APPLICATION OF NEW TECHNOLOGY 新技术应用

doi:10.3969/j.issn. 1672-5166.2022.02.24

深度学习技术在中医领域中的应用

关 菀^① 马志龙^② 徐 春^{③△} 李建强^① 杨吉江^④

文章编号: 1672-5166 (2022) 02-0281-06 中图分类号: R-34; R2-03 文献标志码: A

摘 要 中医在对患者的诊断过程中能得到图像、文本等多种模态的数据,且常常根据患者特点辨证地调整处方。深度学习方法能够深度挖掘数据的内在规律,学习其多层次的特征表示,在许多领域表现出色,在中医领域的应用也取得了诸多进步。本文以不同模态的中医数据作为切入点,深入探讨了深度学习在中医领域和中西医结合领域的应用现状。中医学作为中国传统文化及世界医疗领域的重要组成部分,积累了庞杂的医案数据,且其记载形式通常为古汉语,因而深度学习在该领域的研究仍然道阻且长。

关键词 深度学习 数据挖掘 中医

Application of Deep Learning in Traditional Chinese Medicine (TCM): A Literature Review

GUAN Yu, MA Zhilong, XU Chun, LI Jiangiang, YANG Jijiang

Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

Abstract In the process of diagnosing patients, traditional Chinese medicine can get multi-modal data such as images and texts, and doctors often adjust the prescription according to the characteristics of patients. The deep learning method can extract the internal pattern of data, learn its multi-level feature representation, and achieve distinguished performances in many aspects, including the traditional Chinese medicine field. From the aspects of different modes of traditional Chinese medicine data, this paper discusses the application status of deep learning in the field of traditional Chinese medicine and integrated traditional Chinese and Western medicine. As an essential part of Chinese traditional culture and the world medical field, traditional Chinese medicine contains massive complex prescription data that recorded in ancient Chinese, therefore, the research of in-depth learning in this field is still a long way to go.

Keywords deep learning; data mining; traditional Chinese medicine

基金项目:新疆自然科学基金项目《基于电子病历的概念识别及知识图谱构建技术研究》(项目编号:2019D01A23);新疆高校科学研究 计划项目《基于多粒度临床信息表达的多模式病案发现方法研究》(项目编号:XJEDU2021Y038)

- ①北京工业大学信息学部,北京市,100124
- ②新疆财经大学信息管理学院,新疆维吾尔自治区乌鲁木齐市,830012
- ③新疆财经大学计算机科学与工程学院,新疆维吾尔自治区乌鲁木齐市,830012
- ④清华大学信息科学与技术国家实验室,北京市,100084
- 作者简介: 关菀 (1993—), 女, 博士在读; 研究方向: 机器学习; E-mail: guanyu0010@126.com
- 通信作者: 徐春 (1980—),女,博士,教授;研究方向: 计算机科学; E-mail: xu-yc@163.com

△通信作者

中国卫生信息管理

CHINESE JOURNAL OF HEALTH INFORMATICS AND MANAGEMENT

0 引言

中医典籍中记录着海量的中医 药方数据和药物数据,它们仅数 量种类庞杂,而且当面对具体的思 者时,这些方剂往往是辨证使用且 具个性化的。中医对于患者的医生 行等望诊往往也极度体 是验和个人判断^[1],这些特点。 经验和个人判断^[1],这些特点。 了中医数据信息化难以实现。 学习通过多个处理层组合形,这 学习通过多个处理层组合形, 现大型数据集的复杂结构,即 现大型数据集的特征来获取知识 [2-3], 解决数据丰富与知识贫乏之间的矛盾。近年来,深度学习方法开始越来越多地被运用到中医领域。根据其研究类型,本文将深度学习在中医方面的研究分为图像、文本、多模态以及中西医结合 4 个方向分别进行讨论。

1 深度学习在单一类型中 医数据中的应用

中医数据主要包括望诊数据 集、经典古籍、方剂数据集等。对 于单一类型中医数据,深度学习的 基本应用流程见图 1。

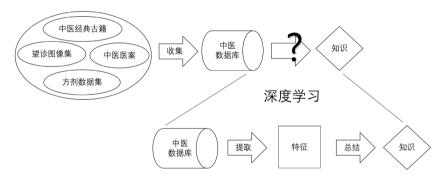


图 1 深度学习在中医领域应用的基本流程

深度学习在中医图像方面的应用主要涉及对图像的目标检测、分割、分类以及重建^[4],包括:手诊、面诊、舌诊、目诊、耳诊等望诊^[5-9],对中草药玻片的检测和分类识别^[10],根据身体三焦温度红外图像对体质的检测和分类判断等^[11-12]。

1.1.1 在望诊图像中的应用

刘婧玮[5] 采用了基于深度学

习卷积神经网络(convolutional neural network, CNN)的 ResNet34 学习模型,对其团队所采集的1706张舌象照片进行腻苔、齿痕舌等特征提取,最终得到的三分类腻苔识别模型准确率达到了90%以上。孙康宁^[6]等人分别纳入中医问诊患者的面部图像和网络公开的公众人物的面部图像,提取感兴趣区域,利用经典CNN模型对中医问诊患者的面色进行了识别与分类研究,并开发了在线的智能

面色诊断系统。李佳佳^[7]基于高血压导致的肝火亢盛在面部上的色部特征建立了图像数据库,并通过引入Xception轻量化深度学习网络,根据该望诊数据库构建了高血压的分类模型。

1.1.2 在其他图像中的应用

孙霞[10]采集了不同批次不 同产地的中草药,对其进行粉碎 后,制作了包括 4 564 张玻片图像 的数据集,并在该数据集上依次 进行了图像增强和基于压缩 - 激 发(squeeze-and-excitation, SE) 模块与特征融合的单激发多盒检 测 (single shot multibox detector, SSD), 最后完成了基于非对称卷 积网络 (asymmetric convolution net, ACNet)的多类型识别,识别 率为94.5%。周子的[11]根据阳虚、 平和等不同体质患者三焦温度各不 相同的特点,在3591张红外图像 的基础上,对ResNet、Inception、 SENet154 等模型叠加注意力模 块, 最终对于阳虚与非阳虚患者的 区分达到了89.64%的准确率。吴 云峰 [12] 利用 7 500 张肺部 CT 图 像,对其进行了图像增强、分割和 三分类操作,以区分正常、病毒性 肺炎和新冠肺炎的肺部, 其提出的 Inception-ResNet 模型最终达到了 96.1%的准确率,为新冠肺炎的基 层诊疗提供了有力的辅助,同时极 大促进了中医的客观和标准化分类 分型诊疗。

1.1.3 目前的局限和困难

医疗图像本就较其他通用自然 领域图像具有更大的获取难度,且

APPLICATION OF NEW TECHNOLOGY 新技术应用

需要该领域的专家有针对性地进行人工标注,自然也更难获取。不同区域、不同医疗机构的数据通常由于隐私性和地域性而不便共享,并且中医诊断较之西医更依赖医生的经验并缺乏流程的规范性。种种问题导致深度学习在此领域难以广泛应用,单一机构数据生成的模型也难以在不同机构间验证和推广应用。

1.2 深度学习在中医文本方面的应用

对浩如烟海的中医古籍中记录的文本自然语言(特别是古汉语)药方数据、药物数据,以及病案进行信息化、智能化的利用,离不开实用的深度学习方法。

1.2.1 实体和关系的抽取

张艺品[13] 从经典中医古籍中抽取并标注了39 100 条具有病症、方剂、中药材实体的非结构化样本数据, 并设计了1个双向长短时记忆神经网络 - 条件随机场 (BiLSTM-CRF)实体抽取模型, 其准确率 (95.47%) 超过了单独的隐马尔可夫模型 (HMM)或条件随机场 模型 (CRF)。李焕 [14] 首先通过 HTML 网页爬取到中医 方剂主治文本数据集,在数据集上构建了1个BERT-BiLSTM-CRF术语识别模型,然后在此基础上引入主 动学习方法,显著减少了BERT-BILSTM-CRF模型对 有标注数据的需求和依赖。高甦[15]以《黄帝内经》为 语料,并融合 701 本中医典籍,形成了 3.84G 的预训 练语料,与条件随机场(CRF)、长短时记忆神经网络 (LSTM)、双向长短词记忆模型(BiLSTM)的单独或组 合结构相比, 其经过预训练的 BiLSTM-CRF 架构能够 对中医病理、认识方法等5个实体达到最好的识别效果 (85.32%)

1.2.2 中医自动诊断架构

任雪^[16] 构建了一种基于主动集成学习的儿科肺炎咳嗽中医诊断架构,纳入了支持向量机(support vector machine, SVM)、决策树、贝叶斯、卷积神经网络(CNN)等机器学习和深度学习方法,对儿童肺炎咳嗽的症候进行了准确率可观的分型。Liang^[17] 将 1个包含多种疾病的纯文本数据集和 1 个原发性高血压的结构化数据集,随机划分为进行无监督学习的训练集和进行有监督学习的测试集,然后基于深度置信网

络 (deep belief networks, DBN)进行无监督训练以获 取特征表达,基于 DBN+SVM 方法以完成有监督学 习。DBN+SVM与任一单一模型相比表现更好[17]。 曹彬 [18] 将基于 CNN、SVM、LSTM 构建的检索模型与 基于 seg2seg (sequence-to-sequence) 构建的生成 回复模型结合,利用爬虫获取的语料库,搭建了1个 可学习的中医领域的在线智能问答系统, 为中医在线问 诊提供了便利。王宁^[19] 基于中医电子病历中的结构化 和非结构化数据,分别利用深度学习和自然语言处理等 方法,实现了中医证型的自动预测和中医处方的自动生 成,并在此基础上对患者的用药情况和最终的治疗效果 进行了基于关联规则的多类相关性学习,得到了较好的 效果。郑子强[20] 依托慢性肾脏病现有的中医诊疗路径, 创新地提出一种结合局部信息和上下文信息的命名实体 识别方法,构建知识图谱,并在此基础之上设计完成了 1个慢性肾病中医知识图谱学习与推理的原型系统,以 实现数据可视化和用户管理等多项功能。

1.2.3 目前的局限和困难

中文文本较之英文文本在实体识别和关系抽取等自然语言处理方面本就具有更大的困难度,而中医古籍文言文的形式因其历史性,更是具有多歧义、多通假、难以结构化等特征。在临床中,医生又会根据个人经验为患者辨证地调整药方。这些特点使得深度学习模型难以推广,而其根据中医文本得到的自动组方尚不具有能真正应用到临床的准确率。

2 深度学习应用于多模态中医数据

针对多模态中医数据的研究较多。多模态数据能够 聚合多源信息,取长补短,使整体特征表达更完善。

胡杨^[21] 根据舌诊的图像数据和问诊的文本数据,不仅分别提出了针对性的语速增强网络和中医组方自动生成网络,而且设计了 1 个基于图的表现优异的语义 - 视觉纠缠网络。王科文^[22] 采集门诊患者的舌诊图像和医生处方,分别构建了舌苔图像集和药材集。首先使用主成分分析和自编码器模型,提取舌苔数据集的特征,并达到将图像向量化的目的,然后合并该舌特征向量和中医特征向量,最后在 1 个全连接神经网络的基础上,

●中国卫生信息管理

CHINESE JOURNAL OF HEALTH INFORMATICS AND MANAGEMENT

根据不同舌苔特征与药材的关联度获得了 1 个中药推荐系统 ^[22]。刘超等人 ^[23] 对 814 位患有冠状动脉临界病变的患者进行了调查研究,采集了他们的冠状动脉造影结果、脉象、舌象、体征等 110 个中医四诊临床特征。其后,结合造影结果数据和临床特征,利用多层前馈神经网络,得到症候要素组合,并最终搭建出 1 个辨证的中医诊断模型,为冠心病的早期诊断提供了技术支持 ^[23]。石艳敏 ^[24] 根据哮喘病患者大鱼际掌纹的阴阳特征表现,引入了改进的 YOLO(you only look once)模型对手掌照片进行识别,达到了 92.5% 的准确率。同时,利用优化的深度置信网络,从文本病案数据中提取出与哮喘病发作高度相关的 20 个主症状,辅助医生对哮喘病的临床诊断 ^[24]。

除了科研领域,深度学习在中医药大数据分析现实应用方面也有很多成功案例。XHT 中医辅助诊疗系统^[25] 是基于 RYY 多模态大数据,利用深度学习等人工智能技术开发出的"云端中医大脑",能够根据记录了望闻问切情况的电子病历和其他数据来辅助医生开方。以中医特殊深度学习算法和海量多模态中医药大数据为基础的"WZ 国医云脑"^[26],具备了感知、连接、语言、决策等能力,能够辅助医生对患者进行诊断。此外,还有赋能基层中医馆的医共体智慧医疗建设^[27],DJ 中医的医联体智慧中医云平台^[28]等。

3 深度学习应用于中西医结合

中西医作为两个不同的医疗体系,二者结合,能够 各取所长,相互交流促进。将深度学习应用于中西医结 合,能够带来更多样、更全面的信息和更好的医疗效果。

3.1 中西医结合带来新发现

Wei^[29]等人以冠心病流行病学为背景,对中医证候与西医蛋白质组之间的关联进行了探讨。他们发现,血瘀证与 10 种炎症因子蛋白有显著相关性,并基于 SVM 方法,识别出 14 个包含证候和蛋白质的模式,且分类准确率大于 90%,这表明这些模式可以作为中医证候与蛋白质组关联的证据,且结果可信^[29]。Zhang^[30] 认为,现代西医技术与中医理论相结合,能够更好地了解

患者的症候。例如: Zhang^[30] 通过内镜探视对胃黏膜下血管网进行观察,发现萎缩性胃炎与血瘀证密切相关,而后利用深度学习方法,结合云计算技术弥补资源不足的缺点,以更好地分析医学图像,诊断脾胃疾病。

3.2 中西医结合预测诊断模型

Chen^[31]等人获取了 1 个含有 85 个病例的急性缺血性卒中数据集。由于数据集较小,他们采用了 3 种机器学习方法(支持向量机、随机森林、梯度提升决策树),在现有危险因素的基础上,结合中医体质和证候,构建了 1 个基于中西医结合的精度高、拟合良好的多维早期预测模型,从而提高预测效率,为脑卒中患者卒中后认知障碍(PSCI)的临床评价提供参考。Wang^[32]等人设计了 1 种基于数据挖掘的双系统中西医临床诊断模型。通过改进的 Apriori 算法,对由信息采集设备采集的中医诊断数据进行处理,利用动量优化 BP 神经网络对用户提交的西医诊断数据进行分析,最终实现对用户数据的定量分析和病理诊断 ^[32]。

但是,由于中西医所属两个完全不同的体系结构, 其融合贯通和之间广泛的关联结合仍然有待深入的研究 和发掘。

4 结语

中医学作为一个古老的医疗体系,已有几千年的历史,积累了浩如烟海的医案和病案。而当代中医除了以典籍为指导外,也积累了众多中医药图像和电子病历,并开始与西医、各种现代技术结合。深度学习作为数据挖掘、机器学习的重要分支,能够处理更大的数据集、更复杂的结构,学习更高级的特征,探索更未知的关系,减少对人工标注的依赖,是现代中医信息化、智能化过程中绕不开的必经之路。目前,中医领域的智能化发展仍然受制于数据集的大小和规范性,受制于古汉语文字的多义、通假等特性。同时,具有多模态数据的患者数量比有单一数据来源的患者数量更少,不同模态间的数据融合仍然具有相当的难度,跨模态学习模型的构建不易实现,而且只通过单模态数据进行辅助诊断也难以达到较好的效果。深度学习离不开大规模高质量的数

APPLICATION OF NEW TECHNOLOGY 新技术应用

据集,因此目前深度学习在中医领域的应用仍然不如其他医疗领域广泛。未来,如能基于中医特色,对中医领域多源多模态的数据采集制定更广泛通用的标准,对中医知识和经验进行更为规范化的结构化整理,针对中医领域下的各类分支方向构建更大规模的公开数据集,同时引入结合区块链技术的联邦学习,以促进脱敏数据的跨地区跨机构知识共用,那么深度学习将在中医领域得到较大的发展。■

参考文献

- TANG J L, LIU B Y, MA K W. Traditional Chinese medicine
 Lancet, 2008, 372(9654): 1938-1940.
- [2] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521: 436-444.
- [3] DENG L, YU D. Deep learning: methods and applications [J]. Foundations and Trends in Signal Processing, 2014, 7(3): 197-387.
- [4] 赵紫娟,赵涓涓,赵琛琦,等.图像智能处理方法在中医中的应用与挑战 [J/OL].太原理工大学学报:1-20[2021-11-09]. http://kns.cnki.net/kcms/detail/14.1220.N.20211018.1350.002.html.
- [5] 刘婧玮.基于卷积神经网络的中医舌象辨识人工智能方法学研究[D].北京:北京中医药大学,2020.
- [6] 孙康宁, 孙琦, 李新霞, 等. 基于卷积神经网络的中医面色提取识别研究[J]. 中华中医药杂志, 2021, 36(7): 4286-4290.
- [7] 李佳佳.基于深度学习的高血压病肝火亢盛证面部色诊研究 [D].北京:中国中医科学院,2020.
- [8] 宋超.基于深度迁移学习的舌象特征分类方法研究 [D]. 南京:南京财经大学,2020.
- [9] 马龙祥,杨浩,宋婷婷,等.基于高分辨率特征的舌象分割算法研究[J].计算机工程,2020,46(10):248-252.
- [10] 孙霞. 基于深度学习的中草药粉末显微特征研究 [D]. 北京:北方工业大学, 2020.
- [11] 周子昀.基于深度学习的中医红外图像体质分类研究 [D]. 成都:电子科技大学,2021.
- [12] 吴云峰 . 基于深度学习的肺炎医学 CT 图像分类算法研究 [D]. 福州 : 福建中医药大学 , 2021.
- [13] 张艺品,关贝,吕荫润,等.深度学习基础上的中医实体

- 抽取方法研究 [J]. 医学信息学杂志, 2019, 40(2): 58-63.
- [14] 李焕.基于深度学习与主动学习的中医术语识别研究 [D]. 北京:北京工业大学, 2019.
- [15] 高甦,金佩,张德政.基于深度学习的中医典籍命名实体识别研究[J].情报工程,2019,5(1):113-123.
- [16] 任雪, 郭艳. 基于主动集成学习的中医智能诊断模型及构建方法 [J]. 中国循证医学杂志, 2019, 19(9): 1118-1123.
- [17] LIANG Z, LIU J, OU A, et al. Deep generative learning for automated EHR diagnosis of traditional Chinese medicine[J]. Computer Methods and Programs in Biomedicine, 2018, 174: 17-23.
- [18] 曹彬 . 基于深度学习的智能中医问答系统的设计 [D]. 广州 : 广东工业大学 , 2019.
- [19] 王宁.基于电子病历的中医临床决策支持研究 [D]. 合肥:中国科学技术大学, 2021.
- [20] 郑子强.面向慢性肾脏病中医医案的知识图谱学习与推理研究 [D]. 成都:电子科技大学, 2020.
- [21] 胡杨.面向智能中医辅助诊疗的多注意力和知识辅助神经 网络设计研究 [D].广州:华南理工大学, 2020.
- [22] 王科文 . 基于深度神经网络的中医药材推荐 [D]. 广州 : 华南理工大学 , 2018.
- [23] 刘超,高嘉良,董艳,等.基于BP神经网络的冠状动脉临界病变患者证候要素及其常见组合中医辨证诊断模型研究[J].中国中医药信息杂志,2021,28(3):104-110.
- [24] 石艳敏 . 基于大数据的中医哮喘病诊疗智能化方法研究 [D]. 青岛 : 青岛科技大学 , 2019.
- [25] 微医云. 悬壶台中医辅助诊疗系统 [EB/OL]. (2017-11-08) [2022-01-14]. http://cloud. guahao. cn/product/1lE1MZzFa.
- [26] 智医疗. "乌镇国医云脑" 发布, 拓开中医 AI 融合之路 [EB/OL]. (2017-12-08)[2022-01-14]. http://www.mhealthchina.org/17692. html.
- [27] 微医. 微医华佗云荣获中医药天池大赛冠军,推动中医数字化发展[EB/OL]. (2020-12-21)[2022-01-14]. https://tech.china.com/article/20201221/122020_679555. html.
- [28] 大经中医 . 中医特色的互联网医院整体解决方案 [EB/OL]. (2018-08-21)[2022-01-14]. http://dajingtcm.com/scheme/2.
- [29] WEI W, ZHAO H, CHEN J, et al. Bridge the gap between syndrome in Traditional Chinese Medicine and proteome in western medicine by unsupervised pattern discovery

(下转第292页)



●中国卫生信息管理

CHINESE JOURNAL OF HEALTH INFORMATICS AND MANAGEMENT

报,2018,9(24):6551-6556.

- [10] 孙佳斐,隋川石. 网络食品交易第三方平台食品安全法律风险防控研究[J]. 法制与社会, 2016(15): 106-107.
- [11] 陈悦,陈超美,刘则渊,等. CiteSpace 知识图谱的方法论功能 [J]. 科学学研究, 2015, 33(2): 242-253.
- [12] NIK V M. Making energy simulation easier for future climate

 Synthesizing typical and extreme weather data sets out of regional climate models (RCMs) [J]. Applied Energy, 2016, 177(9): 204-226.
- [13] APTE J S, MESSIER K P, GANI S, et al. High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data[J]. Environmental science & technology, 2017, 51(12): 6999-7008.
- [14] CHEN X, SHAO S, TIAN Z, et al. Impacts of air pollution and its spatial spillover effect on public health based on China's big data sample[J]. Journal of Cleaner Production, 2017, 142(P2): 915-925.
- [15] LIAKOS K G, BUSATO P, MOSHOU D, et al. Machine Learning in Agriculture: A Review[J]. Sensors, 2018, 18(8): 2674.
- [16] KISSLING W D, AHUMADA J A, BOWSER A, et al. Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale[J]. Biological Reviews, 2018, 93(1): 600-625.
- [17] ENGEMANN K, ENQUIST B J, SANDEL B, et al. Limited sampling hampers "big data" estimation of species richness in a tropical biodiversity hotspot[J]. Ecology and evolution,

2015, 5(3): 807-820.

- [18] AHEARN M C, ARMBRUSTER W, YOUNG R. Big Data's Potential to Improve Food Supply Chain Environmental Sustainability and Food Safety[J]. International Food And Agribusiness Managemant Review, 2016, 19(3): 155-172.
- [19] OZDEMIR V, KOLKER E. Precision Nutrition 4. 0: A Big Data and Ethics Foresight Analysis-Convergence of Agrigenomics, Nutrigenomics, Nutriproteomics, and Nutrimetabolomics [J]. OMICS, 2016, 20(2): 69-75.
- [20] 刘张,千家乐,杜云艳,等.基于多源时空大数据的区际 迁徙人群多层次空间分布估算模型——以COVID-19 疫 情期间自武汉迁出人群为例 [J]. 地球信息科学学报,2020, 22(2):147-160.
- [21] VISCONTI P, Roberto DE F R, VELAZQUEZ R, et al.

 Development of Sensors-Based Agri-Food Traceability

 System Remotely Managed by a Software Platform for

 Optimized Farm Management[J]. Sensors, 2020, 20(13): 3632.
- [22] 李雯,姜仁贵,解建仓,等 . 基于文献计量学的城市洪涝灾害研究可视化知识图谱分析 [J]. 西安理工大学学报,2020, 36(4): 523-529.
- [23] 国务院. 国务院关于印发促进大数据发展行动纲要的通知(国发〔2015〕50号)[EB/OL]. (2015-09-05) [2020-12-30]. http://www.gov.cn/zhengce/content/2015-09/05/content_10137. html.

[收稿日期: 2021-04-09 修回日期: 2022-03-13] (編辑: 耿俊超)

(上接第285页)

- algorithm[C]//IEEE International Conference on Networking, 2008.
- [30] ZHANG Q, BAI C, CHEN Z, et al. Deep learning models for diagnosing spleen and stomach diseases in smart Chinese medicine with cloud computing[J]. Concurrency Computat Pract Exper, 2019(4): e5252.
- [31] CHEN X, WEI C, WU W, et al. Based on machine learning algorithm: Construction of an early prediction model of integrated traditional Chinese and Western medicine for
- cognitive impairment after ischemic stroke[C]//2020 5th International Conference on Universal Village (UV), 2020.
- [32] WANG Z, LI J, ZHANG L, et al. A prediction model of clinical diagnosis by the combination of traditional Chinese and Western medicine based on data mining[C]//2021 International Conference on Intelligent Transportation, Big Data and Smart City (ICITBS), 2021.

[收稿日期: 2021-11-09 修回日期: 2022-03-12] (編辑: 耿俊超)