEE Transactions on Image processing, Image and Vision Computing, Neurocomputing and The Visual Computer.

一种基于自演化预训练的多变量时间序列预测算法

万 晨^{1),2)} 李文中^{1),2)} 丁望祥^{1),2)} 张治杰^{1),2)} 叶保留^{1),2)} 陆桑璐^{1),2)}

²⁾(南京大学计算机科学与技术系 南京 210023)

摘 要 时间序列预测是典型的时间序列分析任务,对于辅助决策、资源配置、提前采取止损措施等方面有重要 意义,在包括电力、气象、交通、商业等领域有广泛应用.近年来,时间序列预测算法一直是机器学习的热门研 究领域,其中多变量时间序列预测是一个具有挑战性的任务.本文研究多变量时间序列预测的局部变量预测精度 问题,即多变量预测需要在提升整体预测性能的同时保证局部单变量的预测精度.针对现有多变量时间序列预测 算法不能保障局部变量预测精度的局限性,我们设计并实现了一种基于自演化预训练的多变量时间序列预测算法 SEPNets.基于预训练的思想,SEPNets首先构建和训练单变量时间序列模型作为后续建模的基准.然后,通过拓展时序卷积网络和长短记忆(LSTM)单元来建模变量间复杂的时序依赖关系.通过将预训练模型和拓展模型进行 融合再训练,SEPNets可以保障多变量时间序列预测的局部变量预测精度,并提升总体的预测性能.我们在5个 真实数据集上对所提模型进行性能评估.实验结果表明,本文提出的 SEPNets 算法比现有算法获得相对最高的预 测精度,同时在保障局部变量预测精度上具有更好的性能.

关键词 时间序列预测;多变量时间序列;机器学习;神经网络;预训练中图法分类号 TP391 **DOI**号 10.11897/SP.J.1016.2022.00513

A Multivariate Time Series Forecasting Algorithm Based on Self-Evolution and Pre-training

WAN Chen^{1),2)} LI Wen-Zhong^{1),2)} DING Wang-Xiang^{1),2)} ZHANG Zhi-Jie^{1),2)} YE Bao-Liu^{1),2)} LU Sang-Lu^{1),2)}

¹⁾(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210023) ²⁾(Department of Computer Science, Nanjing University, 210023)

Abstract Time series forecasting is a typical time series analysis task, which is of great significance for assisting decision-making, resource allocation, and taking stop loss measures in advance. It is widely used in fields including power, weather, transportation, and business. In recent years, time series forecasting algorithms based on machine learning have been a hot research field, among which multivariate time series forecasting accuracy of multivariate time series forecasting, that is, multivariate forecasting needs to improve the overall forecasting performance while ensuring the forecasting accuracy of local univariate forecastings. In view of the above problems, this paper analyzes the limitations of the multivariate time series forecasting algorithm. The multivariate time series forecasting algorithm SEPNets based on self-evolution pretraining is designed

收稿日期:2020-05-19;在线发布日期:2021-07-29.本课题得到本课题得到国家重点研发计划(2018YFB1004704)、国家自然科学基金(61972196,61672278,61832008,61832005)、江苏省重点研发计划项目(BE2018116,BE2017152)、软件新技术与产业化协同创新中心、中德社会计算研究所资助.万晨,硕士研究生,主要研究领域为时间序列分析、深度学习.E-mail:mf1733062@smail.nju.edu.cn. 李文中(通信作者),博士,教授,中国计算机学会(CCF)会员,主要研究领域为分布式计算、社交网络分析、深度学习.E-mail:lwz@nju.edu.cn. 丁望祥,硕士研究生,主要研究领域为时间序列分析、深度学习.张治杰,硕士研究生,主要研究领域为时间序列分析、深度 学习.叶保留(通信作者),博士,教授,中国计算机学会(CCF)会员,主要研究领域为分布计算与并行处理.E-mail:yebl@nju.edu.cn. 陆桑璐(通信作者),博士,教授,中国计算机学会(CCF)会员,主要研究领域为分布计算与并行处理.E-mail:sanglu@nju.edu.cn.

and implemented. Inspired by the pre-trained model, SEPNets first constructs and trains a univariate time series model as a benchmark for subsequent modeling; then expands the complex temporal dependence between the time-series convolutional network (1-dimensional convolution) and the long and short memory (LSTM) unit modeling variables Relationship; using the former as the latter pre-training model for fusion and retraining, SEPNets solves the problem of local variable forecasting accuracy for multivariate time series forecasting. The experimental results show that the SEPNets algorithm proposed in this paper ensures the forecasting accuracy of local variables to a certain extent while obtaining the relatively highest forecasting accuracy.

Keywords time series forecasting; multivariate time series; machine learning; neural networks; pre-train

1 引 言

1.1 研究背景

近年来,时间序列挖掘成为了备受关注的研究 领域,其研究问题包括时间序列的分类、预测、关 联、聚类、回归、摘要、异常检测等.时间序列预 测 (time series forecasting, TSF)的主要目的是基于 历史观测的时间序列数据,分析序列的时序模式(包 括趋势、周期性、波动性等),建立合适的序列模型, 然后使用该模型为对应的时间序列生成未知的未来 数据,即进行预测.TSF 是时间序列建模中通过理解 过去来预测未来的算法^[1].

采用时间序列预测方法、获得未来时刻的预测 值,对于辅助决策、优化资源分配、提前采取止损 措施等具有基础且重要的意义.在电力、气象、交 通、商业等领域,时间序列预测算法有着广泛应用. 例如,亚马逊等大型商业组织依据需求信息及时做 出正确的决策,在不断变化的商业环境中增强商业 竞争优势、降低成本、提高商业利润,所以对整个 供应链中的商品需求量、供应量、价格等进行预测 具有必要性.此外,实时运行的信息系统产生了大 量的时序数据,包括系统性能监测数据等,对这些 监测数据进行预测是很多应用的需求.例如预测网 络流量用于资源规划与异常发现^[2],预测数据中心 的磁盘故障以提前替换^[3]等.因此预测系统性能数 据的趋势可以为现代信息系统的日常运维工作提供 重要的决策依据.

为建立有效的时间序列预测模型以提高预测性能,研究人员做出了许多努力.长期以来,线性统计类算法(例如 AR^[4]、ARIMA^[5])对 TSF 领域产生了影响.由于线性模型在很多实际应用中缺乏适用性^[6],非线性时间序列分析和预测被提出并逐渐应用^[7].

除了经典的统计方法模型,机器学习模型已经 成为时间序列预测领域的主流技术^[6].研究人员将 线性回归模型、支持向量机^[8]、决策树等机器学习 模型应用于时间序列分析. 基于机器学习的时间序 列预测算法把时序预测作为监督学习任务,其中输 入属性为历史观测的时间序列数据,输出标签为(训 练集上的)未来数据.机器学习模型以数据驱动的 方式,学习序列特征,建模历史数据和未来数据之 间的映射关系. 在实际应用中, 基于机器学习的时 间序列预测算法表现优于基于统计的算法[9].为了 改善时间序列预测算法,人们尝试在模型复杂度和 序列特征表示方面进行突破. 以神经网络为代表的 深度学习算法是机器学习研究的新兴领域^[10],其逐 层构建与处理的机制具有显著的模型内特征转换和 足够的模型复杂度的优势^[11].最近在时间序列建模 预测中采用深度神经网络算法成为了研究热点,特 别是在大数据和复杂任务的场景.无论在对模型的 理论理解和在模型优化方面,都取得了令人瞩目的 成果. 例如, 基于递归神经网络的算法通过使用有 状态模型建模时间序列数据的序列特征,获得了出 色的预测性能^[2,12].

尽管基于机器学习的时间序列预测问题已被广 泛研究,多变量时间序列预测仍然是具有挑战性的 研究问题.在多变量时间序列预测任务中,相关的 多个时间序列被并行地统一建模和预测.例如,预 测多个家庭的能源消耗、数据中心中多台服务器的 负载、大型零售商提供的所有产品的需求等.在这 些情况下,可以利用时间序列间的相似性,建模多 变量时间序列依赖关系,以提高整体的预测精度.捕 捉复杂时序模式与变量间依赖关系是多变量时间序 列预测研究的重点和实现精准预测模型的关键^[13].

1.2 研究动因

同一系统中,多个序列组成的多变量时间序列

本质上具有相关性,不同于单变量预测任务,利用 变量间的相关性进行建模,可以提高整体预测精度. 多变量序列预测有着广泛的应用场景,为了获得较 好的预测精度,需要有效建模复杂时序模式与变量 间的依赖关系. 代表性多变量时间序列预测算法包 括时序分析类统计算法 VAR^[14], 基于 LSTM 网络 的时序建模[15,16],和融合卷积神经网络和递归神经 网络的 CNN-GRU-Input(All)-Att^[17]和 LSTNets^[18]算 法等. 然而, 现有的多变量时间序列预测算法主要 关注整体预测目标优化,会带来局部变量预测精度 损失的问题.例如,在多变量预测的回归任务中, 现有算法以所有变量的整体预测误差加权和为最小 化目标,通过优化该目标提高整体预测精度. 然而, 整体优化目标可能忽略局部变量的预测精度损失, 造成牺牲局部变量的精度来提升整体预测精度的 问题. 在实际应用中, 对于一个复杂的多变量系 统,可能会存在某些非常重要的指标,如果为了提 高整体的预测精度而牺牲了重要指标的预测精度, 可能会带来重大损失. 例如, 在亚马逊等电商系统 中记录了许多指标数据,如商品需求量、供应量、 价格、评星等,其中一旦如商品需求量、价格等重 要局部指标的预测精度得不到保障,就会对企业的 决策造成较大的影响.此外,在性能监测系统中需 要监测多种数据,比如 CPU 利用率、内存利用率、 响应时间、吞吐量等. 如果不能保障响应时间和吞 吐量等重要指标的预测精度,可能会导致运维人员 无法针对这些重要的指标提前合理地采取止损措 施.因此,多变量时间序列预测算法在优化整体预 测精度的同时,需要保证局部变量的预测精度不会 造成损失.

我们通过实验评估来展示局部变量预测精度的 损失问题.图 1 给出对部分多变量时间序列预测算 法在 Electricity 数据集(具体描述见第4节)上的 精度评估结果.横坐标表示四种不同的多变量预测 算法,分别为 VAR^[14]、LSTM^[19]、LSTNets^[18]、 Expected (表示理想算法);纵坐标表示对应算法在 所有变量上的精度.其中,对每个变量单独采用单 变量预测算法 AR 进行建模和预测时,其平均精度 记为 P_0 .当采用多变量时间序列预测算法后,四个算 法的整体平均预测精度分别为 P_1, P_2, P_3, P_4 .对每种算 法,蓝色条形表示预测精度上升或不变的局部变量的比例, 橘色条形表示预测精度上升或不变的局部变量的比例, 橘色条形表示预测精度上升或不变的局部变量的比例, 属色条形表示预测精度上升或不变的局部变量的比例, 部变量的预测精度低于单变量预测算法.究其原因,已有算法以提升多变量整体预测精度为优化目标,并不能保障局部变量的精度损失.



图 1 多变量时间序列预测算法的精度估计(Electricity 数据集)

将单变量序列融入多变量序列预测中,其目的 是获得自身预测精度的提升,或者至少不比单序列 自身预测精度差.因此,理想的多变量时间序列预 测算法应如图1的Expected算法所示,在提升整体 预测精度的同时,保障局部变量的预测精度不低于 单变量预测算法的性能.为此,需要研制具有局部 变量性能保障的多变量时间序列预测算法.

1.3 解决思路与贡献

为了解决多变量时间序列预测的局部变量精度 损失问题,针对任务的复杂性以及多变量和单变量 优化目标的差异性,我们采用深度学习模型进行多 时间序列建模,并提出基于自演化预训练的预测算 法.在以上问题中,单序列预测模型的精度是对局 部变量预测性能保障的基准.因此,在多变量时间 序列模型的整体优化中如何有效地融合单序列预测 模型,是解决问题的关键.预训练(pretrain)策略 可以在类似任务的已训练模型的基础上,构建当前 任务的模型,并继续训练.由于当前模型采用了已 训练模型表示(包括网络结构与权值),等价于迁移 已有的知识表示来实现后续任务(downstream task). 预训练机制已被用于计算机视觉(CV)与自然语言 处理(NLP)领域,在节省资源的同时能获得良好的 性能^[20].

基于预训练策略,结合多变量序列预测背景, 本文首先设计单序列基准模型,并在每个单序列上 进行单独训练,作为后续整体任务的"基准知识", 其蕴含了单变量自身演化(Self-Evolution)模式; 然后,设计多层次 CNN-RNN 级联神经网络单元作 为多变量建模模块,先后对多变量间相互依赖 (Inter-Dependency)关系进行局部空域信息(local spatio)建模和时序(temporal)信息建模;最后, 把已训练完成的单变量模型作为预训练模块,与多 变量建模模块的输出一一对应融合,从而利用已训 练完成的单序列模型来优化整体模型的性能.该思 路能够在多变量模型的整体优化中有效地融合已训 练完成的单序列模型,解决局部单变量的预测精度 比对应的单序列模型的预测精度差的问题.

本文的主要贡献总结如下:

本文提出并研究了多变量时间序列预测的局部变量预测精度问题.通过实验观测,评估已有基于机器学习的多变量时间序列预测算法的局限性,这些算法在优化整体预测精度时,可能造成局部单变量的预测精度损失.

 为保障局部变量的预测精度,提出了基于自 演化预训练的多变量时间序列预测算法.基于预训 练策略与卷积网络、长短记忆网络深度序列模型, 结合单变量自演化信息与多变量依赖关系信息进行 建模,实现了多变量时间序列预测的优化算法,同 时兼顾了整体预测精度和局部单变量的预测精度.

•基于五个真实数据集进行对比实验,评估了 所提算法的有效性.实验结果表明,基于自演化预 训练的多变量时序预测算法,具有较好的整体预测 精度,对于局部单变量的预测精度保障性方面远优 于现有的多变量时间序列预测算法.

本文的组织结构如下.第2章介绍相关工作. 第3章介绍基于自演化预训练的多变量时间序列预 测模型和算法.第4章对所提算法进行性能评估和 分析.第5章对全文进行总结.

2 相关工作

多变量时间序列预测方法利用多个时间序列间的相关性,并以此建模提高整体预测性能.预测多 变量时间序列的关键是如何从时间序列中精确捕捉 复杂的时序模式与变量间依赖关系^[13].向量自回归 模型(Vector Autoregression, VAR)是常用的典型 统计时间分析类的多变量时间序列分析算法.VAR 将多变量时间序列看作多维随机过程,在单变量 AR 模型的基础上对变量依赖关系进行线性假设.VAR 描述的随机过程是线性的,用于捕捉多变量序列间 常见的线性模式,适用于平稳序列建模.向量差分 整 合 移 动 平 均 自 回 归 模 型 (Vector ARIMA, VARIMA)是 ARIMA 算法的多变量扩展.该模型能 够以随机过程描述更多的信息,但一般结构复杂优 化困难,其实用效果不如 VAR (VAR 拟合前可进行 简单的差分平稳化).高斯随机过程 (Gaussian Proccess, GP)是另一类多变量时间序列分析的统计 算法^[21].该模型先验假设多变量时间序列之间服从 多维高斯分布,其由均值向量与协方差矩阵共同决 定.尽管这些统计类时间序列分析模型在多个领域 得到应用,但它们假设了线性的时序模式,并且不 适合建模长期依赖关系.

由于深度学习模型在复杂度和序列特征表示方 面表现优秀,目前多变量时间序列预测算法越来越 多地采用深度神经网络进行构建,以捕捉复杂的时 序模式与变量间依赖关系^[6]. 全连接神经网络 (Nerual Networks, NNs) 最早用于对时间序列进行 建模预测[22]. 层间全连接的神经网络能够建模非线 性等时序关系,并且十分灵活.受到序列建模与权 值共享思路的启发,循环神经网络(Recurrent Neural Network, RNN)被提出. RNN 为每个时刻增加了状 态向量*h*,并且增加了层内前一时刻状态到当前时 刻状态的有向连接,这有效建模了序列信息,即当 前时刻的状态不仅取决于当前输入,而且依赖于之 前时刻的信息, 而h, 是时刻 t 前所有信息的表示. 基于递归神经网络的算法通过使用有状态模型建 模时间序列数据的序列特征,并获得较好的预测性 能^{[2][12]}

为了解决 RNN 中梯度消失问题^[23]带来的训练 困难以及长时段依赖关系的捕捉,人们提出了基于 门机制的 Long-Short Term Memory(LSTM)算法^[19]. LSTM 是一种性能出色的深度序列模型,常用作大 数据、复杂任务场景下的序列建模^[24,25],或者作为 基本单元与其他算法融合进行序列建模^[26]. LSTM 还有多种变体,诸如简化门控制机制的 GRU^[27]、关 注序列相位信息的 phi-LSTM^[28]、显示地跨时间步连 接的 skip-LSTM^[29]等.

为了获得多变量时间序列的有效特征表示,研究人员采用自编码器(Auto-Encoder, AE)对多变量时间序列进行编码.Wang等人提出了一种基于优化的时序卷积网络并带有堆叠式自动编码器的预测算法,可以有效表达多变量时间序列中时序关系的复杂性和非线性^[30].基于编解码架构,Qin等人提出基于两阶段注意力(Attention)机制的循环神经网络多变量预测算法DA-RNN^[31],分别在时间序列编码阶段与解码预测阶段加入注意力机制,自适应的选择相关的输入特征表示,有利于捕捉长时间的时序信息与依赖关系.Huang等人提出了一种双重自注意力网络(DSANet),用于高效的多元时间序列

预测,适用于动态周期或非周期序列的预测[32].

除了典型的深度序列模型 RNN,人们发现诸如 卷积神经网络(CNN)等也可以作为序列建模单元^[33]. 为了建模时间序列之间存在的强相关性,Hu等人提 出了一种基于时序卷积网络的多变量时间序列预测 模型^[34].Ong 等人提出了一种时间张量转换网络 (Temporal Tensor Transformation Network)^[12]深度 学习架构,以解决现有算法的强统计假设、高维数 值、手动特征工程和可伸缩性等问题.另外,Assaf 等人证明了 CNN 深度神经网络不仅可以用于多元 时间序列数据预测,还可以用于解释这些预测^[35].

研究人员试图结合时间序列分析算法与深度神经网络来构建混合模型,从而利用两者的优势来提高性能.对于多变量时间序列建模通常存在的显式线性相关性成分和非线性成分信息,Goel等人提出了基于残差学习的混合模型 R2N2^[36].针对多变量时间序列间短期与长期依赖关系的混合问题,Lai等人提出一种长期和短期时间序列网络深度学习框架^[18],使用卷积神经网络(CNN)和递归神经网络(RNN)来提取变量之间的短期局部依赖模式,拓展 skip-LSTM 捕获时间序列趋势的长期模式,进行模型构建.Thongniran 等人提出了将注意力机制(attention)加入 CNN 与 RNN 结合的模型中来进行多变量时间序列的建模^[17].

近年来,国内学者也对多变量时间序列进行了 大量研究.张鹏程等人提出了一种基于多元时间序 列的 QoS 预测方法 MulA-LMRBF^[37],通过网络开 源数据和自测数据的实验结果表明,该方法与传统 方法相比有较好预测效果,更适合动态多步预测. 欧阳红兵等人提出将小波分析与纳入时间序列依赖 特征的长短期记忆(LSTM)神经网络相结合,构建 时间序列数据预测模型^[38].王双成等人在基于高斯 函数估计属性密度的基础上,结合转换数据集构建、 回归变量的离散化、类变量的数量化、属性联合密 度的分解计算等,建立用于时间序列预测的具有超 父结点贝叶斯网络回归模型^[39].此外,李文海等人 提出一种基于 DTW 距离的相似样本度量方法^[40], 通过误差调谐函数对 SVM 的核函数进行调整来提 升预测的精度.

总之,已有基于机器学习的时间序列预测算法 关注于如何有效的捕获复杂时序模式和变量间关 系,并建立整体优化预测模型.在整体模型优化过 程中,往往以所有变量的预测误差和(或者加权和) 为最小化目标.正如本文引言部分分析,现有算法 在优化全部变量的整体预测精度时,不保证每个单 变量时间序列的预测精度,可能导致局部变量预测 精度损失的问题.本文提出的自演化预训练的多变 量时间序列预测算法,可以在提升整体预测精度的 同时,保障单变量时间序列的预测精度不会降低.

3 自演化预训练的多变量时间序列 预测算法

3.1 问题建模

本文将多变量时间序列 $x_t \in \mathbb{R}^N$ 形式化描述如下:

$$\begin{aligned} \mathbf{x}_{t} &= process(\mathbf{x}_{t-T:t-1}) \\ &= merge(\left[SE(\mathbf{x}_{t-T:t-1}), ID(\mathbf{x}_{t-T:t-1})\right]) + \varepsilon_{t} \end{aligned}$$
(1)

其中 N 为多变量时序数据的变量数, T 为通过自相 关函数 ACF 来评估时间序列的周期性而得到的历 史输入序列长度, \mathbf{x}_{t-T} 表示时刻 t 前 T 长度的历史时 间序列. Process(·) 把多变量序列抽象为随机过程, 而 merge(·) 进一步假设这样的随机过程是两部分的 融 合 结 果; SE(·) 表示 单 变 量 的 自 演 化 模 式 (self-evolution), ID(·) 表示多变量间的依赖关系 (inter-dependency); ε_t 为多维随机噪声.

本文针对 SE(·)和 ID(·),进行建模、构建学习模块,然后在 merge(·) 过程进行融合训练,具体介绍如下.

3.2 模型框架

本 文 提 出 基 于 自 演 化 预 训 练 的 神 经 网 络 (Self-Evolution Pretrained Neural Network, 简称为 SEP-Nets)模型,整体框架如图 2 所示.

该框架将具有 N 个变量数的时间序列 $\left[x^{(1)},...,x^{(N)}\right]$ 作为输入,经过两个模块的处理融合后,输出所有变量序列下一时间步的预测值向量 x_i .具体步骤如下.

第一步,将 N 个单序列 $\left[x^{(1)},...,x^{(N)}\right]$ 输入到自 演化预训练模型 SE(·)中,进行单序列建模、预测, 形成对应输出 $\mathbf{x}_{t}^{(se)}$;第二步,将 N 个单序列组成 的二维张量 $\left[x^{(1)},...,x^{(N)}\right]$ 先进行大小不同的卷积操 作,提取变量间依赖关系信息得到多个特征图 (feature map),然后输入到长短记忆网络(LSTM) 中进行时序建模,这样的多变量建模的网络组件 (Nets)形成对应输出 $\mathbf{x}_{t}^{(id)}$;最后,将前两步的 $\mathbf{x}_{t}^{(se)}$ 和 $\mathbf{x}_{t}^{(id)}$ 进行向量元素的——对应相加融合,构建出 多变量预测输出 \mathbf{x}_{t} .以下介绍各部分的细节.

3.3 自演化预训练模型

自演化预训练建模任务是单序列预测任务,为



整体多变量任务建模、训练提供基准,并可以作为 局部单变量预测精度的评估标准.在大量各异的单 序列预测任务的情况下,需要平衡预测精度和模型 复杂度,应对各时序模式各异的单序列,算法必须 具有广泛的适用性.

从简单、有效的时序分析的统计类算法出发, 本文提出基于差分特征构造的线性自回归模型 (enhanced autoregression, EAR).首先,建立线性 自回归模型.对于第*i*个序列有 $x_t^{(i)} = \sum_{j=1}^T w_j^{(i)} x_{t-j}^{(i)}$,其 中 $w_j^{(i)}$ 为线性权值.借鉴差分平稳化的思路,线性 回归模型构造"多阶差分特征".考虑差分阶数 q=1时,有 $(x_t^{(i)} - x_{t-1}^{(i)}) = \sum_{j=1}^T w_j^{(i)} (x_{t-j}^{(i)} - x_{t-j-1}^{(i)})$ 即 $x_t^{(i)} = x_{t-1}^{(i)} + w_1^{(i)} x_{t-1}^{(i)} + \sum_{j=2}^T (w_j^{(i)} - w_{j-1}^{(i)}) x_{t-j}^{(i)}$,可见一阶差

分下预测值含有 $x_{t-1}^{(i)} = q_1^{(i)}$ 常数项(其余项含有权值 变量); 类似地,差分阶数 q = 2时,有常数项 $2x_{t-1}^{(i)} - x_{t-2}^{(i)} = q_2^{(i)}$;类似地,差分阶数 q = 3时,有常 数项 $3x_{t-1}^{(i)} - 3x_{t-2}^{(i)} + x_{t-3}^{(i)} = q_3^{(i)}$.考虑多阶差分信息可 以在平稳、非平稳序列的预测中提高适用性、鲁棒 性,把 $q_1^{(i)}$ 、 $q_2^{(i)}$ 、 $q_3^{(i)}$ 等差分特征显示地融入到线 性自回归模型中,形式化如下:

$$\mathbf{x}_{t}^{(se,i)} = softmax \left(\left[a_{1}^{(i)}, \cdots, a_{Q}^{(i)} \right] \right) \left[q_{1}^{(i)}, \cdots, q_{Q}^{(i)} \right]^{T} \quad (2)$$
$$+ AR(x_{t-T:t-1})$$

其中 $\left[a_1^{(i)}, a_2^{(i)}, \cdots, a_Q^{(i)}\right]$ 为差分特征的融合权值,限制

为非负实数, *softmax*(·)为对应的归一化操作, Q为最大差分阶数; $AR(x_{t-T:t-1}) = \sum_{j=1}^{T} w_{t-j}^{(i)} x_{t-j}^{(i)}$ 为标准的线性自回归项, T为历史输入序列长度.

上述基于"多阶差分特征"的线性自回归模型 为整体多变量任务建模、训练提供基准的同时评估 各个变量的情况.如图 2 中上半部分,第*i*个时间序 列 *x*⁽ⁱ⁾采用 Self Evolution (i)进行建模,由于不同的 单变量时间序列的周期性各不相同,所以选取的历 史窗口长度*T*也可能不相同,因此在图上方所画的 单变量序列长度是不一样的.

3.4 基于CNN-LSTM的多变量依赖关系建模

捕捉复杂动态的时序模式与变量间依赖关系是 多变量序列预测的关键.相关变量序列的相关性往 往体现在变量间存在着局部依赖关系,即单变量序 列在某一时间点的取值依赖于相邻时间段其他相关 序列的取值.本文采用多个卷积核的一维 CNN 层的 时序卷积对多元时间序列之间的复杂依赖关系进行 表示,得到蕴含多变量序的局部依赖关系多层次信 息的特征图(feature map).对于输入序列 $X_{t-T:t-1} = \left[x_{t-T:t-1}^{(1)}, x_{t-T:t-1}^{(2)}, \dots, x_{t-T:t-1}^{(N)}
ight]$ 进行变量维度 上的一维卷积操作,表示如下:

$$C^{(k)} = W^{(k)} * X_{t-T:t-1}$$
(3)

其中W^(k)为第 k 个卷积核, C^(k)是卷积结果, 即从 变量间依赖关系中提取的特征图.

在卷积网络中,通常采用 K 个卷积核 $W^{(1)}, W^{(2)}, \dots, W^{(K)}$ 形成多个通道的特征图 $C^{(1)}, C^{(2)}, \dots, C^{(K)}$.不同于现有算法将所有卷积核设定为相同尺寸,本文采用可变尺寸的卷积核.如图 2 中一维卷积部分,采用

 $W^{(.)} \in R^{1 \times N}, W^{(2,c)} \in R^{3 \times N}, W^{(3,c)} \in R^{5 \times N} (c = 1, 2, ..., C)$ 卷积核组,以此建模多变量序列间的局部依赖关系, 其中 C 为每个卷积核的通道数.与同尺度卷积核相 比,多尺度卷积核显式地描述了不同尺度的关联时 间段特征,如公式(3)计算出的多尺度特征图记为 $C^{(1)}, C^{(2)}, ..., C^{(K)} \in R^{T \times C}$,蕴含着多变量序的局部依 赖关系的多层次信息.上述卷积操作是对变量维度 进行的,所以多尺度特征图依然保留着原有的时序 关系.

完成多变量序列一维卷积的特征提取后, SEPNets采用长短记忆网络 LSTM 对多尺度特征图 进行时序建模,来捕捉多变量依赖关系的时序信息. 如图 3 所示,第k个卷积核所有通道的特征图按时 间维度进行拼接,然后作为对应的第k个长短记忆 网络的输入;对应网络单元的隐藏输出输入到全连 接层,形成输出向量 $h^{(k)} \in \mathbb{R}^{N \times 1}$.同样地, K个相 同的网络结构总计输出 $h^{(1)},h^{(2)},...,h^{(K)}$,据此将所 有向量相加形成多变量依赖关系建模输出结果 $x_t^{(id)}$.

其中 $\hat{x}_{t}^{(se)} = \left[x_{t}^{(se,1)}, x_{t}^{(se,2)}, \dots, x_{t}^{(se,N)}\right]$ 表示自演化预训 练模型的输出向量. $\hat{x}_{t}^{(id)} = \left[x_{t}^{(id,1)}, x_{t}^{(id,2)}, \dots, x_{t}^{(id,N)}\right]$ 表示多变量依赖关系模型的输出向量. 融合方法 如下:

$$\hat{\boldsymbol{x}}_{t} = \boldsymbol{W}^{(se)} \odot \hat{\boldsymbol{x}}_{t}^{(se)} + \boldsymbol{W}^{(id)} \odot \hat{\boldsymbol{x}}_{t}^{(id)}$$
(4)

其中 $W^{(se)}$, $W^{(id)}$ 为融合权值, \odot 表示元素乘法. 这 样的两部分加权融合具有明确实际含义,增加了算 法的可解释性.例如,第*i*个变量的预测值 $\hat{x}_{t}^{(i)} = W^{(se,i)} \times \hat{x}_{t}^{(se,i)} + W^{(id,i)} \times \hat{x}_{t}^{(id,i)}$,那么对于该变 量的预测,两项取值的相对大小便直接反映了两部 分信息在预测中的占比.

完成模型构建后,为了训练模型,我们构造优 化目标对应的损失函数如下: N 维变量预训练模型



图 3 多变量时序依赖关系建模图

的损失函数记为 $\mathcal{L}^{(se)} = \left[\mathcal{L}^{(se,1)}, \mathcal{L}^{(se,2)}, \cdots, \mathcal{L}^{(se,N)}\right]$,其中 $\mathcal{L}^{(se,i)}$ 为真实值与预测值的平方误差:

$$\mathcal{L}^{(se,i)} = \left(\hat{x}_t^{(se,i)} - x_t^{(se,i)}\right)^2$$
(5)

该损失函数用于训练每个单变量预测模型 SE⁽ⁱ⁾(·),而整体的多变量预测损失函数为

$$L = \frac{1}{N} \sum_{i=1}^{N} (\hat{x}_t^{(i)} - x_t^{(i)})^2$$
(6)

整体模型可基于梯度下降法进行优化,其具体流程 如算法1所示.

算法 1. SEPNets 训练过程.

输入: 训练数据集 \mathcal{D} : { $(X_{t-T:t-1}, X_t) \in \mathcal{D}$ }

1. FOR each i = 1, 2, ... N DO

2. 评估 X⁽ⁱ⁾序列的周期性,选取历史窗口 T⁽ⁱ⁾

3. 依据 $T^{(i)}$ 构造单序列训练集 \mathcal{D}^{i}_{train}

依据 ℒ^(se,i) 在 Dⁱ_{train} 上训练得到序列 X⁽ⁱ⁾ 的预训练模型
 SE⁽ⁱ⁾(·)

5. END FOR

6. 合并单序列预训练模型,得整体预训练模型 SE(·)

7. 选取 $T = max(T^{(1)}, T^{(2)}, \dots, T^{(N)})$ 作为整体的历史窗口

8. 依据 T 构造多变量序列训练集 D_{train}

 依据 *L* 在 *D_{train}* 上训练 *SE*(·)和 CNN-LSTM 构建的网络 模型 *Nets*,得到训练完毕的 SEPNets 模型参数.
 输出: SEPNets 模型参数

4 实验验证

4.1 实验数据集

本文基于以下五个真实数据集评估多变量预测 算法的性能.

• 用电量数据集 Electricity^①

该开放数据集记录了某一区域 2012 年至 2014 年约 300个用户的用电量数据,时间间隔为 15 分钟, 每个用户有约 14 万个采样点,取值单位为 kWh.由 于同一区域内,多个用户间具有相邻性,所以对应产 生的用电量序列间具有相关性;不少工作采用此数据 集进行多变量序列预测模型的训练以及评估^{[18] [13]}. 首先,我们对数据集的每个用电量序列进行归一化 处理,以减小不同用户数据不同数量级的影响,方 式如下:

$$x = \frac{x - u_x}{1 + \sigma_x} \tag{7}$$

其中, u_x 为x的均值, σ_x 为x的标准差. 然后,

I) https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014

从 300 多个用户中随机选出 32个用户形成本文分析 的数据集;最后,在时间维度上对数据集进行划分: 将 2012-2013 年历史数据作为训练集,将 2014 年全 年数据作为测试集.

•数据库系统数据集

包括从一家 IT 公司的 Oracle DBMS(数据库管 理系统)收集的三个数据库的真实数据集.

这三个数据库支持着两种不同的服务,并分别 部署在一个国家的两个省份. DBMS 监视 70 个预定 义的关键性能指标 (KPI), 这些 KPI 涵盖了广泛的 性能指标,包括硬件运行状况(例如,CPU利用率), 软件系统性能(例如, SQL 服务响应时间)以及其 他一些与用户相关的指标(例如,活动会话数).监 测系统每 3 分钟记录一次所有 KPI 测量值,并在 相同的时间间隔内生成对应的数据库性能分数.我 们选择的三个数据库的名称分别为 JSDB、GSDB 和 GSDB2(处于隐私保护,我们隐去了公司的名称和 数据库所采集的地点等信息).我们对三个数据库 的 DBMS 采集数据,每隔 30 分钟计算高维系统 KPI 和性能分数的相关性量化值(线性相关系数), 得到70个KPI时间序列与系统性能分数序列的相关 系数时间序列. 所有 KPI 共同描述了同一系统, 所 以不同 KPI 本质上具有相关性, 对应组成 KPI 相关 性度量序列是相关多变量序列,且所有序列的取值范 围都是 [-1,1]. 同样地, 按照时间维度对数据集进行 划分,前面 80%部分构成训练集,而后面的 20%部 分构成测试集. 三个数据集的统计信息如表1所示.

表 1 DBMS 数	的据集统计信息
------------	---------

数据集名称	JSDB	GSDB	GSDB2
时间段	SepNov. 2018	Dec.2018-Jan.201	9 SepNov. 2018
服务类型	Data Analysis	Web	Data Analysis
数据库实例数	15	2	2
数据样本数	28638	12956	14452
性能分数值域	[64,96]	[0,98]	[0,100]
性能分数均值	92.8	87.3	96.6
性能分数标准差	2.42	6.76	13.1
KPI 数	70	70	70

•太阳能数据集 Solar-Energy^①

数据集为 2006 年 Alabama 州选择的 70 个光伏 电站采集的太阳能数据,时间间隔为 10 分钟.同样 地,将数据集按照时间维度进行划分,前面 80% 的数据集作为训练集,而之后的 20%的数据集作为测 试集,然后对数据集的每个光伏电站的数据进行 min-max 归一化处理,使得数据的值映射到[0,1], 归一化方式如下:

$$\boldsymbol{x} = \frac{\boldsymbol{x} - \min_{\boldsymbol{x}}}{\max_{\boldsymbol{x}} - \min_{\boldsymbol{x}}} \tag{8}$$

其中, min_x 为 x 的最小值, max_x 为 x 中的最大值.

4.2 基准算法与评估指标

本文将提出的 SEPNets 算法与下列算法进行 对比:

• Autoregression(AR)^[41]:一个基本的单变量时 间序列建模、预测模型,是具有代表性的随机过程 方法,其将当前时刻的输出变量当做历史时间段变 量的线性加权和进行预测;

•Vector Autoregression (VAR)^[14]:一个典型的统 计类多变量时间序列分析算法,它将多变量时间序 列看作多维随机过程,在单变量 AR 模型的基础上 对变量依赖关系进行线性建模;

•LSTM^[19]:一个基本的深度序列预测模型,它 用 LSTM 基本单元建模多变量时间序列并基于多 时间布历史数据进行预测;

•CNN-GRU-Input(All)-Att^[17]:一个结合卷积神 经网络 CNN(1d)和 GRU 来对多维时间序列进行预 测的深度序列模型,并利用软注意力机制(Soft Attention)来提高预测性能,文章后面将其简称为 CGIA;

•LSTNets^[18]:一个融合卷积神经网络(CNN) 和递归神经网络(RNN)的深度序列模型,能够提 取变量之间的长短期局部依赖关系,建模多变量 预测任务中复杂的时序依赖模式.

此外,我们将第 3.3 节介绍的 SE(·)的实现算法 EAR 也加入到对比算法中,并以单变量模型的预测精度作为各变量精度评估的基准.

评估多变量预测精度时,对于单变量序列预测值的平均绝对误差,第*i*个变量的平均绝对误差 记为 *MAE*⁽ⁱ⁾;所有变量的平均绝对误差的整体平均值记为 *MAE*,即

$$MAE = \frac{1}{N} \sum_{i=1}^{N} MAE^{(i)}$$
(9)

为了计算同一时刻下多变量预测值向量与真实 值向量的相似度,我们采用向量间的余弦距离, 记为 Corr.两个向量 u 和 v 的余弦距离计算公式 如下:

¹ https://www.nrel.gov/grid/solar-power-data.html

$$Corr = 1 - \frac{(\boldsymbol{u} - \overline{\boldsymbol{u}}) \cdot (\boldsymbol{v} - \overline{\boldsymbol{v}})}{\left\| (\boldsymbol{u} - \overline{\boldsymbol{u}}) \right\|_2 \left\| (\boldsymbol{v} - \overline{\boldsymbol{v}}) \right\|_2}$$
(10)

其中 \overline{u} 是u元素的均值, $x \cdot y$ 是向量x和y的点积. Corr \in [0,2],取值越小,预测值向量与真实值向量的相似度距离越小,相似度越大.为了衡量模型预测能力的好坏,我们采用了方程的确定性系数R2,公式如下所示:

$$R2 = 1 - \frac{\sum_{i} (\hat{y}_{i} - y_{i})^{2}}{\sum_{i} (\overline{y}_{i} - y_{i})^{2}}$$
(11)

其中 \hat{y}_i 为预测的值, y_i 为真实值, \overline{y}_i 为 y_i 的均值, *R*2值越大,表示预测能力越好.

为了判断各变量的预测精度与对应单变量模型的预测精度相比是否显著变差,首先,我们计算得到各变量预测值的平方误差,第*i*个变量的平方误差记为集合 $S\mathcal{E}^{(i)} = \left\{ SE^{(i,1)}, SE^{(i,2)}, ..., SE^{(i,N)} \right\}$;然后,先验地假设 $S\mathcal{E}^{(i)}$ 是对应分布的采样结果,该分布为 $x \sim \mathcal{X}^2$ (1)(因为平方误差是误差的平方,并且一般假设误差服从正态分布);最后,基于假设检验,构造统计量,对"各变量的预测精度与对应单变量模型的预测精度相比变差"进行显著性检验.所有变量中预测误差比单变量模型误差显著大的计数为B("差"),剩余变量计数为A("优"),且B + A = N表示所有变量总数.为了评估多变量预测算法与单变量模型相比预测精度的提升情况,我们采用所有变量上的误差降低率,记为*Ratio*,其计算公式如下:

$$Ratio = \frac{1}{N} \sum_{i=1}^{N} \frac{\left(MAE^{(se,i)} - MAE^{(i)}\right)}{1 + MAE^{(se,i)}} \times 100\%$$
(12)

其中 MAE^(se,i)为第 i 个变量单序列模型的预测误差; Ratio 为任意取值的实数,取值越大算法对应的预测精度提升度越大.

4.3 实验环境及参数配置

实验物理机配置为: CPU 核心 i5, 内存 32G 和 GPU GTX1070Ti, 基于 TensorFlow 并使用 Keras 深 度学习框架进行搭建.

根据 3.3 节与 3.4 节对模型的理论描述,我们通过 Keras 搭建了 SEPNets 的模型^①,表 2 展示了其架构和相应的参数.

此外,我们在训练阶段利用了 EarlyStopping 机制,即在模型训练过程中,若模型在数据集中的一部分上表现开始下降时,则停止训练,从而避免继

续训练模型导致过拟合的现象.训练过程中利用 Adam 优化器,并选择了平均绝对误差(*MAE*)作 为损失函数进行模型的训练.

表 2 SEPNets 模型架构及相应参数信息

Operation	Parameters	Inputs	Output						
Self Evolution									
Dense	units:1	x							
Dense	units:1	x	se_pred						
	CNN-LSTM								
Conv1D	filters:1; kernel_size:1	x	<i>c</i> 1						
Conv1D	filters:1; kernel_size:3	x	<i>c</i> 3						
Conv1D	filters:1; kernel_size:5	x	<i>c</i> 5						
LSTM	units:16	<i>c</i> 1	<i>r</i> 1						
LSTM	units:16	<i>c</i> 3	r3						
LSTM	units:16	<i>c</i> 5	r5						
Add		r1,r3,r5	а						
Dense	units:num_var	а	res_pred						
Multiply	units:num_var	se_pred	se						
Multiply	units:num_var	res_pred	res						
Add		se,res	y_pred						

4.4 实验结果及分析

图 4 显示了 SEPNets 模型在五个数据集上随着 训练 epoch 的平均损失变化情况.为了将所有数据 集训练的损失规定在 0 到 0.1 之间进行显示,对其 中的 Electricity 数据集的损失除以 10. 结果显示在 SEPNets 的训练阶段,模型的训练平均损失一直下 降直到收敛.



表 3 对比了五个数据集上不同算法的多变量预测平均绝对误差 MAE、相似度距离 Corr 和 R2, MAE 和 Corr 都是取值越小,表明对应算法多变量预

① https://github.com/weanl/SEPNets

测性能越好, 而 R2 取值越大表明多变量预测算法越 有效.如表所示, 在五个时间序列数据集上, 多变 量算法的预测性能总体上好于单变量算法 AR 和 EAR.在多变量算法中, VAR 和 LSTM 算法在不同 数据集上相对的预测性能表现不同, 但基本上差于 CNN-GRU-Input(All)-Att(CGIA)和 LSTNets 算法. 与其他算法相比,本文提出的算法 SEPNets 无论在 *MAE*, *Corr* 还是 R2 上基本都有较好的提升, 在多 变量预测任务中获得最好的整体预测精度.

若将 SEPNets 第一部分的基于差分特征构造的 线性自回归模型 EAR 替换成简单的自回归模型 AR,经过实验验证,采用 EAR 的 SEPNets 模型在 多变量时序预测性能上将比采用 AR 的 SEPNets 模 型平均提高 2%左右,从结果可以证明多阶差分信 息可以在 AR 的基础上提高模型预测的适用性和鲁 棒性.

表 3 多变量时序预测误差对比表

算法	Electricity		JSDB		GSDB		GSDB2			Solar-Energy					
	MAE	Corr	<i>R</i> 2	MAE	Corr	<i>R</i> 2	MAE	Corr	<i>R</i> 2	MAE	Corr	<i>R</i> 2	MAE	Corr	R2
AR	0.514	0.524	0.232	0.0619	0.126	0.703	0.0537	0.180	0.701	0.0560	0.144	0.712	0.0267	0.0718	0.985
EAR	0.485	0.449	0.340	0.0422	0.121	0.764	0.0501	0.640	0.701	0.0459	0.135	0.720	0.0215	0.0687	0.986
VAR	0.455	0.428	0.361	0.0378	0.119	0.679	0.0502	0.153	0.678	0.0386	0.132	0.718	0.0217	0.0655	0.980
LSTM	0.442	0.431	0.372	0.0374	0.118	0.622	0.0510	0.151	0.639	0.0394	0.126	0.718	0.0234	0.0648	0.982
CGIA	0.434	0.432	0.343	0.0372	0.113	0.762	0.0501	0.152	0.691	0.0391	0.124	0.717	0.0221	0.0636	0.986
LSTNets	0.435	0.435	0.343	0.0372	0.109	0.765	0.0497	0.151	0.696	0.0385	0.120	0.720	0.0220	0.0612	0.987
SEPNets	0.418	0.405	0.379	0.0371	0.107	0.764	0.0493	0.151	0.704	0.0378	0.110	0.722	0.0199	0.0563	0.989

如图 5 所示,我们选择了 VAR,LSTM 和 LSTNets 这三个典型的多变量基准算法和提出的 SEPNets 算 法,以此对比不同算法对于所有变量预测误差比单 变量模型误差显著大的计数 B("差")以及剩余变量 计数为 A("优")的结果.图 6 中的曲线显示了四个 数据集上四种多变量预测算法整体的误差降低率.

图 5(a)展示了四种多变量预测算法在 Electricity 数据集上的性能.从误差降低率来看,VAR 为 2.00%,LSTM 为 2.82%,LSTNets 为 3.25%,SEPNets 为 4.40%,这说明所有多变量算法都一定程度上降 低了预测误差,并且 SEPNets 取得了最好的降低率. 值得注意的是,Electricity数据集的总变量数为 32. SEPNets 算法对应的"差"变量数为 0、"优"变量 数为 32,这表明所有的 32 个变量的预测精度都没 有比单变量模型的预测精度显著地变差;而 VAR 分别为 13 和 19,LSTM 分别为 11 和 21,LSTNets 分别为 4 和 28,这表明现有多变量算法虽然都能 一定程度上降低预测误差,但出现了不少局部单变 量的预测精度比对应的单序列模型的预测精度差 的情况.

图 5 (b)、图 5 (c)、图 5 (d)、图 5 (e)分别 展示了其余的四个数据集结果,他们的总变量数都 为 70. 多变量预测算法中,VAR 和 LSTM 算法在不 同数据集上相对的预测误差降低率表现不同,但都 低于 LSTNets 算法.本文所提的 SEPNets 算法获得 最高的误差降低率,同时对应的"差"变量数和"优" 变量数分别为 0 和 70、1 和 69、1 和 69、0 和 70. 综 上所述, SEPNets 可以极大减少单变量的预测精度 比对应的单序列模型的预测精度差的情况,与现有 多变量时间序列预测算法相比,在保障局部变量预 测精度上有显著的性能提升.

经过实验分析,通过保障局部变量预测精度, 多变量整体预测性能也变得更好,这说明局部变量 的预测与整体变量的预测是相关的,公式(4)对自 演化预训练模型和多变量依赖关系模型两部分结果 按权重进行融合,通过让损失函数 *C* 达到最小,这 样既可以控制局部单变量的精度损失不会变大,同 时也可以控制整体预测精度损失,从而达到我们所 预期的目标.

图 6 (a)和图 6 (b)分别展示了 EAR, LSTM, LSTNets 和 SEPNets 这四种算法在五个数据集上的 训练时间与测试时间.结果表示 SEPNets 的训练时 间和测试时间最长.比如对于数据集 Electricity,在 训练过程中 EAR, LSTM 和 LSTNets 的训练时间分 别为 0.4h, 1.2h, 1.1h, 而 SEPNets 的训练时间为 4.1h; 而在测试过程中 EAR, LSTM 和 LSTNets 总的测试 时间分别为 1.25ms, 1.20ms, 1.14ms, SEPNets 的测 试时间为 2.6ms. 虽然 SEPNets 与其他多变量时间





序列预测算法在对多变量依赖关系建模上是相似的,但不同之处在于还需要先对单变量进行预训练, 所以使得训练和测试时间相比较而言要更长.

5 总结与展望

多变量时序预测是典型的时间序列分析任务, 对于辅助决策、资源配置、提前采取止损措施等方 面有重要意义,在包括电力、气象、交通、商业等 领域有广泛应用.然而,现有多变量时间序列预测 算法关注整体预测目标的优化,会带来局部变量预 测精度损失的问题.针对局部变量预测精度保障问 题,本文提出了一种基于自演化预训练的多变量时 间序列预测算法 SEPNets.首先,基于预训练策略, 设计和训练单序列预测基准模型,作为后续整体任 务的基准知识;然后,设计多层次 CNN-RNN 级联 神经网络模型,对多变量之间的相互时-空依赖关系 进行建模;最后,将已训练的单变量模型作为预训 练模块,与多变量建模模块的输出一一对应融合, 从而利用已训练的单序列模型来优化整体模型的性能,解决局部单变量的预测精度比对应的单序列模 型的预测精度差的问题.本文基于五个真实数据集 进行对比试验,结果表明,基于自演化预训练的多 变量时序预测算法,具有较好的整体预测精度,对 于局部单变量的预测精度保障性方面远优于现有的 多变量时间序列预测算法.

尽管本文的算法取得了较好的实验结果,它仍 然具有以下局限性.首先,算法需要先对单变量进 行自演化预训练,再对多变量依赖关系进行建模, 最后对自演化预训练模型和多变量依赖关系模型两 部分结果进行融合,因此在训练时间和测试时间上, 本算法相较于其他算法消耗的时间更长.其次,本 算法只能支持单步时间序列预测,不能适用于长期 的多步时间序列预测问题.

在未来的工作中,我们会在本算法基础上进一

步研究如何提升训练的和测试的效率,降低时间复杂度.其次,我们将研究在不牺牲局部变量多步预测精度的同时,设计出一个更优的多变量时间序列 多步预测算法.此外,我们拟融合多任务学习等深 度学习技术,进一步优化多变量时间序列的预测精 度和模型的泛化能力.

参考文献

- Raicharoen T, Lursinsap C, Sanguanbhokai P. Application of critical support vector machine to time series prediction// Proceedings of the 2003 International Symposium on Circuits and Systems. Bangkok, Thailand, 2003: 741-744
- [2] Cortez P, RIO M, Rocha M, et al. Multi-scale internet traffic forecasting using neural networks and time series methods. Expert Systems, 2012, 29(2): 143-155
- [3] Botezatu M M, Giurgiu I, Bogojeska J, et al. Predicting disk replacement towards reliable data centers//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA, 2016: 39-48
- [4] Box G E, Jenkins G M, Reinsel G C, et al. Time Series Analysis: Forecasting and Control.NJ, USA: John Wiley & Sons, 2015
- [5] Zhang G P. Time series forecasting using a hybrid arima and neural network model. Neurocomputing, 2003, 50: 159-175
- [6] Ahmed N K, Atiya A F, Gayar N E, et al. An empirical comparison of machine learning models for time series forecasting. Econometric Reviews, 2010, 29(5-6): 594-621
- [7] De Gooijer J G, Hyndman R J. 25 years of time series forecasting. International Journal of Forecasting, 2006, 22(3): 443-473
- [8] Jin-Fang Y, Yong-Jie Z, Dong-Feng W, et al. Time series prediction based on support vector regression. Proceedings of the CSEE, 2005, 25(017): 110-114(in Chinese)
 (杨金芳, 翟永杰, 王东风, 徐大平. 基于支持向量回归的时间序列预测. 中国电机工程学报, 2005, 25(017): 110-114)
- [9] Bontempi G, Taieb S B, Le Borgne Y A. Machine learning strategies for time series forecasting//European Business Intelligence Summer School. Brussels, Belgium: Springer, 2012: 62-77
- [10] Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, UK : MIT Press, 2016
- [11] Zhou Z H, Feng J. Deep forest: Towards An Alternative to Deep Neural Networks//Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Melbourne, Australia, 2017: 3553-3559
- [12] Ong Y J, Qiao M, Jadav D. Temporal tensor transformation network for multivariate time series prediction//International Conference on Big Data. Atlanta, USA, 2020: 1594-1603
- [13] Yin J, Rao W, Yuan M, et al. Experimental study of multivariate time series forecasting models//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing, China, 2019: 2833-2839
- [14] Lütkepohl H. New introduction to multiple time series analysis. Berlin,Germany: Springer Science & Business Media, 2005
- [15] Guo T, Xu Z, Yao X, et al. Robust online time series prediction

with recurrent neural networks//2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA). Montreal, Canada, 2016: 816825

- [16] Lipton Z C, Kale D C, Elkan C, et al. Learning to diagnose with lstm recurrent neural networks. arXiv preprint arXiv:1511.03677, 2015
- [17] Thongniran N, Jitkajornwanich K, Lawawirojwong S, et al. Combining attentional cnn and gru networks for ocean current prediction based on hf radar observations//Proceedings of the 2019 8th International Conference on Computing and Pattern Recognition. New York, USA, 2019: 440–446
- [18] Lai G, Chang W C, Yang Y, et al. Modeling long-and short-term temporal patterns with deep neural networks//The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. Ann Arbor, USA 2018: 95-104
- [19] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735-1780
- [20] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, USA, 2019: 4171-4186
- [21] Wang J, Hertzmann A, Fleet D J. Gaussian process dynamical models//Advances in Neural Information Processing Systems. Vancouver, Canada, 2006: 1441-1448
- [22] Kolarik T, Rudorfer G. Time series forecasting using neural networks. ACM SIGAPL APL Quote Quad, 1994, 25(1): 86-94
- [23] Marino D L, Amarasinghe K, Manic M. Building energy load forecasting using deep neural networks//IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society. Florence, Italy. 2016: 7046-7051
- [24] Narayanan A, Verma S, Ramadan E, et al. Deepcache: A deep learning based framework for content caching//Proceedings of the 2018 Workshop on Network Meets AI & ML. Budapest, Hungary ,2018: 48-53
- [25] Salinas D, Flunkert V, Gasthaus J, et al. Deepar: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting, 2020, 36(3): 1181-1191
- [26] Gensler A, Henze J, Sick B, et al. Deep learning for solar power forecasting—an approach using autoencoder and lstm neural networks //2016 IEEE international conference on systems, man, and cybernetics (SMC). Budapest, Hungary, 2016: 002858-002865
- [27] Cho K, Van Merriënboer B, Bahdanau D, et al. On the properties of neural machine translation: Encoder-decoder approaches// Proceedings of SSST@EMNLP 2014, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation. Doha, Qatar, 2014: 103-111
- [28] Tan Y H, Chan C S. phi-lstm: a phrase-based hierarchical lstm model for image captioning//Asian Conference on Computer Vision. Taipei, China: Springer, 2016: 101-117
- [29] Campos V, Jou B, Giró-I Nieto X, et al. Skip rnn: Learning to skip state updates in recurrent neural networks. arXiv preprint arXiv:1708.06834, 2017
- [30] Wang Y, Liu Z, Hu D, et al. Multivariate time series prediction based on optimized temporal convolutional networks with stacked autoencoders//Asian Conference on Machine Learning. Nagoya, Japan, 2019: 157-172

- [31] Qin Y, Song D, Chen H, et al. A dual-stage attention-based recurrent neural network for time series prediction// Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Melbourne, Australia, 2017: 2627-2633
- [32] Huang S, Wang D, Wu X, et al. Dsanet: Dual self-attention network for multivariate time series forecasting//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing, China, 2019: 2129-2132
- [33] Bai S, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling// 2018 International Joint Conference on Neural Networks. Rio de Janeiro, Brazil, 2018: 1-7
- [34] Hu C, Hu Y, Seo S. A deep structural model for analyzing correlated multivariate time series//2019 18th IEEE International Conference On Machine Learning And Applications. Orlando, USA. 2019: 69-74
- [35] Assaf R, Schuman n A. Explainable deep neural networks for multivariate time series predictions//Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China, 2019: 6488-6490
- [36] Goel H, Melnyk I, Banerjee A. R2n2: residual recurrent neural networks for multivariate time series forecasting. arXiv preprint arXiv:1709.03159, 2017



WAN Chen, master student. His research interests include time series analysis, and deep learning.

LI Wen-Zhong, Ph.D., professor. His research interests include distributed computing, social network analysis, and deep learning.

DING Wang-Xiang, master stu-

Background

Time series forecasting is a typical time series analysis task, which is of great significance for assisting decisionmaking, resource allocation, and taking stop loss measures in advance. It is widely used in fields including power, weather, transportation, and business. In recent years, time series forecasting algorithms based on machine learning have been a hot research field, among which multivariate time series forecasting is a challenging task. This paper focuses on the local variable forecasting accuracy of multivariate time series forecasting, that is, multivariate forecasting needs to improve the overall forecasting performance while ensuring the forecasting accuracy of local univariate forecastings. In view of the above problems, this paper analyzes the limitations of the multivariate time series forecasting algorithm. The multivariate time series

- [37] Peng-Cheng Z, Li-Yan W, Shun-Hui J, et al. Web service qos forecasting approach using multivariate time series. Journal of Software, 2019(6): 1742-1758(in Chinese)
 (张鹏程,王丽艳,吉顺慧,等.多元时间序列的 Web Service QoS 预测方法.软件学报, 2019(6): 1742-1758)
- [38] Hong-Bing O, Kang H, Hong-Ju Y. Prediction of financial time series based on lstm neural network. Chinese Journal of Management Science, 2020, 28(4): 30-38(in Chinese)
 (欧阳红兵,黄亢,闫洪举.基于 LSTM 神经网络的金融时间 序列预测.中国管理科学, 2020, 28(4): 30-38)
- [39] Shuang-Cheng W, Gao-Rui, Rui-Jie D. With super parent node bayesian network ensemble regression model for time series. Chinese Journal of Computers, 2017, 40(12): 2748-2761 (in Chinese) (王双成,高瑞,杜瑞杰. 具有超父结点时间序列贝叶斯网络
- 集成回归模型. 计算机学报, 2017, 40(12): 2748-2761) [40] Li W H, Jia-Yu C, Chen-Yang X. Prediction method of cyclic time series based on dtw similarity. Computer Science, 2019, 46(05): 157-162(in Chinese) (李文海, 程佳宇, 谢晨阳. 基于 DTW 相似判定的周期性时
- 间序列预测方法. 计算机科学, 2019, 46(05): 157-162) [41] Hamilton J D. Time Series Analysis. Volume 2. New Jersey,
- USA: Princeton New Jersey, 1994

dent. His research interests include time series analysis, and deep learning.

ZHANG Zhi-Jie, master student. His research interests include time series analysis, and deep learning.

YE Bao-Liu, Ph.D., professor. His research interests include distributed computing and parallel processing.

LU Sang-Lu, Ph.D., professor. her research interests include distributed computing and parallel processing.

forecasting algorithm SEPNets based on self-evolution pre-training is designed and implemented. Inspired by the pre-trained model, SEPNets first constructs and trains a univariate time series model as a benchmark for subsequent modeling; then expands the complex temporal dependence between the time-series convolutional network (1-dimensional convolution) and the long and short memory (LSTM) unit modeling variables Relationship; using the former as the latter pre-training model for fusion and retraining, SEPNets solves the problem of local variable forecasting accuracy for multivariate time series forecasting. The experimental results show that the SEPNets algorithm proposed in this paper ensures the forecasting accuracy of local variables to a certain extent while obtaining the relatively highest forecasting accuracy.