▶ 现代计算机

XIANDAI JISUANJI 第29卷第2期(总第770期) 半月刊(1984年创刊) 2023年1月25日出版

主管单	位	中山大学		
主办单	位	广州中山大学出版社有限公司		
出版单	位	广东现代计算机杂志社有限公司		
发	行	广东省报刊发行局(全国公开发行)		
印	刷	广州一龙印刷有限公司		
社	长	黄少伟		
主	编	石玉珍		
编	委	邹岚萍 熊锡源 李 文 石玉珍 梁嘉5		
地	址	广州市海珠区新港西路135号		
		中山大学内(510275)		
电	话	020-84112089(编辑部)		
网	址	www.moderncomputer.cn		
电子邮	电子邮箱 tougao@moderncomputer.cn			

ISSN 1007-1423	邮发代码: 46-121	
CN 44-1415/TP	定价: 30.00元	邮履





2023年1月 第29卷 第 2 期 (总第770期)

第29卷 第2期 (总第770期

M

现代计算机

💮 中山大學出版社 主办





MOD

ISSN 1007-1423 CN 44-1415/TP

n

中国期刊数据库CNKI全文收录期刊 中国学术期刊(光盘版)收录期刊 中文科技期刊数据库全文收录期刊 中国核心期刊(遴选)数据库收录期刊 中国学术期刊综合评价数据库收录期刊

- ◆研究与开发: 计算机发展和软、硬件开发的理论研究
- ◆ 图形图像: 重点为与图形图像相关的理论及实践研究
- ◆ 开发案例: 基于某方面的计算机开发案例研究与分析
- ◆ 实践与经验: 计算机应用的实例及心得

#### 版权声明

1. 本刊版权属于杂志社所有,其他报刊或网站如需转载,须经本刊 同意,注明转载自本刊并付作者稿酬。

2. 本刊来稿恕不退还,请自留底稿。请勿一稿多投。来稿文责自负, 严禁抄袭。对侵犯他人版权或其他权利的稿件,本刊概不承担连带责任。

3. 对所投稿件,本刊编辑有权根据刊物的需要进行删改或调整。

4. 凡是刊登在本刊的稿件,即表示作者同意稿件在《现代计算机》 网站、中国期刊数据库CNKI、中国学术期刊(光盘版)、中文科技期刊 数据库、中国科技期刊(遴选)数据库、中国学术期刊综合评价数据库 等媒体发布。

### 现代计算机

XIANDAI JISUANJI

第29卷 第2期(总第770期)2023年1月25日出版

## 目 次

### 研究与开发

一种自适应无人机集群网络恢复方法 石运阳,华 翔,张金金	(1)
基于时间序列的民用运输航空器碳排放预测研究 向小军,杨志晗,赵赶超	(14)
一种改进 DeepLabV3+的岩屑图像语义分割算法 罗崇兴,师明元,王正勇,滕奇志	(23)
融合多尺度卷积的端到端宫颈细胞分割 王文涛, 王嘉鑫, 张 根, 陈大江	(32)
一种基于YOLOv5算法的布匹瑕疵检测系统 邓 景,李成海,丁兆栋,杜光辉,陆 可	(41)
基于混合遗传算法的成品油二次配送优化	(50)
基于时间跨度注意力机制的多变量时间序列预测方法 李文豪,严华	(56)
基于 Pareto 支配的高维多目标优化算法的分析与研究 操心慧, 许丽娟	(62)
基于词向量与TextRank的政策文本关键词汇抽取方法研究	
······························ 李 晨,赵燕清,于俊凤,张铭君, DMYTRO LANDE	(68)
基于 ResNet-18 网络的城市生活垃圾识别方法研究	
·································金张根,曹 杨,于红绯,孙才华,刘 克	(73)
基于光照估计滤波的太赫兹图像融合研究	(78)

### 实践与经验

基于拟合的混响室莱斯K因子预测及信道重建	张雪莹,	赵翔	(82)
Web应用中间件性能测试系统设计与验证	何冬辉,	杨攀飞	(88)
基于 SPSS 的管制工作负荷与航班架次关系分析 ······	郭东鑫,	李科扬	(95)
基于胶囊模型的短文本细粒度情感分类		邵辉	(99)

#### 开发案例

一种面向铁路领域在线客服内容违规和风险的应急管理方法	-				
	•••••	皮尔达伟斯·	巴吐尔,	刘捷	(103)
基于区块链和"时间银行"的互助式服务平台的设计和开发	~				
	·向佳	欣,王宏杰,	梁桂萍,	赖沛鑫	(110)
私有云平台数据云上云下备份体系设计					
孙建刚, 高 颖,	,杨庆	甫,常雨竹,	董耀众,	李伟良	(116)

# Modern Computer

(Vol. 29, No. 2; Jan. 25, 2023)

### CONTENTS

## Research and Development

An adaptive network restoration method for UAV cluster	(1)
Forecast study on carbon emissions of civil transport aircraft based on time series	(14)
An improved cuttings image semantic segmentation algorithm based on DeepLabV3+ ·····	(23)
End-to-end cervical cell segmentation with multi-scale convolution	(32)
A fabric defect detection system based on YOLOv5 algorithm	(41)
Optimization of vehicle routing problem of petroleum products distribution based on	
hybrid genetic algorithm	(50)
A time span attention based deep neural network for multivariate time series forecasting	(56)
Analysis and research of many-objective optimization algorithm based on Pareto-domination	(62)
Research on keyword extraction of policy using word vector and TextRank	(68)
Research on municipal solid waste identification method based on ResNet-18 network	(73)
Terahertz image fusion based on illumination estimation filtering	(78)

### Practice and Experience

Rician <i>K</i> -factor prediction and channel reconstruction of reverberation chamber based on fitting	(82)
Performance measurement system of Web application servers: design and verification	(88)
Analysis of the relationship between controller workload and flight sorties based on SPSS	(95)
Fine-grained sentiment classification of short text based on capsule model	(99)

### **Development Solution**

An emergency management method for online customer service content violations and risks			
in the railway field ·····	(103)		
Design and development of mutual assistance service platform based on			
blockchain and "time bank"	(110)		
The private cloud platform data cloud backup system design	(116)		

#### 研究与开发

文章编号:1007-1423(2023)02-0001-13

DOI: 10.3969/j.issn.1007-1423.2023.02.001

### 一种自适应无人机集群网络恢复方法

石运阳,华 翔\*,张金金

(西安工业大学电子信息工程学院,西安 710021)

摘要:为解决无人机集群网络因节点失效造成网络性能和网络可靠性下降的问题,提出了一种自适应 无人机集群网络恢复方法。通过信号稳定性和链路负载度对集群内部的通信代价进行评估,并结合图论构 建无人机集群网络通信模型。同时,通过鲸鱼算法对灰色滚动预测进行优化,利用优化后的预测模型对通 信代价进行提前评估。算法根据通信代价矩阵的预测值找到待恢复节点之间的最短路径,并利用负载均衡 算法对恢复后的路径进行优化,实现无人机网络的性能恢复。仿真结果表明,该方法在无人机集群网络失 效时可以快速进行拓扑重构,恢复通信节点之间的链路连接,并使网络具有良好的生存性和抗毁性。

关键词: 无人机集群网络; 通信代价; 改进预测算法; 拓扑修复

#### 0 引言

无人机集群是由多架无人机组成的一个完整的有机系统,集群内部的无人机通过无线自组织网络建立连接,构建成一个整体的作战单元,从而具备在复杂多变环境中执行危险任务的能力。无人机集群在执行任务过程中,无人机个体的移动会导致网络中的链路频繁改变,恶劣环境的影响也会使得部分节点失效<sup>[1]</sup>,这些都会导致集群网络的间歇连接和动态拓扑,使得网络结构受损,降低无人机集群网络的拖进通信<sup>[2]</sup>。因此,无人机集群网络的拓扑修复应当是在网络出现结构性损伤后,自适应调整局部节点之间的链路,以保证网络整体的连通性。

目前,通信节点自适应进行链路重选的方法 主要有两种,分别是基于网络结构特征的链路重 选机制以及基于网络负载程度的链路重选机制。

基于网络结构特征的链路重选机制是利用

几何方法对集群网络拓扑结构进行分析。马学 森等<sup>[3]</sup>针对无线自组网最优传输路径问题,提 出基于蚁群算法的路径寻优和恢复算法,该算 法有效降低了节点的能量消耗,在出现节点死 亡时可以对最优路径进行快速恢复。姚玉坤 等<sup>[4]</sup>针对自组网网关节点失效问题,通过马尔 可夫链路状态预测模型,提出了基于无人机--地 面控制站链路状态预测的网关选择算法。李玉 龙等<sup>[5]</sup>针对贪婪路由协议不能准确反映节点位 置的问题,提出了一种基于移动预测和链路保 持时间的路由协议MP-GPSR,综合考虑节点的 移动位置和链路保持时间来选择下一跳,避绕 路由空洞,降低了传统边界转发的路径冗余。

基于网络负载的链路重选机制是确保无人 机节点能有效传输其负载对象。对于随机路径 移动模型,康巧琴等<sup>[6]</sup>提出了基于效用值转发 的路由快速恢复算法。综合利用网络时延、节 点效用值和下一跳数三个指标得到最优的下一

**收稿日期:** 2022-12-04 修稿日期: 2023-01-10

基金项目:陕西省重点研发计划项目(2023-YBGY-227);陕西省自然科学基础研究计划资助项目(2023-JC-QN-0705); 西安市科技计划项目(2022JH-RYFW-0138);碑林区科技计划项目(GX2216)

作者简介:石运阳(1998—),男,江苏南通人,硕士研究生,主要研究方向为无人机网络修复;\*通信作者:华翔(1979—), 女,陕西西安人,教授,主要研究方向为短距离通信,E-mail:huaxiang@xatu.edu.cn;张金金(1992—),女,陕西西安人,讲师, 主要研究方向为集群无人系统通信关键技术、推荐系统

跳节点。该算法平均跳数少、成功概率高、平 均时延低。肖军弼等<sup>[7]</sup>将 SDN 网络架构引入故 障恢复策略,根据域间域内跳数、带宽总量、 已使用的链路带宽进行通信代价权值评价,生 成域间相交最小的备用路径。孙明杰等<sup>[8]</sup>提出 了基于蚁群优化的路由算法,该算法将蚁群信 息素与路由算法相结合,大大减少了路由拥塞 和链路断路的情况。

然而,相较于传统移动自组网,无人机集 群自组织网络节点之间的相对速度更大,链路 质量变化也更加频繁<sup>[9]</sup>,因此对于无人机集群 网络恢复算法的设计应当着重考虑无人机的移 动性和拓扑时变性。传统的网络恢复算法可以 可靠地寻找下一跳备选节点,但是在无人机领 域会在时效性方面存在一定的不足。因此如何 设计出适用于无人机高速动态拓扑的自组织网 络恢复算法是目前无人机集群网络研究领域的 热点和难题。

针对以上问题,本文提出一种自适应无人 机集群网络恢复算法。首先,每个无人机节点 利用鲸鱼算法优化后的灰色滚动模型对节点之 间的通信代价进行预测。然后,当集群中的部 分节点失效后,待恢复节点通过最短路径算法 和负载均衡算法根据预测结果找到节点之间的 最短路径,实现集群网络的自适应损伤恢复。 最后,对算法进行仿真分析,验证恢复算法的 时效性、恢复后网络的有效性和抗毁性。

#### 1 无人机集群网络拓扑映射模型

无人机集群编队在飞行过程中往往是整体 朝着一个既定目标进行移动<sup>[10]</sup>。本文基于复杂 网络理论对无人机集群网络进行建模,采用图 论方法对其进行抽象,将无人机的节点与链路 映射到复杂网络模型中,考虑了无人机之间信 号的功率强度、相对速度以及节点负载度等因 素,将无人机的连通性恢复问题转化为复杂网 络的边重连问题。

#### 1.1 集群网络通信代价权值指标

考虑无人机集群网络拓扑时变的特点,结 合无人机节点属性参量因子,由单位时间内无 人机之间的通信关系建立无人机集群通信网络 结构模型。将整个无人机通信网络抽象为图 *G*(*V*,*E*,*W*),其中,*V*(*G*)表示所有节点的集合, *E*(*G*)表示所有边的集合,|*V*|代表节点的个数, |*E*|表示通信链路的数量,根据单位时间内无人 机之间的通信拓扑结构,可以将无人机的链路 映射到图*G*中的连边*e*(*i*,*j*)。 ₩表示无人机之 间通信代价的权重矩阵,*w*<sub>i</sub>表示无人机编号*i*和 *j*的通信代价权值指标,通过归一化后的信号稳 定性指标以及链路负载度指标参数确定。

信号稳定性与节点之间的功率强度和相对速 度有关。功率强度指标主要通过无人机相互之间 Hello消息的信号强度确定。对于节点*i*,定义其 接收到节点*j*的信号功率大小为*P<sub>ij</sub>*,根据自由空 间传播模型,节点间功率强度计算方法为

$$P_{ij} = \frac{P_s G_s G_r \lambda^2}{(4\pi)^2 d^2 L} \tag{1}$$

式中: *P*<sub>s</sub>为无人机节点的额定天线功率; *C*<sub>s</sub>、*C*<sub>r</sub>为无人机天线的接收增益参数和发射增益参数; *L*为信道传输路径损耗的参数; λ为波长,以上变量均为定值。



图 1 无人机节点之间的相对运动

无人机之间的相对速度可以根据多普勒效 应计算得到。设f为无人机节点发射功率的原频 率,f'为节点j接收到节点i的载波频率,可知  $\Delta t = \frac{1}{f}$ 。由多普勒效应,无论节点j相对于节点i远离或是接近,节点i和j的相对速度为

$$v_{ij} = \frac{c \left| f - f' \right|}{f \cos \angle abc} \tag{2}$$

结合余弦定理和公式(1)中的自由空间传播 模型,利用节点之间的功率强度对公式(2)中的 三角函数进行替换,可以推导出节点*i*和*j*的相 对速度:

$$v_{ij} = \frac{fR_m \sqrt{2c \left| f - f' \right| \frac{\sqrt{P_{jb} P_{\min}}}{f^2 R_m} - (P_{jb} + P_{jc})}}{\sqrt{P_{\min}}} (3)$$

根据公式(1)和公式(3)分别得到节点*i*和*j*的功率强度和相对速度后,两个节点之间的信号稳定度*LS<sub>i</sub>*就可以根据以下公式得到:

$$LS_{ij} = \lambda \frac{P_{ij}(d) - \max\left(\min\left(P_{ij}\right), P_{ih}\right)}{P_s - \max\left(\min\left(P_{ij}\right), P_{ih}\right)} + \varphi \frac{v_{ij\max}}{v_{\max}}$$
(4)

式中: i = j是对应链路两端的发送节点和接收 节点。 $P_{ij}$ 是节点 $i \pi j$ 间的信号功率强度, $v_{ij}$ 是 节点 $i \pi j$ 的相对移动速度, $P_s \pi v_{max}$ 分别表示无 人机的发射功率与节点最大速度。 $\lambda \pi \varphi$ 分别为 功率和速度的权重因子,  $\Xi \lambda + \varphi = 1$ 。

*P*<sub>th</sub>为设定的对应链路最小阈值,其对应于 处在无人机通信边缘且相对距离恰好为最大通 信半径减去通信时延*t*<sup>de\_min</sup>的邻居节点。设置链 路最小阈值可以避免节点选择生存时间非常短 的链路。

$$P_{th} = \frac{P_s G_s G_r \lambda^2}{(4\pi)^2 (R_m - 2v_{\max} t^{de_{\min}})^2 L}$$
(5)

链路负载度指标映射到边权重时,定义为 两端节点的负载均值:

$$LC_{ij} = \frac{M_i + M_j}{2} \tag{6}$$

通信代价权值代表两架无人机之间通信困 难程度,相对移动速度越快、信号强度越小、 链路负载程度越高,无人机之间的通信代价就 越大。根据公式(4)和公式(6),信号稳定性越 大、链路负载越小,通信代价越小,因此定义 通信代价权值*w*<sub>ij</sub>为链路负载度指标与信号稳定 性指标的比值:

$$w_{ij} = \begin{cases} \frac{k_{LC} * LC_{ij}}{k_{LS} * LS_{ij}} & P_{ij} > P_{th} \\ inf & \text{else} \end{cases}$$
(7)

其中: k<sub>Lc</sub>和k<sub>Ls</sub>分别为链路负载度指标和信号稳 定性指标的权值。可以看出,当信号稳定性越 高、节点负载度越小时,节点*i*与*j*之间的通信 代价越小;当节点信号强度低于阈值强度后, 通信代价为无限大。

#### 1.2 无人机集群初始网络拓扑结构

无人机集群系统一般采取密集编队模式, 每个通信个体与最近几个邻居进行通信。考虑 一个包含n架无人机的无人机集群,假设所有的 无人机具有相同的结构和运动能力,无人机的 通信半径为R<sub>m</sub>。设定无人机集群中包含一架领 航无人机以及n-1架跟随无人机,运动方式为 参考群组移动模型。领航无人机在接收到上位 机的命令后,会沿着任务信息所规划的航迹进行 飞行,跟随无人机根据邻居节点自动调整位置, 保证每个无人机的运动轨迹与集群整体一致,因 此领航无人机无需与集群中所有无人机实时进行 数据同步,任意一个无人机都可以成为领航节 点,每个无人机只需要与邻居无人机保持相似的 运动轨迹,即可实现集群运动的同步。

设定无人机集群整体以速度 $v_0$ 朝着一个既 定目标飞行,在飞行过程中,由于自然因素和 无人机自身性能因素,每架无人机速度会在一 定范围内波动,定义无人机i在时刻t的速度为  $v_i(t)$ ,在网络拓扑模型G(V, E, W)中,任意节点 i的速度为

$$v_i(t) = v_0 + v_{rand}, \forall i \in V(G)$$
(8)

每个无人机在速度偏离整体速度后都会自适应进行速度调节,因此在所有时间的速度收敛于v<sub>0</sub>,可以用以下公式进行描述:

$$\lim v_i(t) = v_0, \quad \forall i \in V(G) \tag{9}$$

无人机集群运动状态如图2所示,选定16 号节点为领航节点,在图中用黄色标出,其余 蓝色节点为跟随节点。考虑到节点之间的通信 链路为双向链路,因此集群网络拓扑结构应当 具有对称性,网络所映射的图论模型也应当为 无向权重图。无人机与其最近的几个邻居节点 建立双向链路,构成如图2所示的网络拓扑。

在无人机集群执行任务过程中,当无人机 集群中节点数较多时,节点因发生故障而出现 失效的情况在所难免。此时,网络拓扑会因该 节点的移除而发生改变,进而影响整个网络的 连通性。如图3(a)所示,如果14号和28号节点 发生故障,无人机集群网络映射到图论中的数 学模型则转化为图3(b)。在图3(b)中,无人机 集群网络内部虽然能够进行信息交互,但是22 号节点承担了右下方通信子集与其他节点之间 的通信,如果22号节点发生故障,则无人机集 群网络被分割为两个不连通的子集,此时,集 群内部节点无法进行信息同步,集群整体的网 络性能发生了严重的下降。因此,本文从节点 移除对网络连通性造成的影响角度来考虑无人 机飞行自组织网络的连通性维护问题,通过链 路预测算法与局部拓扑修复算法,实现网络的 连通性恢复。



图 2 无人机集群运动方式与初始网络拓扑

#### 2 网络修复算法

灰色预测是一种能够针对样本数据量小、 数据变化不规律的时间序列进行预测的方法, 具有预测速度快、预测拟合度高、参数估计简 单和预测结果可检验等优点。本文采用灰色滚 动预测算法,对节点之间的链路状态信息进行 实时预测。但由于灰色预测算法在非线性数据 序列预测方面存在准确性不足的问题,因此本 文利用鲸鱼优化算法,考虑了"新信息优先" 原则,对灰色预测算法时间响应函数的初始值 进行改进,提出了鲸鱼权值优化-灰色滚动 (whale optimization algorithm-weight of grey model, WOA-WGM)预测算法。

相较于传统的灰色预测算法,本文提出的 鲸鱼权值优化-灰色滚动预测算法有效提高了预 测的精确度,增加了预测值与样本数据序列之 间的关联性,并具有良好的非线性时间序列预 测能力。此外,通过Dijkstra算法对局部拓扑恢 复路径进行寻优,考虑链路负载方差,并通过 鲸鱼算法对寻优后的路径进行负载均衡处理, 实现局部拓扑的最优恢复,提高恢复后网络的



图 3 节点发生故障后网络拓扑变化图

鲁棒性。

#### 2.1 灰色滚动预测模型

GM(1,1)灰色预测是一种可以对不确定时 间序列进行拟合和估计的有效工具,其数据样 本空间允许少到4个,非常适合于链路质量的快 速估计<sup>[11]</sup>。无人机集群网络拓扑结构变化频繁, 需要及时对下一时刻的链路状态进行预测,因 此在灰色预测模型中加入时间窗口,不断去旧 值、添新值,保证样本数据的实时性。

设原始时间序列为 $W^{(0)} = \{w^{(0)}(1), w^{(0)}(2), \cdots, w^{(0)}(n)\}$ 其中 $w^{(0)}(k)$ 是在时刻k时的数据。 $w^{(0)}$ 的 累加生成序列为 $W^{(1)} = \{w^{(1)}(1), w^{(1)}(2), \cdots, w^{(1)}(n)\},$ 其中 $w^{(1)}(k) = \sum_{i=1}^{k} w^{(0)}(i)$ 。对于GM(1,1)预测模型,其灰色微分方程为

$$w^{(0)}(k) + az^{(1)}(k) = b \tag{10}$$

式中,

 $z^{(1)}(k) = 0.5w^{(1)}(k) + 0.5w^{(1)}(k-1), k = 2, 3, \dots, n;$ a和b分别为发展系数和灰作用量。将时间序列  $W^{(0)}, W^{(1)}$ 代入公式(10),可以得到如下公式:

$$\begin{cases} w^{(0)}(2) + az^{(1)}(2) = b \\ w^{(0)}(3) + az^{(1)}(3) = b \\ \cdots \\ w^{(0)}(n) + az^{(1)}(n) = b \end{cases}$$
(11)

将公式(11)转化为矩阵形式,即可得到:

$$B = \begin{bmatrix} -z^{(1)}(2) & 1\\ -z^{(1)}(3) & 1\\ \vdots & \vdots\\ -z^{(1)}(n) & 1 \end{bmatrix}, Y = \begin{bmatrix} w^{(0)}(2)\\ w^{(0)}(3)\\ \vdots\\ w^{(0)}(n) \end{bmatrix}$$
(12)

求解上述公式,可以得到参数a和b的最小 二乘估计:  $[a, b]^{T} = (B^{T}B)^{-1}B^{T}Y_{\circ}$ 

对输入的时间序列建立用于预测的微分方程:

$$\frac{\mathrm{d}w^{(1)}}{\mathrm{d}t} + aw^{(1)} = b \tag{13}$$

将参数 a 和 b 的预测值代入公式(13),并将 w<sup>(1)</sup>(1) = w<sup>(0)</sup>(1) 作为初始条件代入,得到公式 (13)求解后的时间响应函数:

$$w^{(1)}(k+1) = \left[w^{(1)}(1) - \frac{b}{a}\right]e^{-ak} + \frac{b}{a}, \ k = 1, 2, \cdots, n$$
(14)

最终通过对 $\hat{w}^{(1)}(k+1)$ 累减求得 $\hat{w}^{(0)}(k+1)$ 的预测值,得到预测值序列:

$$\hat{w}^{(0)}(k+1) = \left(1 - e^{a}\right) \left[ w^{(0)}(1) - \frac{b}{a} \right] e^{-ak}, k = 1, 2, \cdots, n$$
(15)

由此,即可得到*k*+1时刻的链路稳定性预测 值。

为预测未来数据 $\hat{w}^{(0)}(k+2)$ ,移除旧数据序 列中的 $w^{(0)}(1)$ 并且使用最新的数据序列 $w^{(0)}(2)$ ,  $w^{(0)}(3), \dots, w^{(0)}(k+1)$ 作为滚动GM(1,1)灰色预 测模型的新序列,随后再次执行新的预测过程。

#### 2.2 鲸鱼权值优化模型

#### 2.2.1 时间响应函数优化策略

经典灰色预测的时间响应函数的初始条件 为输入时间序列的第一个参数,当进行滚动预 测时,会导致对信息的适应能力下降,从而导 致灰色预测模型的预测性能降低。针对这个问 题,本文通过对时间响应函数的初始序列进行 加权处理,利用鲸鱼算法在每次滚动预测时对 权值进行寻优。经典灰色预测时间响应初始条 件为 $w^{(1)}(\beta)$ ,对其进行加权处理,具体为  $w^{(1)}(\beta)=\alpha^{n-1}w^{(1)}(1)+\alpha^{n-2}w^{(1)}(2)+\cdots+\alpha^{n-n}w^{(1)}(n)=$ 

$$\sum_{k=1}^n lpha^{n-k} w^{(1)}(k)$$

优化后的时间响应函数为

$$\hat{w}^{(1)}(k) = \left[ w^{(1)}(\beta) - \frac{b}{a} \right] e^{-a(k-\beta)} + \frac{b}{a}, \quad (17)$$
  
$$k = 2, 3, \cdots, n, n+1, \cdots$$

(16)

其中: $\alpha^{n-k}(0 < \alpha < 1)(k = 1, 2, \dots, n)$ 是动态权 重系数, $\beta(1 \le \beta \le n)$ 是时间输入系数。

新提出的初始条件也充分考虑了影响模型 准确性的历史信息,由于 $\alpha$ 满足 $0<\alpha <1$ ,因此 参数 $\alpha$ 次数越高,权重越小,即k值越大,相应 的 $x^{(1)}(k)$ 加权值越大:

 $\alpha^{n-1} < \alpha^{n-2} < \alpha^{n-3} < \cdots < \alpha^{1} < \alpha^{n-n}$ 

生成系数 α 和β 的最佳值是通过最小化预测 值和实际值之间的平均绝对百分比误差来计算。 为此,优化初始条件下的最佳生成系数由以下 目标函数确定:  $\min_{\alpha,\beta} \frac{1}{n} \sum_{k=1}^{n} \frac{\left| \hat{w}^{(0)}(k) - w^{(0)}(k) \right|}{w^{(0)}(k)} \times 100\%$   $B = \begin{bmatrix} -z^{(0)}(2) & 1 \\ -z^{(0)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} w^{(0)}(2) \\ w^{(0)}(3) \\ \vdots \\ w^{(0)}(n) \end{bmatrix}$ s.t.  $\left\{ \hat{w}^{(1)}(k) = \begin{bmatrix} \sum_{i=1}^{n} \alpha^{n-i} w^{(1)}(i) - b/a \\ \begin{bmatrix} i \\ i \end{bmatrix} (1 - e^{\hat{a}}) e^{-\hat{a}(k-\beta)} \\ \begin{bmatrix} a, b \end{bmatrix}^{\mathrm{T}} = (\mathbf{B}^{\mathrm{T}} \mathbf{B})^{-1} \mathbf{B}^{\mathrm{T}} \mathbf{Y} \\ \alpha \in (0, 1), k = 1, 2, \cdots, n \end{cases}$ (18)

其中: $\hat{w}^{(0)}(k)$ 是拟合数据值; $w^{(0)}(k)$ 是实际数据 值;n是输入数据的数量。

#### 2.2.2 鲸鱼算法权值寻优

逼近目标函数的最小值有助于获得权重系 数的最优值。由于目标函数的非线性特性,无 法采用常规方法求解,而智能优化算法可以简 单、快速地解决非线性优化问题。其中,鲸鱼 算法收敛速度快、局部搜索能力强,可以快速 得到近似最优解,因此,本文采用鲸鱼优化算 法对模型参数进行优化,使得该模型的建模误 差减小,进而获得模型非线性参数的最优值。

鲸鱼算法是通过模拟鲸鱼捕获猎物的方式 来实现寻优,鲸鱼算法求解过程主要经历三个 阶段:搜寻猎物阶段、环绕包围以及起泡网狩 猎<sup>[12]</sup>。鲸鱼算法位置更新方式中,环绕包围捕 猎和螺旋式路径捕猎概率*p*等同,因此位置更新 公式为

 $\vec{X}(t+1) =$  $\begin{cases} \vec{X}^{*}(t) - \vec{A} \cdot \vec{D} & p < 0.5, \left| \vec{A} \right| < 1 \\ \left| \vec{X}^{*}(t) - \vec{X}(t) \right| \cdot e^{\beta t} \cdot \cos\left(2\pi t\right) + \vec{X}^{*}(t) & p \ge 0.5, \left| \vec{A} \right| < 1 \\ \vec{X}_{r}(t) - \vec{A} \cdot \left| 2\vec{r} \cdot \vec{X}_{r}(t) - \vec{X}(t) \right| & \left| \vec{A} \right| \ge 1 \end{cases}$ (19)

其中: $\vec{X}(t)$ 为座头鲸的位置0; $\vec{X}^{*}(t)$ 表示当前 鲸鱼位置的最优解,参数 $\vec{r}$ 是区间[0,1]中随机 数,参数l是区间[-1,1]中的随机数, $\beta$ 为常数, 通常用以定义对数螺旋,包围收缩的参数 $\vec{A}$ 和 $\vec{D}$ 如下式:

$$\begin{cases} \vec{D} = \left| 2\vec{r} \cdot \vec{X}^{*}(t) - \vec{X}(t) \right| \\ \vec{A} = 2f(t) \cdot \vec{r} - f(t) \\ f(t) = 2 - 2t/T \end{cases}$$
(20)

式中,T表示最大迭代次数。

在利用鲸鱼算法对参数进行权值寻优之前, 首先需要设定待求参数α和β的上界和下界,并 对算法参数进行初始化,然后生成初始鲸鱼位 置序列,通过三种策略迭代更新鲸鱼个体适应 度最高的位置,最后,达到最大迭代次数,输 出最优解。具体步骤如下:

第一步:初始化鲸鱼算法的参数*r*,*p*,*l*,*b*, 设置鲸鱼算法的种群数量*N*、最大迭代次数*T*, 确定算法的复杂度;

第二步:初始化鲸鱼种群位置 $X = \{x_1, x_2, \dots, x_N\}$ ,并设置待求参数 $\alpha$ 和 $\beta$ 的范围区间。其中,权重参数 $\alpha$ 的范围为 $\alpha \in (0, 1)$ ,时间输入系数 $\beta$ 的范围为 $\beta \in [1, n]$ ;

第三步:根据算法的成本函数计算鲸鱼个体的成本值,选择成本值最低的鲸鱼个体作为最优解,鲸鱼优化算法的成本函数如下:

$$Fitness = \min_{\alpha,\beta} \frac{1}{n} \sum_{k=1}^{n} \frac{\left| \left[ \sum_{i=1}^{n} \alpha^{n-i} w^{(1)}(i) - b/a \right] (1 - e^{a}) e^{-a(k-\beta)} - w^{(0)}(k) \right]}{w^{(0)}(k)}$$
(21)

第四步:更新鲸鱼优化算法的参数;

第五步:根据公式(19)对鲸鱼个体位置进 行更新。鲸鱼个体位置更新分为三种:搜寻猎 物阶段、环绕包围以及起泡网狩猎。根据更新 的参数选择相应的方法进行位置更新,并对鲸 鱼个体的成本值进行计算,更新最优解;

第六步:判断算法是否达到最大迭代次数, 若未达到,则转第三步,否则输出最优解作为 参数α和β的最优值。

利用鲸鱼算法找到最优解后,即可确定初 始条件的权重参数和时间参数的最优值,将其 代入灰色预测算法,对输入序列进行预测。在 滚动预测过程中,需要不断对输入序列去旧值、 添新值,并在每次滚动预测过程中,利用鲸鱼 优化算法对新的初始条件进行优化,保证数据 预测的准确性。预测算法全部过程如下: 第一步:采用实际观测值 $w^{(0)}(1), w^{(0)}(2), \cdots, w^{(0)}(c)$ 作为灰色滚动预测算法的输入数据。根据 上述成本函数,使用鲸鱼优化算法计算参数 $\alpha$ 和 $\beta$ 。由此可以获得预测数据 $\hat{w}^{(0)}(c+1)$ ;

第二步:为预测未来数据 $\hat{w}^{(0)}(c+2)$ ,移除旧 数据序列中的 $w^{(0)}(1)$ ,并且使用最新的数据序列  $w^{(0)}(2), w^{(0)}(3), \dots, w^{(0)}(c+1)$ 作为滚动 GM(1,1) 灰色预测模型的新序列;然后利用鲸鱼算法重 新计算参数 $\alpha$ 和 $\beta$ 的新值,随后,可以再次执行 新的预测过程;

第三步:重复上述步骤,直到预测出所有 剩余数据点。

### 2.3 Dijkstra和鲸鱼优化网络拓扑修复 算法

本文设计的网络拓扑恢复算法将网络结构 和网络负载都纳入考量。针对网络结构,通过 预测后的通信代价矩阵,选择通信代价最小的 链路进行重建,这样有助于网络整体的可靠性。 但是, 仅考虑通信代价进行拓扑修复可能会导 致少量链路的负载过大,从而引起整个网络的 传输性能下降,因此网络恢复算法应当能够在 最小通信代价的基础上对拓扑进行负载均衡处 理,优化集群网络内部的链路数量于链路负载, 保证集群内部信息能够进行有效传输的同时对 拓扑结构进行改善。考虑到无人机移动速度快、 通信链路质量变化频繁的特点,对局部拓扑恢 复寻优的问题很难采用某种解析方法去解决, 而鲸鱼算法可以在有限计算时间内寻找优化问 题的最优解或者次优解,因此本文提出将Dijkstra算法<sup>[13]</sup>和鲸鱼优化算法相结合,来对网络拓 扑进行修复。

使用本文预测算法预测得到链路权值后, 需要对备选节点进行筛选,选择合适的节点建 立新的链路。首先通过Dijkstra算法寻找通信代 价最小的链路,核心思想是选定一个起始节点 和目的节点,其余待恢复节点为必经节点,利 用基于回溯法的Dijkstra算法寻找起始节点与目 标节点之间的最佳路径。如图4所示,虚线代表 预测的链路,实线双箭头代表节点之间存在双 向稳定链路,在当前时刻,节点Z失效,节点X 与节点Y链路断开,对这两个节点执行网络恢 复策略。



图 4 无人机节点链路重选示意图

根据公式(6)可知:通信代价越高,节点之间的相对速度和距离就越大,因此取通信代价最小的路径作为最优恢复路径。设节点X到节点Y的路径边集为*E<sub>xy</sub>*,则X到Y的最小通信代价路径为

$$E_{\min} = \min\left(\sum_{e \in E_{xy}} w_e\right)$$
(22)

由公式(9)可以计算无人机链路负载权值, 要实现无人机恢复后的局部拓扑负载优化,本 文通过计算恢复后拓扑负载的方差来衡量链路 负载均匀度:

$$D = \sum_{N} \left( \sum_{e \in E_{xy}} LC_e - AVG \right)^2$$
(23)

其中: *LC*<sub>e</sub>代表局部拓扑中无人机链路e的负载权 值; *N*代表无人机局部拓扑的链路数目; *AVG*代 表局部拓扑的负载权值均值,由以下公式确定:

$$AVG = \frac{\sum_{e \in E_{xx}} LC_e}{N}$$
(24)

综合考虑无人机的通信代价 E 和链路负载 均匀度 D,对于局部拓扑 G<sub>part</sub>,将以无人机网络 中的通信代价 E 和负载均匀度 D 作为目标函数, 通过权重 k 来均衡无人机,使得在架构成本和负 载均衡之间做出权衡,最终设计方案应使得目 标函数尽可能小,即拓扑设计方案的成本更低, 负载更均衡:

 $\min f(G_{\text{part}}) = E(G_{\text{part}}) + k*D(G_{\text{part}})$  (25)

根据无人机拓扑恢复的目标函数,结合无 人机网络中对于链路的相关约束,可以得到以 下无人机拓扑恢复方法的目标函数:

$$\min f(G_{\text{part}}) = E(G_{\text{part}}) + k*D(G_{\text{part}})$$

$$AVG = \frac{\sum_{e \in G_{\text{part}}} LC_e}{N}$$

$$E(G_{\text{part}}) = \sum_{e \in G_{\text{part}}} w_e$$

$$D(G_{\text{part}}) = \frac{\sum_{e \in G_{\text{part}}} (LC_e - AVG)^2}{N}$$

$$0 < LC_e < LC_{th}, e \in G_{\text{part}}$$

(26)

将上述优化目标函数代入到鲸鱼算法中进 行寻优,设置鲸鱼算法的成本函数为

$$Fitness = \min_{G_{part}} \left[ \sum_{e \in G_{part}} w_e + k^* \frac{\sum_{e \in G_{part}} \left[ LC_e - \frac{\sum_{e \in G_{part}} LC_e}{N} \right]^2}{N} \right]$$
(27)

根据算法的成本函数,计算鲸鱼个体的成本 值,以成本值最低的鲸鱼个体作为当前的最优 解,并通过收缩包围、螺旋运动和随机游走三种 方式,对鲸鱼算法的参数选择相应的方法进行位 置更新,并对鲸鱼个体的成本值进行计算,更新 最优解,最终实现对无人机局部的拓扑恢复。

#### 3 实验结果与仿真分析

#### 3.1 实验参数设置

本实验通过 Python 仿真平台进行算法的仿 真,并对本文提出的 WOA-WGM 预测算法进行 仿真分析,以验证不同网络损伤类型下集群网 络的生存能力和抗毁能力。具体的仿真参数设 置如表1所示。

参数	值
仿真区域	15000 m×15000 m
节点个数	30
最大通信半径R <sub>m</sub>	500 м
无人机移动速度	[13 m/s, 20 m/s]
无人机运动方式	参考点组移动模型
节点失效模式	随机失效/蓄意攻击
天线类型	全向天线

表 1 仿真参数设置

为了对本文提出的网络恢复算法进行仿真 评估,本文选取 BA 无标度网络算法以及 LPN (link prediction based on neighbors)一跳邻居链路 预测算法这两个经典恢复算法对比分析网络恢 复性能。在BA网络修复算法中,需要进行连通 性修复的节点会自发寻找通信范围内重要度最 高的节点进行边重连,修复后的网络具备一定 的无标度网络特性。LPN网络算法会对每个节 点的通信代价矩阵<sup>[9]</sup>进行预测,根据预测结果, 选择链路权值最好的节点进行重连,实现网络 的连通性恢复。在仿真实验中,首先分析本文 网络恢复算法在集群最大速度不同的情况下恢 复网络连通性所用的时间,验证本文算法的时 效性;其次,验证无人机集群网络在随机失效 下的网络恢复能力,验证算法恢复后网络的生 存性;最后,评估蓄意攻击下无人机集群网络 的恢复能力,验证恢复后网络的抗毁性。

#### 3.2 仿真实例分析

根据以上介绍的仿真环境及参数设置,对 无人机集群整体位置进行仿真模拟,每个无人机 参考领航节点,以速度v<sub>0</sub>+v<sub>rand</sub>在区域内部进行 飞行。每个无人机与自身若干相邻节点建立链 路,组成网络的初始拓扑结构。在进行了13次飞 行仿真后,得到如图5所示的网络拓扑结构。



图 5 无人机集群网络拓扑结构

在13次仿真过程中,每个无人机通过与邻 居节点进行信息共享,根据自身与邻居节点的



图 6 通信代价预测矩阵 🕅

信号功率强度、相对速度和负载程度计算出对应的通信代价权值 $w_{ij}$ 。集群中的无人机根据权值生成集群网络的通信代价矩阵W,以预测窗口大小为基准代入多个时刻的历史样本数据,通过WOA-WGM算法预测得出通信代价的预测矩阵 $\hat{W}$ ,如图6所示。

其中: A<sub>1</sub>, A<sub>2</sub>, …, A<sub>30</sub>代表无人机节点编号, inf代表无人机节点之间不存在通信链路。

采用表1的参数进行仿真实验,随机产生失效节点ID。设集群网络中的失效节点集合为 FN,则集群网络中拓扑结构受到影响的无人机 节点向其邻居节点发送网络恢复请求,集群内 部通过信息共享得到待修复节点集合RN。集合 FN和RN包含的节点如下所示:

 $FN = \{A_{14}, A_{22}, A_{28}\},\$ 

 $RN = \{A_8, A_{13}, A_{15}, A_{16}, A_{20}, A_{23}, A_{27}, A_{29}\}$ 

节点编号为14、22、28的无人机失效后, 集群网络拓扑结构如图7所示。在图中,绿色节 点代表连通性受影响节点集合*RN*,红色节点代 表失效节点集合*FN*,阴影部分表示需要对该部 分的局部拓扑结构进行修复。

当节点A<sub>14</sub>,A<sub>22</sub>,A<sub>28</sub>失效后,集合 RN 中的节 点按序依次寻找最短路径。集合中的待恢复节 点根据通信代价矩阵预测值 ŵ,通过最小通信 代价路径算法遍历出待恢复节点之间的路径, 以此实现网络局部拓扑结构的修复,结果如表2 所示,通过最小通信代价修复后的拓扑图如图8 所示,通过修复算法找到节点之间的最小通信 代价路径用点线表示。



图 7 无人机集群网络拓扑结构受损



图 8 最短路径拓扑修复图

队列	最短路径	路径权重
$A_8 - A_{13}$	$A_8 - A_{13}$	6.0
$A_8 - A_{19}$	$A_8 - A_{13} - A_{19}$	14.3
$A_8 - A_{15}$	$A_8 - A_{15}$	7.1
$A_8 - A_{16}$	$A_8 - A_{15} - A_{16}$	16.3
$A_8 - A_{20}$	$A_8 - A_{20}$	8.9
$A_8 - A_{23}$	$A_8 - A_{15} - A_{16} - A_{23}$	26.2
$A_8 - A_{27}$	$A_8 - A_{20} - A_{27}$	17.1
$A_8 - A_{29}$	$A_8 - A_{15} - A_{21} - A_{29}$	23.6
$A_{13} - A_{19}$	$A_{13} - A_{19}$	8.3
$A_{13} - A_{20}$	$A_{13} - A_{20}$	8.9
$A_{13} - A_{15}$	$A_{13} - A_{20} - A_{15}$	18.8
$A_{13} - A_{16}$	$A_{13} - A_{20} - A_{15} - A_{16}$	27.9
$A_{15} - A_{27}$	$A_{15} - A_{27}$	7.5
$A_{19} - A_{20}$	$A_{19} - A_{20}$	9.6
$A_{15} - A_{16}$	$A_{15} - A_{16}$	9.1
$A_{23} - A_{29}$	$A_{23} - A_{29}$	6.3
$A_{23} - A_{27}$	$A_{23} - A_{29} - A_{21} - A_{27}$	21.0

表 2 节点之间最短路径表

利用最小通信代价算法进行了初步修复后, 需要继续对其进行负载优化。在实验中,假设 每个无人机都向邻居节点发送长度为1帧的数 据,长度为256 byte,周期为4 ms。经过仿真实 验,比较了直接寻找最小通信代价方案和优化 设计方案,得到了链路数、负载均值和负载方 差的对比数据,如表3所示。拓扑修复图如图9 所示,图中待修复节点根据最短路径对局部拓 扑结构进行修复,避免因节点失效造成网络整 体重新进行拓扑重构。同时,本文算法提出的 网络修复方案在链路数量上进行了一定的优化, 网络体系代价有了明显的降低,负载更加均匀, 提高了集群网络的鲁棒性和健壮性。

表 3 仿真对比结果

	链路条数/条	负载均值/帧	负载方差/帧
最小通信代价 方案	10	4.2	0.68
优化设计方案	7	3.4	0.23



图 9 WOA-WGM 实现无人机集群网络拓扑修复

#### 3.3 网络恢复时效性分析

网络恢复时效性分析是指在无人机集群网络受到损伤后,恢复网络的连通性所需要的时间。图10中分别表示了不同速度与损伤情况下,本文算法与BA算法和LPN算法恢复网络的时效性对比,为消除可能存在的随机误差,采集的实验数据均为进行10次仿真后求得的平均值。

在图 10(a)中,设定无人机集群在3个节点 失效的情况下,不同算法在不同的速度恢复网 络连通性所需的时间,可以看出,在无人机集 群最大速度小于10 m/s时,经典BA算法和LPN 算法由于其计算复杂度低,可以很快找到合适 的链路进行拓扑修复。而随着无人机的最大移 动速度增大,三种恢复算法耗时均存在不同程 度的增加, BA算法和LPN算法增加最为明显, 这是因为随着无人机移动速度的增加, 节点之 间的信号强度和相对速度变化越频繁, BA算法 由于没有预测机制,在恢复网络连通性时需要 不断试错,LPN算法虽然可以对邻居节点的链 路进行预测,但是预测结果精确度不足,且不 能对多跳链路进行预测, 难以实现整个集群网 络的连通性恢复。在图10(b)中,设定无人机集 群在15 m/s的最大速度下,不同算法在损失不 同数量的节点修复网络拓扑所需的时间,同样, 经典 BA 算法和 LPN 算法都随着失效规模的增 大,恢复耗时急剧增加,而本文算法由于加入 了改进的预测机制和多节点路径恢复机制,对 下一时刻的无人机速度和链路距离进行了预判, 对于速度的变化较为敏感,能够使无人机在高 速移动时选择到高稳定链路,从而减少了不必 要的路由开销,实现网络连通性的快速恢复。



(a) 不同速度下恢复算法的时效性



(b)不同失效节点数目下恢复算法的时效性

图 10 无人机集群网络恢复算法时效性对比

#### 3.4 网络恢复生存性分析

网络恢复生存性是指无人机集群网络恢复 算法在节点随机失效情况下,恢复网络连通性 的能力。本文通过网络最大连通度和平均路径 长度比例对网络性能进行分析<sup>[14]</sup>。

网络最大连通度*G*表示最大连通子图节点数占总节点数的比例,表达式为

$$G = \frac{V'}{V} \tag{28}$$

对于带权无向图,平均路径长度与边权值 相关,平均路径长度*L*表达式如下:

$$L = \frac{\sum_{i \neq j} w_{ij}}{n(n-1)}$$
(29)

则平均路径长度比例*L*\*为受到影响后集群 平均路径长度与初始平均路径长度的比值:

$$L^* = \frac{L'}{L} \tag{30}$$

在仿真实验中,随机选择集群中的无人机 失效,分析无人机集群在不同失效规模下使用 不同算法恢复后网络的生存性。结果如图11所 示。图11(a)为网络最大连通度随失效规模变化 的曲线,可以看出,在失效比例小于0.4时, BA算法集群内部的最大连通度要高于LPN算 法,这是由于BA恢复算法选取度值最大节点实 现边重连,恢复后的网络相较于LPN算法能够 更好地保证网络整体连通性。LPN算法由于优 先选取通信代价最小的链路实现边重连,因此 容易陷入局部最优而忽略集群整体的连通性能, 导致集群内部被分割为多个不连通的子集,降 低网络的连通性。而本文算法考虑了所有网络 拓扑受到影响的节点之间的连通性恢复,因此 在网络拓扑连通性恢复过程中,节点在失效比 例小于0.5时,能够一直保持集群内部的连通性 不受影响,即使在网络大规模失效,本文算法 在网络整体的连通性能恢复上也更优于BA算法 和LPN算法。



(a) 最大连通度





图 11 无人机集群网络生存性指标

图 11(b)为平均路径长度比例随失效规模的 变化曲线,当节点失效比例小于 0.2 时,BA 网 络恢复算法所平均路径长度比例优于 LPN 算法, 这也验证了 BA 算法在执行恢复策略时,能够使 得网络中连通性下降的节点自动在通信范围内 选择连通度最高的节点实现网络重连,网络整体的连通性更好。然而当节点失效比例超过0.2时,BA算法的平均路径长度比例大幅降低,可以看出,当随机失效作用到关键节点时,BA算法所恢复的集群网络连通性会受到很大影响,网络中出现大量分割的连通子集,导致网络整体的生存性下降。本文算法相较于BA算法和LPN算法,平均路径长度比例下降更为平缓,在不同失效规模的情况下均优于传统的BA算法和LPN算法。综合最大连通度和平均路径长度比例,算法能够尽力维持网络整体的连通性,且平均路径长度比例在60%以上,此时网络处于相对可靠的状态,这是因为本文提出的算法在网络出现损伤后,会及时修复所有连通度下降节点之间的网络拓扑,补偿损失的节点连通路径。

#### 3.5 网络恢复抗毁性分析

网络抗毁性分析是无人机集群网络在遭到 恶意攻击作用下,集群内部保持连通性能的能 力。在仿真实验中,通过选取无人机集群网络 中节点度值较大的节点进行失效仿真,分析网 络恢复算法在不同失效规模下使用不同算法恢 复后网络的抗毁性。结果如图 12所示。

分析图 12(a)中网络最大连通度随失效规模 变化的曲线,可以看出 BA 算法的网络性能恢复 能力是三种恢复算法中性能最差的,这是由于 BA 算法恢复后的网络具有一定的无标度网络特 性,在随机失效时具有良好的鲁棒性,但是对 于蓄意攻击是脆弱的。LPN 算法虽然相较于 BA 算法在面对蓄意攻击时具有更好的恢复性能, 但是网络最大连通度还是出现了大幅下降,难 以维持集群整体连通性。本文算法在失效比例 小于 0.5 时,依然能够维持集群整体的连通性, 且在失效规模继续增加时,最大连通度下降曲 线较为平缓,优于传统的 BA 算法和LPN 算法。

从图 12(b)可以看出, BA 算法和 LPN 算法 在面对蓄意攻击时,平均路径长度比例在较小 的失效规模下发生了急剧下降,网络整体已经 处于多个不连通子集状态。而本文算法可以对 通信代价实时预测,快速选择合适的节点建立 链路,同时通过最小通信代价策略初步对缺失 链路进行恢复,并通过负载均衡策略对链路进 行优化,因此从图中可以看出,不同失效规模 下,平均路径长度变化较为平缓,不会由于新 增链路造成网络平均路径长度发生急剧变化, 实现了无人机集群网络的有效恢复,减少了不 必要的冗余链路,保证了网络的可靠性。







#### (b) 平均路径长度比例

#### 图 12 无人机集群网络抗毁性指标

结合图 11 和图 12 中的数据,本文提出的网 络恢复算法在随机失效和蓄意攻击下相较于传 统的网络恢复算法能够更好地维护网络性能, 并且在失效比例不超过 0.5 的情况下,能够稳定 地维持集群整体的连通性。在节点失效比例在 0.5 至 0.7 时,集群网络内部大范围节点都已失 效,本文提出的算法在最大连通度和平均路径 长度比例方面都明显优于 BA 算法和 LPN 算法, 表明本文算法在集群遭到大规模损伤时仍然能 够尽力维持集群中的极大连通子集,提高网络 的生存性和抗毁性。

#### 4 结语

针对高动态拓扑下无人机节点失效造成的 网络性能和可靠性下降的问题,本文根据信号 稳定度指标和链路负载度指标计算无人机之间 的通信代价,构建无人机集群网络通信模型, 并通过鲸鱼权值优化-灰色滚动链路预测算法对 通信代价指标进行预测,降低了无人机高速移 动对修复算法实时性的影响,最后利用最短路 径算法和负载均衡算法,实现了网络连通性恢 复并优化了网络恢复后的拓扑结构。实验结果 表明,该方法在无人机集群网络出现节点失效 时可以快速进行拓扑重构,恢复通信节点之间 的链路连接,实现网络连通性恢复,并使恢复 后的网络具有良好的生存性和抗毁性。

#### 参考文献:

- [1] AI J, GENG L, AN Y, et al. The study of resilience on unmanned swarms systems [C] //Proceedings of the 2021 IEEE International Conference on Unmanned Systems (ICUS). Beijing, China, 2021: 576-579.
- [2] ZHANG Q, FENG G, QIN S, et al. Distributed Topology Control based on Swarm Intelligence In Unmanned Aerial Vehicles Networks [C] // Proceedings of the 2020 IEEE Wireless Communications and Networking Conference (WCNC). Seoul, Korea, 2020: 1-6.
- [3] 马学森,曹政,韩江洪,等.改进蚁群算法的无线传感器网络路由优化与路径恢复算法[J].电子测量与仪器学报,2015,29(9):1320-1327.
- [4] 姚玉坤,宋威威,李其超,等. 链路预测模型下的无 人机网关选择算法[J]. 小型微型计算机系统, 2020,41(2):338-343.
- [5] 李玉龙,黄国策,张衡阳,等.移动预测的无人机自 组网路由协议[J].空军工程大学学报(自然科学 版),2015,16(3):61-65.

- [6] 康巧琴,袁丁,严清. Mesh网络中基于效用转发的 网络编码算法[J]. 计算机应用研究, 2021, 38(9): 2782-2787, 2791.
- [7] 肖军弼,孟祥泽,田爱宝,等.SDN网络中基于拓扑 分域的故障恢复方法[J]. 计算机工程与设计, 2021,42(6):1501-1508.
- [8] 孙明杰,周林,于云龙,等.无人机自组网中基于蚁 群优化的多态感知路由算法[J].系统工程与电子 技术,2021,43(9):2562-2572.
- [9] 梅家栋,南建国.无人机自组网基于加权的稳定分 簇算法[J]. 计算机应用研究, 2021, 38 (11): 3411-3416.
- [10] ZHANG Z, ZHAO Y, LIU J, et al. A general deep learning framework for network reconstruction and dynamics learning [J]. Applied Network Science, 2019,4(1):1-17
- [11] DING S, HIPEL K W, DANG Y. Forecasting China's electricity consumption using a new grey prediction model[J]. Energy, 2018, 149: 314-328.
- [12] CHAKRABORTY S, SHARMA S, SAHA A K, et al. SHADE-WOA: a metaheuristic algorithm for global optimization [J]. Applied Soft Computing, 2021,113:107866.
- [13] QING G, ZHENG Z, YUE X. Path-planning of automated guided vehicle based on improved Dijkstra algorithm [C] //Proceedings of the 2017 29th Chinese Control and Decision Conference (CCDC). Chongqing, China, 2017:7138-7143.
- [14] 陈浩.复杂条件下固定翼无人机集群编队控制研究
   [D].长沙:国防科技大学,2020.

### An adaptive network restoration method for UAV cluster

#### Shi Yunyang, Hua Xiang<sup>\*</sup>, Zhang Jinjin

(School of Electronic Information Engineering, Xi'an Technological University, Xi'an 710021)

Abstract: For solving the problem of degraded network performance and reliability caused by node failure in UAV cluster network, an adaptive UAV cluster network restoration method is proposed in this work. The communication cost within the cluster is evaluated through signal stability and link load degree, and the network communication model of UAV cluster is built based on graph theory. The whale algorithm is used to optimize the grey rolling prediction, and the optimized prediction model is used to evaluate the communication cost in advance. According to the predicted value of the communication cost matrix, the algorithm finds the shortest path between the nodes to be recovered, and optimizes the restored path by using the load balancing algorithm to realize the performance recovery of the UAV network. The simulation results show that the proposed method can quickly reconstruct the topology, restore the link between communication nodes, and make the network have good survivability and destruction-resistance when the UAV cluster network fails.

Keywords: UAV cluster network; communication cost; improved prediction algorithm; topological restoration

文章编号:1007-1423(2023)02-0014-10

DOI: 10.3969/j.issn.1007-1423.2023.02.002

### 基于时间序列的民用运输航空器碳排放预测研究

向小军1,杨志晗2\*,赵赶超2

(1. 中国民用航空飞行学院科研处, 广汉 618300; 2. 中国民用航空飞行学院飞行技术学院, 广汉 618300)

摘要:随着中国民航业的高速发展,运输航空器的碳排放问题逐渐引起关注。采用时间序列的方法建 立了传统的差分整合移动平均自回归(ARIMA)模型以及优化的长短期记忆网络(LSTM)模型,对航空器碳 排放量、碳排放强度以及吨公里碳排放量进行了预测,通过鲸鱼优化算法(WOA)对LSTM中的学习率和隐 藏节点数进行优化,避免了人为选择参数的主观性和盲目性,有利于提高模型预测的准确性。通过对比两 种模型的均方根误差(RMSE)和平均绝对误差(MAE),ARIMA模型在航空器碳排放预测中有较好表现, WOA-LSTM模型在碳排放强度、吨公里碳排放的预测中有较好表现。

关键词:时间序列;ARIMA;WOA-LSTM;碳排放

#### 0 引言

近年来,我国民航业得到巨大发展,虽受 到"新冠"疫情的影响, 2021年民航运输总周 转量仍相较于十年前增长48%,机队规模扩大 近1.3倍,在《"十四五"民用航空发展规划》 的指导下,民航业还将继续朝着民航强国的目 标飞速发展。与此同时,随着民航业发展,航 空燃油消耗量每年也呈增长态势,由燃油消耗 带来的碳排放与日俱增。据统计,我国民航业 碳排放约占全国碳排放的1.6%,可以预见,未 来民航业将成为主要的能源消费和碳排放来源 之一。中国已正式宣布将力争在2030年前实现 碳达峰、2060年前实现碳中和<sup>[1]</sup>,由于航空业 研发投入大、周期长、难度高等特点,很难像 其他产业一样在短时间内完成低碳转型,因此, 民航业减排将成为我国完成"双碳"目标的重 要挑战。在这样的背景下,预测我国民航业碳 排放量并分析减排措施对于早日完成行业内 "双碳"目标有一定参考意义。

目前,对于碳排放预测的研究主要分为三 个方面。一是结合历史数据,建立回归预测模 型,通过若干易统计数据来推测碳排放量。Wei 等[2]通过相关性分析,分析了二氧化碳排放与 影响因素之间的关系,然后使用优化后的最小 二乘支持向量机预测了碳排放量,该方法有效 提高了碳排放的预测精度。Faruque 等<sup>[3]</sup>建立了 二氧化碳排放受电能消耗和国内生产总值 (GDP)影响的预测模型,对比分析了卷积神经 网络(CNN)、卷积神经网络-长短期记忆(CNN-LSTM)、LSTM 和密集神经网络(DNN)四种深度 学习方法的预测精度。陈文婕等[4]基于私家车 轨迹数据,通过逆地理编码与人工神经网络对 各区域私家车碳排放量进行预测,并从效率、 效果和公平三个角度评估各区域减排潜力,对 我国道路交通领域制定减排策略提供了参考。 高金贺等<sup>[5]</sup>在STIRPAT模型的基础上,建立了 由遗传算法优化的支持向量机(GA-SVR)预测模 型,选取人口总量、人均GDP、机动车保有量、

**收稿日期:** 2022-11-04 修稿日期: 2022-12-23

基金项目:中国民用航空飞行学院面上基金项目(J2021-015)

作者简介:向小军(1972—),男,四川广安人,硕士,教授,研究方向为飞行运行与飞行性能;\*通信作者:杨志晗 (1995—),男,山东菏泽人,硕士研究生,研究方向为飞行运行、计算机仿真,E-mail:420495780@qq.com;赵赶超(1992—), 男,河南洛阳人,硕士,讲师,研究方向为飞行运行及安全技术 旅客周转量、货物周转量、城镇化率和碳排放 强度作为影响指标,对北京市1995—2019年期 间数据进行分析,结果表明该模型具有良好的 拟合回归效果。

二是进行情景分析或建立碳达峰预测模型, 对各行业碳达峰情况进行预测。李宝成等[6]先 是利用排放系数法对河南省交通运输领域碳排 放进行测算,分析能耗与碳排放量、碳排放强 度等方面的关系;之后结合情景分析法预测碳 达峰时间和碳排放峰值,最后提出一系列节能 减排措施。韩博等<sup>[7]</sup>构建了一套符合民航特征 的碳排放综合预测模型,对2019—2050年民航 碳排放量开展预测分析,结果表明,民航业很 难实现2030年碳达峰愿景,预测民航业将于 2045年实现碳达峰,在此基础上,可持续性航 空燃料等新技术的发展将会使民航碳达峰时间 提前4—7年。胡荣等[8]运用情景分析和蒙特卡 洛模拟方法预测了厦门机场航空器碳排放达峰 的可能性、峰值与影响因素,结果表明,在基 准情景下未来15年内碳排放年均增长率为3.1%, 期间没有达峰可能性。许继辉等<sup>19]</sup>基于LEAP构 建自上而下的中国民航业能源系统模型, 探讨民 航业中长期低碳发展的技术路径,预计2060年左 右中国人均乘机次数翻两番,现有政策情景下, 民航业碳排放有望在2046年左右达峰,峰值水平 约为3.5亿吨。李心怡等<sup>[10]</sup>首先采用Kaya恒等式 和对数平均迪氏指数分解法(LMDI)分解中国民航 运输碳排放的影响因素,其次建立Tapio 解耦模 型分析民航运输碳排放量与各影响因素的关联强 度,最后运用改进可拓展的STIRPAT模型实现中 国民航运输碳排放预测,预计基准情景下2050年 之前未有碳排放峰值出现。

三是借助计算机深度学习等方法建立时间 序列预测模型,对碳排放进行短期或中长期预 测。张金良等<sup>[11]</sup>首先根据 Kaya 恒等式的扩展, 分析出影响碳排放的主要因素为人口、经济、 产业结构、能源消费强度以及消费结构;然后 分别建立线性回归、径向基函数神经网络、 ARIMA 以及 BP神经网络模型,对比得到最优模 型;最后基于最优模型在基准发展、产业优化、 技术突破、低碳发展这4个情景下对未来 30年 碳排放进行预测。徐丽等<sup>[12]</sup>对 1997—2015年中 国居民能源消费碳排放现状进行分析,基于 ARIMA模型对 2016—2025年居民能源消费碳排 放进行预测研究,该模型预计,到2025年居民 能源消费碳排放将达7.87亿吨。张昱等[13]设计 并实现了一种基于 Attention 机制的 CNN-LSTM 时序预测方法,将此方法应用于某大型园区八 个组团的建筑热负荷多步预测中,结果表明此 方法有较高的精确度。魏光普等[14]对内蒙古包 头市 2002—2020 年城区三种建筑类型能源消耗 量进行分析,以循环神经网络模型为基础构建 LSTM预测模型,并对2021-2030年的包头市建 筑碳排放量进行预测,结果表明包头市建筑行 业碳排放将于2022年达峰。胡剑波等[15]基于 LSTM神经网络模型预测出我国碳排放强度变化 趋势,同时,建立ARIMA-BP神经网络模型作 为验证模型对碳排放强度进行直接预测,两模 型对比后预测结果相差 2.03 个百分点。Yang 等[16]先是使用自上而下的方法对往返上海的所 有客运航班进行碳排放估算,后建立 ARIMA 模 型对途中碳排放进行5年预测,最终得到最佳的 ARIMA模型阶数。柴建等<sup>[17]</sup>基于单变量(ETS、 ARIMA模型)和多变量(Bayes多元回归)两个维 度对航空燃油效率及航空运输总周转量进行分 析和预测,最终综合以上两种模型结果得出航 空燃油效率、航空总周转量与人均 GDP、城镇 化率之间的关系。

综上可以看出,国内外学者对碳排放的回 归、达峰等预测已相当成熟,该类研究主要集 中在综合碳排放、高耗能产业碳排放等方面。 对于交通运输业的碳排放预测主要集中在公路 运输上,而民航业碳排放研究相对较少。本文 基于时间序列预测的方法建立ARIMA模型以及 WOA-LSTM组合模型,对我国民航业航空器碳 排放量、碳排放强度等数据预测至2035年并对 比结果,最终根据预测数据提出合理减排措施。

#### 1 模型与方法

#### 1.1 差分整合移动平均自回归模型

差分整合移动平均自回归模型通常被称为 ARIMA(autoregressive integrated moxving average) 模型,是由Box和Jenkins于20世纪70年代初提 出的一种时间序列预测分析方法<sup>[18]</sup>。在描述 ARIMA模型时,首先要了解AR模型、MA模型 和ARMA模型。 AR(autoregressive)模型是自回归模型的简称,数学模型表达式如公式(1):

$$X_{i} = c + \sum_{i=1}^{p} \varphi_{i} X_{i-i} + \varepsilon_{i}$$
(1)

其中:  $X_i$ 是当前值; c是常数项; p是阶数;  $\varphi_i$ 是 自相关系数;  $\varepsilon_i$ 被假设为平均数等于0,标准差 等于 $\delta$ 的随机误差值;  $\delta$ 被假设为对于任何的t时 刻都不变。该模型反映了在t时刻的目标值与前  $t - 1 \cong t - q$ 时刻的目标值之间存在线性关系。

MA(moving average)模型是移动平均模型的简称,数学模型表达式如公式(2):

$$X_{i} = c + \sum_{i=1}^{q} \theta_{i} \varepsilon_{i-i} + \varepsilon_{i}$$
(2)

该模型反映了在t时刻的目标值与前t - 1至 t - q时刻的误差值之间存在线性关系。

ARMA(autoregressive moving average)模型是 自回归移动平均模型的简称,描述的是AR模 型与MA模型的结合,具体数学模型表达式如 公式(3):

$$X_{\iota} = c + \sum_{i=1}^{p} \varphi_{i} X_{\iota-i} + \varepsilon_{\iota} + \sum_{i=1}^{q} \theta_{i} \varepsilon_{\iota-i} \qquad (3)$$

ARIMA 模型则是在 ARMA 模型的基础上, 将数据通过差分转化为平稳数据,再将因变量 仅对它的滞后值以及随机误差项的现值和滞后 值进行回归所建立的模型。

模型参数包括*p*—趋势自回归阶数;*d*—趋势差分阶数;*q*—趋势移动平均阶数。其中*p*和*q*由表1确定。

表 1	ARIMA( <i>p</i> ,	d,	q)阶数确定

模型	自相关函数	偏自相关函数
AR(p)	衰减趋于零 (几何型或振荡型)	p阶后截尾
$\operatorname{MA}(q)$	q阶后截尾	衰减趋于零 (几何型或振荡型)
ARMA(p,q)	q阶后衰减趋于零 (几何型或振荡型)	q阶后衰减趋于零 (几何型或振荡型)

#### 1.2 长短期记忆人工神经网络算法

长短期记忆网络(long short term memory, LSTM)是在循环神经网络(recurrent neural network, RNN)的基础上进行改进的一种算法,由 Seep Hochreiter 等学者于 1997 年提出,该算法 可以在一定程度上解决RNN算法的梯度消失和 梯度爆炸问题<sup>[19]</sup>。

作为一种神经网络模型,LSTM 拥有三种类型的门结构:遗忘门、输入门和输出门,这些特殊的"门"结构可以在细胞状态中增加或丢弃信息。LSTM处理信息的流程如下:

首先通过"遗忘门"决定从细胞状态中丢弃的信息。该遗忘门读取前序输出与当前输入, 再通过 sigmoid 函数做非线性映射,之后输出一个维度在[0,1]之间的向量*f*,最后与细胞状态 相乘,表达式如公式(4):

$$f_{t} = \boldsymbol{\sigma} \left( W_{f} \cdot \left[ h_{t-1}, x_{t} \right] + \boldsymbol{b}_{f} \right)$$
(4)

其中: *W<sub>t</sub>*为权值; *h<sub>t-1</sub>*为前序输出; *x<sub>t</sub>*为当前输入; *C<sub>t-1</sub>*为细胞状态; *b<sub>t</sub>*为偏置向量。

其次通过"输入门"确定何种新信息被存 放在细胞状态中。输入门包含了决定和值被更 新的 sigmoid 层和创建新候选向量的 tanh 层,数 学表达式如公式(5)和(6):

$$i_{t} = \sigma \left( W_{i} \cdot \left[ h_{t-1}, x_{t} \right] + b_{i} \right)$$
(5)

$$\tilde{C}_{i} = \tanh\left(W_{c} \cdot \left[h_{i-1}, x_{i}\right] + b_{c}\right)$$
(6)

其中: $i_i$ 为 sigmoid 层确定的更新值; $\tilde{C}_i$ 为 tanh 层 产生的候选值向量。

再次,经过上两步骤后,对细胞状态进行 更新。过去的状态与遗忘门输出的向量 $f_i$ 相乘以 及输入门中的 $i_i \times \tilde{C}_i$ 相加后得到新的候选值, 数学表达式如公式(7):

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{7}$$

最后通过输出门确定需要的输出值。数学 表达式如公式(8)和(9):

$$o_{t} = \sigma \left( W_{o} \left[ h_{t-1}, x_{t} \right] + b_{o} \right)$$

$$\tag{8}$$

$$h_t = o_t \times \tanh(C_t) \tag{9}$$

整个流程如图1所示。



#### 1.3 鲸鱼优化算法

鲸鱼优化算法(WOA)是2016年由澳大利亚 格菲斯大学的Mirjalili等学者提出的一种模拟座 头鲸狩猎行为的元启发式优化算法,该算法采 用随机或最佳搜索代理来模拟捕猎行为,并使 用螺旋模拟座头鲸的气泡网攻击机制,具有机 制简单、参数少、寻优能力强等特点<sup>[20]</sup>。

标准WOA模拟了座头鲸特有的搜索方法和 围捕机制,主要包括:围捕猎物、气泡网捕食 和搜索猎物三个重要阶段。WOA中每个座头鲸 的位置代表一个潜在解,通过在解空间中不断 更新鲸鱼位置,最终获得全局最优解。

#### 1.3.1 围捕猎物

座头鲸的搜索范围是全局解空间,需要先 确定猎物的位置以便包围。由于最优设计在搜 索速度中的位置不是先验已知的,因此WOA算 法假定当前的最佳候选解是目标猎物或接近最 优解。在定义了最佳搜索代理之后,其他搜索 代理将尝试向最佳搜索代理更新它们的位置。 这一行为由公式(10)和(11)表示:

$$D = \left| \boldsymbol{C} \cdot \boldsymbol{X}^{*}(t) - \boldsymbol{X}(t) \right| \tag{10}$$

$$\boldsymbol{X}(t+1) = \boldsymbol{X}^{*}(t) - \boldsymbol{A} \cdot \boldsymbol{D}$$
(11)

其中: *t*表示当前迭代次数; *A*和*C*是系数向量; *X*<sup>\*</sup>(*t*)是目前得到的最佳解的位置向量; *X*(*t*)向 量是位置向量。

如果存在更好的解决方案,那么应该在每 次迭代中更新*X*<sup>\*</sup>(*t*),其中向量*A*和*C*的计算方 式如公式(12)和(13);

$$\mathbf{A} = 2a \times r_1 - a \tag{12}$$

$$C = 2 \times r_2 \tag{13}$$

在整个迭代过程中a由2线性降到0, $r_1$ 和 $r_2$ 是[0,1]中的随机向量。

#### 1.3.2 气泡网捕食

座头鲸捕食主要有收缩包围和气泡网捕食 两种机制。采用气泡网捕食时,座头鲸与猎物 的位置更新用对数螺旋方程表达式如公式(14) 和(15)表示:

$$X(t+1) = D' \times e^{bl} \times \cos(2\pi l) + X^*(t)$$
(14)  
$$D' = |X^*(t) - X(t)|$$
(15)

其中: D'是当前搜索个体与当前最优解的距离; b是螺旋形状参数; l是值域为[-1,1]均匀分布 的随机数。

由于靠近猎物过程中有两种捕食行为,因此 WOA 根据概率来选择气泡网捕食或者收缩包围,位置更新如公式(16)所示:

$$X(t+1) = \begin{cases} X^{*}(t) - A \cdot D, & p < 0.5 \\ D' \times e^{bl} \times \cos(2\pi l) + X^{*}(t), & p \ge 0.5 \end{cases}$$
(16)

其中, *p*为捕食机制选择概率, 值域为[0,1]的随机数。

随着迭代次数*t*的增加,参数A和收敛因子 a逐渐减小,若|A|<1,则各座头鲸逐渐包围当 前最优解,在WOA中属于局部寻优阶段。

#### 1.3.3 搜索猎物

为保证所有座头鲸能在解空间中充分搜索, WOA根据座头鲸彼此之间的距离来更新位置, 达到随机搜索的目的。因此,当|*A*≥1时,搜索 个体会游向随机座头鲸,如公式(17)和(18)表 示:

$$D^{''} = \left| C \cdot \boldsymbol{X}_{\text{rand}}(t) - \boldsymbol{X}(t) \right| \tag{17}$$

$$\boldsymbol{X}(t+1) = \boldsymbol{X}_{\text{rand}}(t) - \boldsymbol{A} \cdot \boldsymbol{D} \tag{18}$$

其中: *D*<sup>"</sup> 为当前搜索个体与随机个体的距离; *X*<sub>and</sub>(*t*)为当前随机个体的位置。

#### 1.4 WOA-LSTM时间序列模型

LSTM的训练效果以及训练过程中的拟合速 度与初始参数的设置密切相关,其中学习率和 隐藏节点数直接影响了神经网络的训练精度和 收敛速度。对于学习率来说,若初始学习率设置 过大,会导致偏离值较大且后期无法拟合,反之 收敛速度会很慢。对于隐藏节点数来说,设置过 少会欠拟合,过多会导致过拟合<sup>[21]</sup>。通过鲸鱼优 化算法,全局向局部搜索寻优,确定最佳学习率 和隐藏节点个数,从而优化神经网络。

统计产品与服务解决方案(SPSS)是世界领 先的统计分析软件之一,广泛应用于自然科学、 技术科学、社会科学等各个领域,具有操作简 单、编程简单、功能强大等特点,世界上许多 有影响的报刊杂志纷纷就 SPSS 的自动统计绘 图、数据的深入分析、使用方便、功能齐全等 方面给予了高度的评价。矩阵实验室(Matlab)是 一款商用数学软件,常用于数据分析、无线通 信、深度学习、图像处理、信号处理等领域, 具有高效的数值计算功能以及丰富的应用工具 箱,为用户提供了大量方便实用的处理工具。

本文将基于SPSS建立ARIMA时间序列预测 模型,同时基于MATLAB建立WOA-LSTM时间 序列预测模型,采用均方根误差(RMSE)、平均 绝对误差(MAE)以及平均绝对误差百分比 (MAPE)作为评估指标来对比分析两种模型的预 测结果。

#### 2 实例分析

本文涉及的中国民航运输业交通与经济数据主要来源于中国民用航空局发展计划司编著的1992—2020《从统计看民航》,其收录了1991—2019年中国民航业发展的统计数据。

#### 2.1 数据预处理

由于不同指标量纲不同,不存在可比性, 故需要将原始数据标准化,以消除量纲的影响。 标准化表达式如公式(19):

$$Z_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \tag{19}$$

其中:X<sub>i</sub>为原始数据,Z<sub>i</sub>为标准化后的数据。

#### 2.2 数据选择

根据中国民用航空局发布的年度《民航行 业发展统计公报》,可以得出航空燃油消耗量与 航空器碳排放之间的关系为1:3.15,即平均每 消耗1吨燃油产生3.15吨二氧化碳排放。本文结 合1991—2019年中国民航统计数据,选取航空 器产生的总碳排放、碳排放强度(单位GDP所排 放二氧化碳)、吨公里碳排放作为分析预测目 标,并将1991—2014年24组数据作为训练集, 2015—2019年5组数据作为测试集,其中训练 集相关数据如表2所示,测试集相关数据如表3 所示。

将中国民航1991—2019年航空器产生的碳 排放量、碳排放强度、吨公里碳排放变化趋势 用折线图表示,依次如图2~图4所示。

表2 1991―2014年中国氏航娠排放相天到	奴抈	Ē
-------------------------	----	---

年份	总碳排放量/吨	碳排放强度/ (吨/万元)	吨公里碳排放/ 千克
1991	4171248.90	3.54	1.301
1992	5346312.30	3.18	1.248
1993	6661194.75	2.89	1.301
1994	7407086.40	2.11	1.268
1995	8550309.60	2.06	1.197
1996	9492628.95	1.97	1.178
1997	10316857.95	1.98	1.190
1998	11893848.75	2.34	1.279
1999	12202183.35	1.93	1.150
2000	15565214.70	2.19	1.271
2001	16870395.15	1.92	1.195
2002	18902349.90	1.64	1.146
2003	19053650.39	1.50	1.116
2004	24846335.33	1.47	1.076
2005	27659374.47	1.33	1.059
2006	31516934.40	1.19	1.031
2007	35591575.95	1.17	0.974
2008	36998211.60	1.06	0.982
2009	41396414.85	1.17	0.969
2010	48239084.25	0.95	0.896
2011	51858018.45	0.81	0.898
2012	56291426.10	0.82	0.922
2013	62940874.50	0.92	0.937
2014	69803399.61	0.95	0.933

#### 表 3 2015—2019年中国民航碳排放相关数据

年份	总碳排放量/吨	碳排放强度/ (吨/万元)	吨公里碳排放/ 千克
2015	78902381.57	1.17	0.926
2016	88669360.08	1.20	0.921
2017	99960833.39	1.16	0.923
2018	109068713.80	1.06	0.904
2019	116209582.70	1.07	0.899



图 2 1991—2019年航空器产生的碳排放量趋势图







可以看出, 1991-2019年航空器产生的碳 排放呈逐年增加的趋势,且过去十年为快速增 长期; 而碳排放强度和吨公里碳排放整体处于 下降趋势,碳排放强度相较于吨公里碳排放过 去三十年变化幅度较大,且近十年稍有起伏, 吨公里碳排放近年处于稳中带降的趋势。以上 三项数据与我国大力发展民航业和民航业技术 变革等因素密切相关, 民航业的飞速发展带来 航班量与飞行小时数的增加,从而航空燃油消 耗量增加,进一步导致碳排放量增加;民航业 新技术的应用使燃油效率提高,因此吨公里碳 排放量整体处于下降态势。

#### 预测结果与讨论 3

#### 3.1 基于ARIMA模型预测结果

首先对1991-2019年航空器碳排放量进行 自相关(ACF)、偏自相关(PACF)分析,如图5 所示, ACF出现拖尾, 即相关数据不具备平稳 性,对其进行二阶差分后,ACF与PACF均出现 截尾,如图6所示,此时数据具有平稳性,可进 行预测分析,此时模型为ARIMA(1,2,1)。同 理,对碳排放强度、吨公里碳排放分析后模型 分别为ARIMA(0,1,0)与ARIMA(1,1,1)。





4

6

#### 差分前航空器碳排放 ACF 与 PACF 分析 图 5





其次对三项数据预测至2035年,依次如图 7-图9所示。经过ARIMA模型预测,航空器碳 排放在未来十六年一直保持增长态势,2035年 碳排放将达到3.88亿吨,未见"达峰"出现; 碳排放强度于2020年后持续上升,2035年将上 升至与1991年持平水平;吨公里碳排放则出现 稳定下降趋势,至2035年吨公里碳排放仅为 0.804千克。



#### 图 7 ARIMA 模型航空器碳排放演化趋势图 (1991—2035年)









最后模型对应 RMSE、MAE 以及 MAPE 如表 4 所示。

表4 ARIMA	奠型 RMSE、	MAE、MAF	ΡE统计
----------	----------	---------	------

模型	RMSE	MAE	MAPE
航空器碳排放	163.531	109.492	3.731
碳排放强度	0.209	0.146	9.178
吨公里碳排放	0.046	0.033	2.993

#### 3.2 基于WOA-LSTM模型预测结果

设置 WOA 算法中种群个数为 20,迭代 10 步后得到最优的 LSTM 学习率与隐藏节点数,不 仅提高了参数选择与调整的效率,同时提高了 预测模型的准确度。通过计算机深度学习的方 法,基于 WOA-LSTM 建立的三项数据的模型测 试集拟合效果如图 10所示。



由图 10可知,所建时间序列预测模型拟合 度良好,使用该模型对 2020—2035 年数据进行 预测,演化趋势如图 11~图 13 所示。基于 WOA-LSTM模型预测的航空器碳排放量虽同样 为上升趋势,但增加较为平缓,甚至出现"达 峰"趋势;碳排放强度与吨公里碳排放变化都 较为平缓,吨公里碳排放在未来出现细小波动。



#### 图 11 WOA0-LSTM 模型航空器碳排放演化趋势图 (1991—2035年)



#### 图 12 WOA-LSTM模型碳排放强度演化趋势图 (1991—2035年)



#### 图 13 WOA-LSTM模型吨公里碳排放演化趋势图 (1991—2035年)

同样, 计算基于 WOA-LSTM 所建立三种模

型的RMSE、MAE以及MAPE值,如表5所示。

表 5 WOA-LSTM 模型预测结果

模型	RMSE	MAE	MAPE	
航空器碳排放	570.881	519.805	5.038	
碳排放强度	0.007	0.007	0.912	
吨公里碳排放	0.005	0.004	0.328	

#### 3.3 模型对比分析

将 ARIMA 模型与 WOA-LSTM 模型进行对 比,可以看出 ARIMA 模型对航空器碳排放的预 测比 WOA-LSTM 模型更加精准,而对于碳排放 强度及吨公里碳排放的预测略逊于 WOA-LSTM 模型,尤其是碳排放强度预测的 MAPE 值为 9.173,明显不如 WOA-LSTM 模型准确。综合来 看,经过 WOA 优化的 LSTM 时间序列预测模型 的预测效果好于传统 ARIMA 模型。

#### 4 结语

本文使用SPSS与MATLAB建立了ARIMA及WOA-LSTM时间序列预测模型,用于对中国民航业航空器碳排放、碳排放强度等数据的未来预测,其中WOA-LSTM使用鲸鱼优化算法对LSTM模型中的学习率和隐藏节点数进行优化,避免了人为确定参数的主观性和盲目性。通过对ARIMA与WOA-LSTM模型的对比,ARIMA模型在航空器碳排放以及吨公里碳排放预测中具有良好的效果,WOA-LSTM模型则在碳排放强度和吨公里碳排放的预测中有更为优越的拟合度和准确性。

根据预测结果可知,在当前政策没有较大 变动的情况下,未来民航业碳排放将会持续上 升,这将对我国在环境保护方面的总体规划造 成较大冲击,为我国完成2030 "碳达峰"目标 造成一定的困难。同时,由于行业特殊性,民 航业技术迭代周期长、新技术应用慢等特点, 也给民航业的节能减排工作带来很大困难。基 于本文预测结果与民航业特点,总结出以下几 点民航业节能减排方面的建议:

(1) 调整优化机队规模结构。从1991—2019年

吨公里碳排放数据的演化可以看出,老旧机型 相较于新机型在燃油效率方面有一定劣势,应 加快撤下高排放老旧飞机。

(2)提升空管运行效率,优化航班航线。积极推进空管新技术研发应用以及飞行程序新标准的制定,全面提升航班整体运行效率。

中国是一个发展中的人口大国,根据 《"十四五"民用航空发展规划》的部署,我 国民航业未来还将继续高速发展,随之而来的 二氧化碳排放问题值得特别关注。本文所建立 的碳排放时间序列模型可以对相关研究提供一 定的参考,同时我们将继续关注民航业绿色发 展等问题,为后续研究做好准备。

#### 参考文献:

- [1] 习近平.继往开来,开启全球应对气候变化新征程:
   在气候雄心峰会上的讲话[N].北京:人民日报,
   2020-12-13(1).
- [2] WEI S W, WANG T, LI Y B. Influencing factors and prediction of carbon dioxide emissions using factor analysis and optimized least squares support vector machine [J]. Environmental Engineering Research, 2017,22(2):175-185.
- [3] FARUQUE O, RABBY A J, HOSSAIN A, et al. A comparative analysis to forecast carbon dioxide emissions[J]. Energy Reports, 2022(8):8046-8060.
- [4] 陈文婕,吴小刚,肖竹.中国四大经济区域道路交通 碳排放预测与减排潜力评估:基于私家车轨迹数据 的情景模[J].经济地理,2022,42(7):44-52.
- [5] 高金賀,郑宝珠,周伟昊,等.基于GA-SVR 的城市 交通运输碳排放预测研究[J].东华理工大学学报 (自然科学版),2022,45(3):269-274.
- [6] 李宝成,王芳,彭晨,等.以碳达峰引导河南省交通运输减排降碳研究[J].物流科技,2022,45(7): 72-75.
- [7] 韩博,邓志强,于敬磊,等.碳达峰目标下中国民航 CO2与NOx减排协同效益分析[J].交通运输系统 工程与信息,2022,22(4):53-62.
- [8] 胡荣,王德芸,冯慧琳,等.碳达峰视角下的机场航

空器碳排放预测[J]. 交通运输系统工程与信息, 2021,21(6):257-263.

- [9] 许继辉,王克.中国民航业中长期碳排放预测与技术减排潜力分析[J].中国环境科学,2022,42(7): 3412-3424.
- [10] 李心怡,赵瑞嘉,高成男,等.中国民航运输碳排放 解耦分析和达峰预测[J].环境污染与防治,2022, 44(6):729-733,739.
- [11] 张金良, 贾凡. 中国火电行业多模型碳达峰情景预测[J]. 电力建设, 2022, 43(5):18-28.
- [12] 徐丽,曲建升,李恒吉,等.中国居民能源消费碳排 放现状分析及预测研究[J].生态经济,2019,35
   (1):19-23,29.
- [13] 张昱,陈广书,李继涛,等.基于Attention机制的 CNN-LSTM时序预测方法研究与应用[J].内蒙古 大学学报(自然科学版),2022,53(5):516-521.
- [14]魏光普,康瑜,范浩文,等.重工业城市建筑业碳排 放核算与预测研究[J].生态经济,2022,38(9): 43-48.
- [15] 胡剑波,赵魁,杨苑翰.中国工业碳排放达峰预测及 控制因素研究:基于BP-LSTM神经网络模型的实 证分析[J].贵州社会科学,2021(9):135-146.
- [16] YANG H J, O' CONNELL J F. Short-term carbon emissions forecast for aviation industry in Shanghai [J]. Journal of Cleaner Production, 2020(275):1-12.
- [17] 柴建,张钟毓,李新,等.中国航空燃油消费分析及 预测[J].管理评论,2016,28(1):11-21,41.
- [18] 朱丹青.基于数据挖掘的供应链需求预测模型研究[D].武汉:华中科技大学,2020.
- [19] 张丽婷,李鹏飞,庞文静,等.基于季节性自回归积 分滑动平均与深度学习长短期记忆神经网络的降 水量预测[J].科学技术与工程,2022,22(9): 3453-3463.
- [20] 闫旭, 叶春明, 姚远远. 量子鲸鱼优化算法求解作业
   车间调度问题[J]. 计算机应用研究, 2019, 36(9):
   975-979.
- [21] 苏鹏飞,徐松毅,于晓磊.基于WOA-LSTM的窄带 通信网网络时延预测算法[J].河北工业科技, 2022,39(1):9-15.

(下转第67页)

DOI: 10.3969/j.issn.1007-1423.2023.02.003

### 一种改进DeepLabV3+的岩屑图像语义分割算法

罗崇兴1,师明元2,王正勇1\*,滕奇志1

(1.四川大学电子信息学院,成都 610065; 2.河北省地质矿产勘查开发局第六地质大队,石家庄 050080)

摘要:岩屑图像的准确分割与识别是地质勘探人员研究地层特征和地质建模的重要基础和保障。针对 DeepLabV3+在岩屑图像语义分割时存在的推理速度慢、分割效果不理想的问题,提出了一种改进的Deep-LabV3+岩屑图像语义分割模型。首先,利用轻量级 MobileNetV3提取岩屑图像语义特征,以减少参数量和 计算量;其次,采用结合注意力机制的空洞空间卷积池化金字塔模块(atrous spatial pyramid pooling, ASPP)以 更好地拟合岩屑图像特征;然后,借鉴特征金字塔思想将编码过程提取的不同尺度特征融合至解码过程, 从而获取更精准的岩屑图像语义分割结果;最后,在自制的岩屑图像数据集上和现有的7种方法进行对比实 验,实验结果表明,改进的 DeepLabV3+模型的类别平均像素准确率为0.78,平均交并比为0.68,相比于其 他的网络模型具有更好的分割表现。

关键词:岩屑图像;语义分割;DeepLabV3+;注意力机制

#### 0 引言

在矿产资源勘探和开发工作中,岩屑录井 是非常重要的一步工作,对岩屑进行准确识别 是地质勘探人员研究地层特征和地质建模的重 要基础和保障。在岩屑采集现场,多种因素的 影响,使得采集过程中获取到的岩屑并不只是 一种岩性,而是混合着多种岩性。为更好地分 析岩屑岩性,需要对岩屑图像进行语义分割。 随着科学技术的发展,深度学习开始广泛应用 于岩屑图像语义分割,深度学习的引入加快了 岩屑图像处理速度,提高了岩屑识别的准确率。

自2012年,AlexNet模型<sup>[1]</sup>在ImageNet比赛 上大放光彩,在那之后,深度学习网络模型层 出不穷。2014年,Simonyan等<sup>[2]</sup>提出VGG网络, 采用连续的小尺寸卷积核代替了较大卷积核, 却存在需要计算更多的参数,对内存和时间要 求高的问题。之后,Long等<sup>[3]</sup>在VGG-16网络原 有结构基础上提出全卷积网络(fully convolutional networks, FCN),采用卷积层代替了传统 卷积中的全连接层, 使图像能进行像素级别的 语义分割,但是还存在分割结果粗糙等问题。 Chen 等<sup>[4]</sup>针对语义分割中存在的池化导致信息 丢失,标签之间的概率关系未利用的问题,基 于FCN网络提出了DeepLab V1网络,在该网络 中引入了全连接条件随机场(conditional random field, CRF),以改善原始分割结果不精细的问 题,同时利用空洞卷积使网络在不改变参数量 和计算量的情况下扩大网络感受野,获取图像 更多的信息。在那之后, DeepLab 系列不断发 展, 衍生出 DeepLab V2<sup>[5]</sup>, DeepLab V3<sup>[6]</sup>等网络 模型。2015年, Ronneberger等<sup>[7]</sup>基于对称的编 码器-解码器结构提出了U-Net网络,以方便融 合分辨率相同的特征。2015年,剑桥大学团队 借鉴 FCN 网络和 U-Net 网络, 提出 SegNet 模 型<sup>[8]</sup>,该模型的编码结构构建了最大池化层索 引存储,以内存为代价换取轻微的精度损失,

**收稿日期:** 2022-10-08 修稿日期: 2022-11-02

基金项目: 国家自然科学基金资助项目(62071315)

作者简介:罗崇兴(1997—),男,云南西双版纳人,硕士研究生,研究方向为电路与系统、计算机视觉;师明元(1987—),男,河北冀州人,硕士,工程师,研究方向为矿产地质;\*通信作者:王正勇(1969—),女,四川达州人,博士,硕士 生导师,从事领域为图像处理与模式识别、计算机视觉、智能系统等,E-mail:690728634@sina.com;膝奇志(1961—),女,四 川成都人,博士,教授,研究方向为智能信息系统、信号与信息处理

解决边界信息丢失问题。2017年,香港大学和 商汤科技联合提出PSPNet网络<sup>[9]</sup>,提出了金字 塔池化模型,融合不同尺度和不同区域之间的 信息,通过全局先验信息有效获取高质量的语 义分类结果。Lu等<sup>[10]</sup>首次将图卷积模型应用于 分类,通过卷积特征图构建图结构,并将其转 化为图节点分类问题,解决了卷积神经网络在提 取特征的过程中局部位置信息损失的问题。北京 大学联合商汤科技基于类的动态图卷积自适应提 出CDGCNet模型<sup>[11]</sup>,利用构造图的动态图卷积结 果学习特征聚集和权重分配,融合原始特征和精 炼特征获取最终预测。He等<sup>[12]</sup>提出了一种基于协 作学习的多源领域框架,将多标记源模型适应到 无标记源中,并用于语义分类。

在岩屑分割领域,2020年,万川等<sup>[13]</sup>对U-Net 模型进行了改进,并应用于岩屑图像分割中, 运用金字塔池化模块聚合不同区域的上下文特征 信息,更好地利用了全局信息。2022年,严良平 等<sup>[14]</sup>在VGG16的基础上提出了一种深度图像引导 的岩石颗粒分割算法,利用深度图像的三维距离 信息,提高了岩石图像的分割精度。

为更好地对岩屑图像进行语义分割,本文 在DeepLabV3+算法上进行改进,并提出了一种 改进DeepLabV3+的岩屑图像语义分割算法。

#### 1 相关算法原理

DeepLabV3+算法采用编码器-解码器结构。 编码器主要用于岩屑图像特征提取,解码器是将 编码器提取的特征映射到高维空间,以实现图像 像素级别的语义分类,通过反卷积操作不断恢复 图像的空间维度,从而实现图像语义分割。

DeepLabV3+的网络结构如图1(a)所示,其 编码器部分由Xception<sup>[15]</sup>和ASPP模块<sup>[5]</sup>组成, Xception是利用若干个大小不同的卷积提取输入 特征,在减小计算量的同时获取不同感受野, 经过Xception模型,图像分辨率降低为原始图 像的1/16,之后再送入ASPP模块。ASPP模块 如图1(b)所示,由不同采样率的空洞卷积并联 组成,将结果融合在一起之后,利用1×1的卷 积降低输出的通道数,以实现用多个比例捕捉 图像上下文信息的目的。DeepLabV3+网络的解 码部分,首先是将输出的特征图进行上采样操 作,将特征图尺寸扩大四倍,之后将该特征图 和低级特征进行拼接融合,最后再进行上采样, 实现图像语义分割,得到最终预测结果。

#### 2 改进的 DeepLabV3+

#### 2.1 MobileNetV3

原始的 DeepLabV3+模型使用 Xception 特征 提取网络作为主干网络,但 Xception 对模型的 参数规模和运算速度控制不佳,使得在进行岩 屑图像语义分割的过程中存在参数量大、推理 速度慢的问题,所以本文采用 MobileNetV3<sup>[5]</sup>代 替原网络中的 Xception 模块来提取岩屑图像特 征,在不降低模型精度的同时提高模型的速度。 该模型借鉴了 MobileNetV1<sup>[16]</sup>的深度可分离卷 积、MobileNetV2<sup>[17]</sup>的具有线性瓶颈的逆残差结 构,并且引入了 Squeeze and excitation<sup>[18]</sup>的轻量 级注意力结构,采用了一种新的非线性激活函



数h-swish, 计算公式如下:

h-swish [x] = 
$$x \frac{\text{ReLU6}(x+3)}{6}$$
 (1)

该函数能减少运算量并提高模型性能。

MobileNetV3 网络模型的基本模块 bneck 结构如图 2 所示,首先利用1×1的卷积进行升维操作,之后经过5×5的深度可分离卷积提取图像特征,然后利用注意力机制调整每个通道的权重,最后再通过1×1的卷积降维。



#### 图 2 bneck 结构

#### 2.2 注意力机制 EPSANet

在神经网络中,模型学习能力越强,需 要存储的信息也就越大,这会引起信息超载。 为了更好地引导模型聚焦于岩屑图像中的重 要信息,本文在ASPP模块引入了注意力机制 EPSANet<sup>[19]</sup>。如图3所示,EPSANet能够有效捕 获不同尺度特征图的空间信息,同时丰富特征 空间,建立长期依赖关系,进而学习更丰富的 特征表示。

EPSANet首先将输入特征图*X*拆分为*S*个部分( $X_0, X_1, \dots, X_{s-1}$ ),然后利用多尺度卷积核分组卷积提取不同尺度特征图的空间信息,计算公式如下:

$$F_i = \operatorname{Conv}(K_i \times K_i, G_i)(X_i)$$
(2)

然后将这些特征图拼接起来,计算公式如下:

$$F = \operatorname{Cat}(F_0, F_1, \cdots, F_{S-1})$$
(3)

最后对这些拥有不同尺度信息的特征图进 行通道注意力权重加权,计算公式如下:

$$Z_i = \text{SEWeight}(F_i) \tag{4}$$

为更好地交互多尺度通道信息,利用Softmax 进一步标定权重信息,计算公式如下:



图 3 EPSANet

$$att_i = \text{Softmax}(Z_i) = \frac{\exp(Z_i)}{\sum_{i=0}^{s-1} \exp(Z_i)}$$
(5)

最后将对应的权重和特征图进行通道级别的相乘,计算公式如下:

$$Y_i = F_i \odot att_i \tag{6}$$

在ASPP模块中引入EPSANet得到EPSANet-ASPP模块,如图4所示,能够引导模型在融合 特征时更好地聚焦于岩屑图像中的重要信息, 降低对无关信息的关注。



图 4 EPSANet-ASPP 模块

#### 2.3 改进后的网络模型

改进 DeepLabV3+的岩屑图像语义分割算法 的网络结构如图 5(a)所示,该模型的编码器部 分由 MobileNetV3 和 EPSANet-ASPP 模块组成, 首先经过 MobileNetV3 提取岩屑图像特征,在不 降低模型精度的同时提高模型的速度,以得到 不同尺度的图像特征;然后将最小尺寸的图像 特征送入EPSANet-ASPP模块,EPSANet-ASPP模 块如图5(b)所示,引入EPSANet机制的EPSANet-ASPP模块能够帮助网络更好地提取出特征中的 重要信息,从而更高效地捕获岩屑图像中不同 尺度信息。

由于在特征提取的过程中,低层次的特征 包含较多的细节信息,更容易反映图像特征, 同时目标位置更加准确,但是缺乏语义特征信 息。高层次特征具有更多的语义信息,却只有 较少的细节信息和位置信息。有效融合低级特 征和高级特征能更有助于岩屑图像语义分割, 因此,在解码器部分,本文借鉴 FPN<sup>[20]</sup>的多尺 度特征融合思想,在编码器和解码器之间构建 横向连接,使模型在对特征图进行上采样时能 够融合编码器中 MobileNetV3 提取的不同尺度特 征,进而获取更多的图像细节信息和更准确的 目标位置信息,得到更精细的岩屑图像语义分 割结果,实现岩屑图像语义分割。

#### 3 实验结果及分析

为更好地评估本文算法性能,本文介绍了 几种常用的语义分割评价指标,在相同的实验 配置条件下,对比分析了7种不同的语义分割网 络在岩屑图像语义分割数据集上的表现,分别 为Danet<sup>[21]</sup>、FCN<sup>[3]</sup>、DeepLabV3+<sup>[22]</sup>、PSPNet<sup>[9]</sup>、 UNet<sup>[7]</sup>、UPerNet<sup>[23]</sup>、HRNet<sup>[24]</sup>。实验证明,本



(a)改进后的DeepLabV3+网络结构

(b)ERSANet-ASPP模块

图 5 改进后的 DeepLabV3+模型及 ERSANet-ASPP 模块图

文算法具有更高的语义分割精度,优于其他对 比网络。

#### 3.1 实验环境

本实验在 Linux 操作系统下进行,实验使用 显卡型号为 NVIDIA GeForce RTX 2080 Ti,处理 器为 Inter(R) Core(TM) i7-9700 CPU,使用的编 程语言是 Python3.8.0,深度学习框架为 PyTorch 1.8.0, CUDA Version为11.1,内存为32 GB。

#### 3.2 实验数据集

为验证本文算法的有效性,首先利用岩屑 图像采集设备采集岩屑图像,得到尺寸为4000× 3750的原始岩屑图像,部分原始岩屑图像如图6 所示,共采集岩屑图像126张,然后通过图像裁 剪、数据增强等方法将原始岩屑图像制作成 512×512的岩屑图像语义分割数据集共9682张, 其中7746张用于训练模型,1936张用于测试模 型。实验数据集共有20个类别,其中包含背景和 19种岩屑样本,岩屑种类分别是油斑粉砂岩、紫 红色安山岩、深灰色安山岩、灰绿色安山岩、褐 色流纹岩、灰绿色砂岩、含砾粗砂岩、紫灰色泥 质粉砂岩、灰色粉砂岩、褐色粉砂岩、灰绿色凝 灰岩、方解石、石灰岩、页岩、石英、灰泥岩、 灰黑色泥岩、深灰黑色泥岩、棕红泥岩,部分岩 屑样本展示如图7所示。



图 6 原始岩屑图像



#### 图 7 部分岩屑样本展示

3.3 实验结果及分析

#### 3.3.1 评判标准

语义分割是对图像进行像素级别的分类,

常用的评价指标有像素准确率(pixel accuracy, PA)、类别平均像素准确率(mean pixel accuracy, mPA)、交并比(intersection over union, IoU)和 平均交并比(mean intersection over union, mIoU)。

PA是指类别预测正确的像素数目占总像素的比例,计算公式如下:

$$PA = \frac{C_{ii}}{\sum_{i=1}^{20} C_{ii}}$$
(7)

*mPA* 是每个类别的*PA* 的求和再平均的值, 计算公式如下:

$$mPA = \frac{1}{20} \times \sum_{i=1}^{20} \frac{C_{ii}}{\sum_{i=1}^{20} C_{ii}}$$
(8)

*IoU*是类别预测结果和真实结果的交集与并 集的比值,计算公式如下:

$$IoU = \frac{C_{ii}}{\sum_{j=1}^{20} C_{ij} + \sum_{i=1}^{20} C_{ij} - C_{ii}}$$
(9)

*mloU*是每个类别*loU*的和再平均的结果, 计算公式如下:

$$mIoU = \frac{1}{20} \times \sum_{i=1}^{20} \frac{C_{ii}}{\sum_{j=1}^{20} C_{ij} + \sum_{i=1}^{20} C_{ij} - C_{ii}} \quad (10)$$

其中, *C<sub>ij</sub>*代表在岩屑图像中预测分类为*i*类、真 实分类为*j*类的总像素数,如当*i*=*j*,代表在岩 屑图像中像素点的预测为*i*类,同时实际值也是 *i*类,这类结果表示为真;当*i*!=*j*,代表在岩屑 图像中像素点的预测是*j*类,但是实际分类值为 *i*类,这类结果为假。

#### 3.3.2 消融实验结果与分析

为验证改进的 DeepLabV3+岩屑图像语义分 割模型的有效性,分别对 MoblieNetV3模块,注 意力机制 EPSANet 模块和编解码联系模块进行 消融实验。选取 DeepLabV3+网络模型作为基线 模型。在自制的岩屑数据集上性能对比结果如 表1所示。

从表1可以看出, 原始的DeepLabV3+网络 在自制岩屑数据集上的评价指标 mPA为0.68, mIoU仅有0.57; 用MoblieNetV3 替换Xception进 行特征提取后, mPA比基线模型提升了0.05, mIoU比基线模型提升了0.03, 参数量减少了 10.17 M, 速度提升了1.75 s; 在ASPP中加入注 意力机制EPSANet后, mPA比基线模型提升了 0.02, mIoU比基线模型提升了0.04; 在改进了 编解码联系模块后, mPA比基线模型提升了 0.03, mIoU比基线模型提升了 0.06。实验表明, 本文所提出的几个改进方法在性能上相较于原 始的 DeepLabV3+网络均有不同程度的提高。

表 1	消融实验的 mPA、mloU、参数量、单张图像处理
	时间对比

模型	mPA	mIoU	Parms/M	处理时间/s
Baseline	0.68	0.57	43.69	5.80
Baseline + MoblieNetV3	0.73	0.60	33.52	4.05
Baseline + EPSANet	0.70	0.61	45.31	5.95
Baseline + 编解码联系	0.71	0.63	44.97	6.27
本文模型	0.78	0.68	37.26	5.31

#### 3.3.3 与其他模型的比较

表2是不同算法在自制的岩屑图像语义分割 数据集下性能对比结果,其中加粗数字是横向 比较最优结果。从表2可以看出本文算法在岩屑 图像数据集上的 mPA 为 0.78, mloU 为 0.68,分 割的整体性能优于其他对比网络,具有更高的 分割精度。

为更直观地比较不同算法的性能表现,图8 展示了不同算法在岩屑图像上语义分割结果。 对比了7种不同的语义分割算法,图中左上方框 中的岩屑是灰绿色砂岩,从图中可以看出Deep-LabV3+、UNet、UperNet、HRNet算法将其错误 识别为灰绿色安山岩。右下方框中的岩屑是深灰 色安山岩,Danet、FCN、DeepLabV3+、UNet、 UperNet算法未能很好地区分该岩屑和背景。从 图 8 可以看出,本文改进的DeepLabV3+算法语 义分割结果优于对比算法。

图9展示了不同算法在密集岩屑图像上的语 义分割结果,图中方框中的岩屑是灰绿色凝灰 岩,Danet、DeepLabV3+、PSPNet、UNet、Uper-Net、HRNet算法将该岩屑的部分像素错误分类 为灰绿色安山岩。从图中可以看出,本文算法 性能优于其他对比算法。

					1			
				<b></b> 侯型				
	Dane <sup>[22]</sup>	FCN <sup>[3]</sup>	DeepLab V3+ $\lfloor 23 \rfloor$	PSPNet <sup>[9]</sup>	UNet <sup>[7]</sup>	UPerNet <sup>[24]</sup>	HRNet <sup>[25]</sup>	本文算法
背景	0.84	0.84	0.83	0.84	0.82	0.80	0.85	0.88
类别一	0.62	0.46	0.66	0.53	0.67	0.56	0.65	0.64
类别二	0.56	0.63	0.40	0.68	0.53	0.48	0.53	0.61
类别三	0.73	0.49	0.41	0.50	0.52	0.47	0.76	0.63
类别四	0.62	0.56	0.59	0.39	0.46	0.36	0.61	0.63
类别五	0.42	0.63	0.68	0.53	0.49	0.56	0.50	0.65
类别六	0.75	0.56	0.55	0.67	0.70	0.59	0.74	0.71
类别七	0.60	0.77	0.75	0.72	0.60	0.66	0.64	0.78
类别八	0.55	0.48	0.26	0.58	0.40	0.59	0.56	0.46
类别九	0.65	0.58	0.52	0.66	0.50	0.48	0.65	0.55
类别十	0.67	0.62	0.42	0.68	0.63	0.61	0.69	0.57
类别十一	0.59	0.46	0.44	0.54	0.35	0.43	0.44	0.56
类别十二	0.34	0.48	0.29	0.31	0.23	0.33	0.26	0.39
类别十三	0.45	0.56	0.65	0.62	0.42	0.52	0.79	0.77
类别十四	0.85	0.87	0.80	0.70	0.88	0.72	0.85	0.91
类别十五	0.58	0.42	0.54	0.60	0.56	0.39	0.61	0.77
类别十六	0.70	0.58	0.60	0.65	0.46	0.56	0.63	0.74
类别十七	0.46	0.36	0.68	0.40	0.50	0.46	0.54	0.79
类别十八	0.64	0.50	0.72	0.54	0.62	0.51	0.60	0.85
类别十九	0.75	0.77	0.66	0.45	0.54	0.53	0.44	0.71
mPA	0.71	0.65	0.68	0.70	0.67	0.61	0.70	0.78
mIoU	0.62	0.58	0.57	0.58	0.54	0.53	0.62	0.68

表 2 不同算法的性能对比



(a)原图



(f)PSPNet



(b)真实结果

(g)UNet



(c)Danet



(h)UPerNet



(d)FCN







(e) Deep Lab V3+



(g)本文算法

图 8 不同模型的语义分割对比图组1



图 9 不同模型的语义分割对比图组2

#### 4 结语

本文针对岩屑图像语义分割问题,利用岩 屑采集设备采集岩屑图像制成数据集,并提出 了基于 DeepLabV3+算法的改进方法。本文算法 首先采用 MobileNetV3作为主干网络,以快速提 取岩屑图像不同尺度的特征,接着采用融合了 EPSANet的 ASPP 网络,用于有效获取岩屑图像 的重要信息并更好地融合图像特征,最后借鉴 FPN 思想,在解码过程中融合编码提取的不同 尺度特征,从而得到更精准的岩屑图像语义分 割结果。实验表明,本文算法在自制的岩屑数 据集上 mPA 为 0.78, mIoU 为 0.68,分割的整体 效果优于其他对比网络,岩屑颗粒的边缘分割 也更加精确。

#### 参考文献:

- [1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017,60(6):84-90.
- [2] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. arXiv Preprint arXiv:1409.1556,2014.
- [3] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional networks for semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.

- [4] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs [J]. Computer Science, 2014(4):357-361.
- [5] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4):834-848.
- [6] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation [EB/OL]. arXiv Preprint arXiv: 1706. 05587,2017.
- [7] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C] //Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015:234-241.
- [8] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(12):2481-2495.
- [9] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2881-2890.
- [10] LU Y, CHEN Y, ZHAO D, et al. Graph-FCN for image semantic segmentation [C] //Proceedings of the International Symposium on Neural Networks. Springer, Cham, 2019:97-105.

- HU H, JI D, GAN W, et al. Class-wise dynamic graph convolution for semantic segmentation [C] // Proceedings of the European Conference on Computer Vision. Springer, Cham, 2020; 1-17.
- [12] HE J, JIA X, CHEN S, et al. Multi-source domain adaptation with collaborative learning for semantic segmentation [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 11008-11017.
- [13] 万川,王正勇,何海波,等.基于改进P-Unet模型的 岩屑颗粒识别[J].信息技术与网络安全,2020,39 (11):56-61.
- [14] 严良平,潘月梁,姜雄彪,等.深度图像引导的岩石 颗粒分割方法[J].应用科技,2022,49(2):87-93.
- [15] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251-1258.
- [16] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3 [C] //Proceedings of the IEEE/ CVF International Conference on Computer Vision, 2019:1314-1324.
- [17] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. arXiv Preprint arXiv:1704.04861,2017.
- [18] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: inverted residuals and linear bottlenecks
   [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510 -4520.

- [19] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [20] ZHANG H, ZU K, LU J, et al. EPSANet: an efficient pyramid squeeze attention block on convolutional neural network [EB/OL]. arXiv Preprint arXiv:2105.14447,2021.
- [21] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:2117-2125.
- [22] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation [C] // Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.
- [23] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C] // Proceedings of the European Conference on Computer Vision, 2018: 801-818.
- [24] TETE XIAO, YINGCHENG LIU, BOLEI ZHOU, et al. Unified perceptual parsing for scene understanding [C] //Proceedings of the European Conference on Computer Vision, 2018.
- [25] SUN K, XIAO B, LIU D, et al. Deep High-Resolution Representation Learning for Human Pose Estimation [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019:5693-5703.

### An improved cuttings image semantic segmentation algorithm based on DeepLabV3+

Luo Chongxing<sup>1</sup>, Shi Mingyuan<sup>2</sup>, Wang Zhengyong<sup>1\*</sup>, Teng Qizhi<sup>1</sup>

(1. College of Electronics and Information Engineering, Sichuan University, Chengdu 610065;2. The Sixth Geological Brigade of Hebei Bureau of Geology and Mineral Exploration and Development, Shijiazhuang 050080)

Abstract: Accurate segmentation and recognition of cuttings image is an important basis and guarantee for geological explorers to study stratigraphic characteristics and geological modeling. An improved DeepLabV3+ cuttings image semantic segmentation model is proposed to solve the problems of slow reasoning speed and unsatisfactory segmentation effect in DeepLabV3+ cuttings image semantic segmentation model. Firstly, lightweight MobileNetV3 is used to extract the semantic features of cuttings images to reduce the number of parameters and computational cost. Secondly, Atrous spatial pyramid pooling (ASPP) module combined with attention mechanism is used to better fit the cuttings image features. Then, the different scale features extracted from the encoding process are fused to the decoding process by using the feature pyramid idea, so as to obtain more accurate semantic segmentation results of cuttings images. Finally, compared with the existing 7 methods on the self-made cuttings image dataset, the experimental results show that the improved DeepLabV3+ model has a category average pixel accuracy of 78% and an average intersection ratio of 68%. Compared with other network models, it has better segmentation performance.

Keywords: cuttings image; semantic segmentation; DeepLabV3+; attention mechanism

文章编号:1007-1423(2023)02-0032-09

DOI: 10.3969/j.issn.1007-1423.2023.02.004

### 融合多尺度卷积的端到端宫颈细胞分割

王文涛<sup>1,2</sup>, 王嘉鑫<sup>1,2</sup>, 张 根<sup>1,2</sup>, 陈大江<sup>1,2</sup>

(1. 中南民族大学计算机科学学院, 武汉 430074; 2. 湖北省制造企业智能管理工程技术研究中心, 武汉 430074)

摘要: 宫颈癌是目前唯一一种病因明确的妇科恶性肿瘤,通常采用宫颈细胞筛查来进行早期排查以及 治疗。在细胞筛查过程中,为了发现早期的宫颈异常细胞,需要从显微细胞图像中准确分割出细胞核与细 胞质。现有宫颈细胞分割方法存在计算量大、精确率低、数据不平衡所导致的学习困难等问题。为了提高 宫颈细胞分割算法的性能,引入了多尺度卷积的思想,采用一种U型编码器-解码器模型,设计了一个端到 端的细胞分割算法IR U-Net。主要包括: ①利用多尺度卷积结构加宽网络,避免人工选择卷积核,能提取 多尺度特征,同时在卷积结构中加入残差连接,来减少梯度消失等现象; ②通过使用Leaky-ReLU减少"神 经元死亡"导致的网络稀疏特征多、难以收敛的问题; ③采用改进的损失函数Focal-Dice Loss 以缓解数据不 平衡的问题。仿真实验结果表明,改进后的模型相比对照算法在精度上有所提高,分割性能得到改善。

关键词: 宫颈细胞; 图像分割; 端到端; 多尺度卷积; 损失函数

#### 0 引言

宫颈癌是女性最常见的恶性肿瘤之一,据 世界卫生组织估计,目前全世界约有100万妇女 患有宫颈癌。幸运的是,宫颈癌是目前唯一明 确病因、可预防、早发现可治疗的癌症<sup>[1]</sup>。宫 颈细胞涂片筛查是早期检测宫颈癌的重要细胞 学筛查方法之一,但传统的细胞涂片需要专业 病理学家进行手动筛查,分析过程繁琐、耗时, 长时间工作下容易出错,因此,计算机辅助筛 查在该领域有着重大意义。

细胞学筛查需要观察细胞形态特征,如形状、颜色、尺寸等,分辨是否为异常细胞<sup>[1]</sup>, 计算机辅助筛查系统通过自动分割核质边界、 特征提取和自动分类等方式来自动筛查异常细 胞。而细胞分割作为首要任务,其精确度决定 了后续分类筛查的准确率,但细胞涂片中的细 胞形态极其不规则,染色质不均匀导致细胞边 界不清晰,想要在涂片中准确分割出核质是非 常困难的。目前对细胞进行语义分割的方法主 要有以下两类:

(1)基于形态学的传统细胞语义分割方法。 2011年,Plissiti等<sup>[2]</sup>使用分水岭与基于形态学先 验的方法,先对图像核质区域进行粗分割,再 使用距离相关规则以及支持向量机(support vector machine,SVM)的像素分类方法进行分割的 细化。2015年,Chalfoun等<sup>[3]</sup>通过计算局部对比 度来检测像素强度变化较大的区域,即可能是 细胞主体的区域,再使用局部对比度阈值来分 割出细胞边界,并且使用了一种迭代算法将细 胞边缘的光晕去除。但传统的分割方法只分析 了底层特征,无法提取高级的结构特征,存在 精确率低、泛化能力差、效率低等问题。

(2)基于深度学习的细胞语义分割方法。随着深度学习的快速发展,其在细胞分割领域中的运用越来越广泛,深度学习避免了人工提取

**收稿日期:** 2022-10-06 修稿日期: 2022-10-14

作者简介:王文涛(1967—),男,湖北武汉人,博士,副教授,研究方向为计算机视觉、网络、大数据与数据挖掘; 王嘉鑫(1999—),女,湖北成宁人,硕士,学生,研究方向为计算机视觉;张根(1998—),男,湖北武汉人,硕士,学生,研究方 向为计算机视觉;陈大江(1999—),男,湖北荆州人,硕士,学生,研究方向为数据挖掘
特征的局限,并提供更高的精确度和更快的速 度,其中卷积神经网络(convolutional neural networks, CNN)是最常用的模型之一。 2018年, Liu 等<sup>[4]</sup>使用了一种利用像素先验信息的神经网 络 Mask R-CNN(mask regional convolutional neural network), 先通过基于残差网络(residual network)和特征金字塔(feature pyramid network)的 特征提取网络确定细胞的感兴趣区域,再进行 区域卷积获得粗分割的掩膜图,最后通过一个 局部全连通条件随机场对分割图进行细化,得 到了更高的准确度。2021年, Roy等<sup>[5]</sup>利用一个 基于编码器-解码器的语义分割模型 DeepLabv3 进行细胞分割, 编码器包含了空洞卷积与多尺度 并行的空间金字塔池化模块,可以提取丰富的语 义信息,通过简单有效的双线性上采样解码器模 块进行空间信息的恢复,有效地提高了精度与运 行效率。此外, DenseNet<sup>[6]</sup>、D-MEM<sup>[7]</sup>、ICPN<sup>[8]</sup> 等模型也被用来提高分割性能。虽然这些算法 在一定程度上提高了准确率,但通常具有适应 特定任务或是数据的网络结构特点,且数据不 平衡时模型性能较差。

在众多卷积神经网络模型中,全卷积网络 (fully convolutional networks)<sup>[9]</sup>是医学图像分割 领域中的研究热点之一,在各种全卷积网络体 系中,U-Net<sup>[10]</sup>模型是其中应用最为泛用的模型 之一。U-Net模型是一个像素到像素、端到端的 网络,编码器与解码器之间有跳跃连接层,只 需要较少的训练数据就保留很多特征信息,然 而,向标准的U-Net结构中直接添加更多层,会 使网络太深,导致梯度消失,难以收敛。

本文受 U-Net 模型的启发,结合 Inception-Res<sup>[11]</sup>结构的优点,采用了一个改进的端到端模型 IR U-Net(Inception-Res U-Net),主要贡献如 下:①通过Inception-Res结构增加网络宽度的 同时减少了冗余计算,并能提取多尺度特征, 提高网络精确度;②通过使用Leaky-ReLU减少 "神经元死亡"导致的网络稀疏特征多、难以收 敛的问题;③采用改进的损失函数Focal-Dice Loss来解决细胞图像部分边界模糊、目标区域 大小差异大、学习困难的问题。

# 1 IR U-Net模型

本节将介绍模型的整体结构以及内部模块 的具体结构,整个网络以U-Net作为主干网络模型,内部多尺度特征提取与特征融合模块为Inceptioni-Res结构,模型为不包含全连接层的端 到端模型,输入输出为分辨率相同的图像。

### 1.1 模型整体结构

图1展示了模型的整体结构,模型上半部分为编码器,用于特征提取;下半部分为解码器, 用于像素还原。解码器中上采样的输出将与相 应编码器同层特征图进行拼接,作为解码器下 一层输入,将原始模型中编解码器部分的3×3 卷积模块替换为Inception-Res模块。

模型的输入图像分辨率为224×224,通道 数为3,编码器中最大池化层(MaxPooling)对特 征图进行下采样,使特征图分辨率缩小一半, 经过四个Inception模块和池化层后,特征图缩 小至14×14的大小。解码器部分的上采样层 (UpSampling)将特征图分辨率还原至原来的一 倍,经过四个Inception模块和上采样层后,图 像将恢复到与输入分辨率相同,其中四次跳跃 连接加强了浅层与深层特征的融合,使得分割 结果更为精细。最后经过一个包含1×1卷积层 与 sigmod 激活层的分类器,对像素进行分类, 输出一个三通道的语义分割结果。



图 1 模型结构图

# 1.2 Inceptioni-Res 模块

本文采用了一种改进的Inceptioni-Res结构, 如图2所示,该结构与原始卷积结构不同,使用 了三个不同分支结合的卷积核,主要目的是使 用不同分支不同大小的卷积核输出一个聚合特 征图,多分支的优点主要在于网络能够灵活调 整出对训练有益的卷积核大小,并形成密集的 聚合特征图,配合残差结构,使学习效率增加。 结构中多个1×1卷积核能够改变输入维度以减 少训练参数,使得学习更加容易。



图 2 改进的 Inceptioni-Res 结构

# 2 结构与损失函数优化

### 2.1 批标准化模块

在本文模型中,每个卷积层后加入了批标 准化模块(batch-normalization)<sup>[12]</sup>,当数据通过 卷积层后,其特征分布可能会发生偏移,当卷 积层增加时,偏移会加重,这会产生梯度消 失等现象,批标准化可以在保留卷积层的同 时,使得数据始终保持标准正态分布,加快 训练速度。批标准化的计算首先是对输入数 据 $B = \{x_1, x_2, \dots, x_i\}$ 进行标准化:

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \tag{1}$$

$$y_i = \gamma \hat{x}_i + \beta \tag{2}$$

其中: 对输入数据*B*先计算均值 $\mu_B = \frac{1}{m} \sum_{i=1}^{m} x_i$ 与 方差 $\sigma_B = \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu_B)^2$ , 然后使用式(1)进行 标准化, 使得数据呈标准正态分布, 最后通过 式(2)将标准化后的数据再进行一个线性变换, 虽然这里重新对数据进行偏移,但神经网络可 以学习变换参数γ与β来自动衡量标准化操作是 否对优化产生效果。

## 2.2 激活函数

批标准化后,使用了 ReLU(rectified linear unit)激活函数的变体 Leaky-ReLU<sup>[13]</sup>进行一个输 入端到输出端的非线性映射。原始 ReLU 激活函 数在训练时,值小于0的神经元的梯度会一直为 0,容易产生"神经元死亡"的问题,这样会导 致网络稀疏特征多,难以收敛。而 Leaky-ReLU 函数会给负轴微小的斜率,使得神经元的梯度 不会完全消失,如图 3 所示。



### 图 3 ReLU函数与Leaky-ReLu函数

Leaky-ReLU函数的表达式如下:

$$y = \max(0, x) + k\min(0, x)$$
 (3)

其中:k为一个固定参数且 $k \in (0,1)$ ,一般取  $0.01^{[13]}$ 。

# 2.3 改进损失函数

语义分割常用的损失函数为交叉熵损失函数(cross entropy loss)与Dice系数损失函数<sup>[14]</sup>, 公式如下:

$$L_{\rm CE} = -\frac{1}{N} \sum_{i} \sum_{c=1}^{M} y_{ic} \log(p_{ic})$$
(4)

其中: M为类别数量,当样本i的真实类别为c, 则 $y_{i\epsilon}$ 取1,否则取0, $p_{i\epsilon}$ 为样本i属于类别c的预 测概率。

$$L_{\text{Dice}} = \frac{2\sum_{i=1}^{N} y_{ic} p_{ic} + \varepsilon}{\sum_{i=1}^{N} y_{ic} + \sum_{i=1}^{N} p_{ic} + \varepsilon}$$
(5)

其中: y为真实标签; p为预测概率; c为类别; ε为平滑系数。 何凯明等<sup>[15]</sup>提出了改进的交叉熵损失函数 焦点损失(Focal Loss),公式如下:

$$L_{\text{Focal}}(p_{t}) = -\alpha_{t}(1-p_{t})^{\gamma}\log(p_{t})$$
(6)

其中:

 $\alpha_{\iota} = \begin{cases} \alpha, & \text{if } y = 1 \\ 1 - \alpha, & \text{otherwise} \end{cases}, \quad p_{\iota} = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{otherwise} \end{cases}, \\ \alpha_{\chi} \gamma 均为调节因子, \quad \underline{1} \alpha \in [0, 1], \quad \gamma > 0, \quad \alpha \Pi \\ 来调节正样本损失的重要程度, \quad \gamma \Pi 来调节难样 \\ 本损失的重要程度。$ 

交叉熵损失会对图像中所有的像素点进行 平等地计算,若图像中存在区域非常小的像素 类别,则容易被大范围的背景区域干扰,导致 分割不准确。Dice损失函数是一个区域相关的 损失函数,损失值不受背景区域大小的影响, 所以Dice损失函数善于挖掘前景区域,相比于 交叉熵损失函数,更适用于类别不平衡的情况。 本文通过焦点损失与Dice损失函数结合解决数 据不平衡的问题。改良的损失函数公式如下:

$$L_{\rm FD} = \sum_{c} w_c (1 - L_{\rm Dice}^{\beta}) \tag{7}$$

其中: $w_c$ 为c类别的权重; $\beta$ 为调节因子,且  $\beta \in (0,1)$ ,使得模型更加关注 Dice 损失较小的 样本,本文中 $\beta$ 一般取 0.5。

# 3 实验结果及分析

本节将介绍实验数据、预处理过程、评价 指标、与其他分割算法的性能对比实验分析以 及展示模型各模块影响的消融实验分析。

## 3.1 实验数据与预处理

本文的实验均在公开的 Herlev 宫颈细胞数

据集<sup>[16]</sup>上进行,该数据集是由丹麦赫列夫大学 医院(Herlev university hospital)采集的样本制作 而成,其具体组成如表1所示。

表 1 Herlev 宫颈细胞数据集组成

类别	细胞类型	图像数量/张	细胞种类
1	浅表鳞状上皮细胞	74	
2	中层鳞状上皮细胞	70	正常宫颈细胞
3	柱状上皮细胞	98	
4	轻度鳞状异常细胞	182	
5	中度鳞状异常细胞	146	已造合石伽叻
6	重度鳞状异常细胞	197	开吊吕钡细胞
7	原位癌细胞	150	

该数据集由917张单个宫颈细胞图像组成, 共有七类细胞,每一张图像都对应着一张语义 标注GT(groud truth)图像,如图4所示,所有的 类别以及GT图都是由权威专家人工标注得来, GT图像中浅蓝色区域代表细胞核,深蓝色区域 代表细胞质,灰色区域代表细胞ROI(region of interest),红色区域为背景区域。图像的分辨率 为0.201微米/像素,平均大小为156×140像素。 其中最长边达到768像素,而最短边仅有32像 素,尺寸差异范围较大。

由于数据集每张图像大小形状不一,为了 在统一输入尺寸的同时不改变分割区域的相对 位置,在输入模型前对图像进行零填充,并将 分辨率大小统一调整为224×224像素。同时, 本文采用基于像素的语义分割,GT图中的细胞 主体都在其ROI内,无需先确定细胞的ROI,所



图 4 Herlev 数据集部分图像

· 36 ·

以将ROI并入背景中,最终的语义图像共有三 类像素,分别为细胞核、细胞质与背景区域, 图5为数据处理前后对比。



#### 图 5 图像处理前后对比

为了便于训练,将像素进行编码,表2为本 文训练所使用的像素类别索引。

类别	索引	RGB值
背景	0	(0, 0, 0)
细胞质	1	(0, 0, 255)
细胞核	2	(255, 0, 0)

表 2 像素类别索引

深层网络的训练通常需要大量数据进行学 习,否则容易产生过拟合,但医学图像数据难 以获取,本文使用的公开数据集仅有917张图 像,因此本文对现有数据集进行旋转、水平或 垂直翻转等仿射变换,将数据量扩充至原来的 六倍,使得模型具有更好的泛化能力。

### 3.2 评价指标

为了对语义分割结果进行评估,本文采用 平均像素精度(mean pixel accuracy, MPA)与平均 交并比(mean intersection over union, MIoU)作为 评价指标,公式如下:

$$MPA = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=1}^{k} p_{ij}}$$
(8)

$$MIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=1}^{k} p_{ij} + \sum_{j=1}^{k} p_{ji} - p_{ii}}$$
(9)

其中, 假设共有 *k*+1 个类别(0, 1,…, *k*), *p*<sub>*i*</sub>表示 类别为*i*的像素预测正确的数目, *p*<sub>*i*</sub>表示类别为*i*  的像素被预测为j的数目, p<sub>i</sub>代表类别为j的像素 被预测为i的数目。平均像素精度为每个类别中 分类正确的像素总数与每个类别的像素总数之 比的均值,平均交并比为每个类别真实标签与 预测结果之间交集与并集像素数量比值的均值, 上述指标在用于评价分割模型性能时,值越大 代表性能越好。

# 3.3 实验结果分析

本文仿真实验平台为Windows 10,处理器 为 Intel i3-8100 CPU, 12 GB内存,显卡为 NVIDIA GeForce RTX 2070,8 GB显存,在机器 学习平台 Tensorflow 1.13.1上进行网络训练,优 化方法采用的是 Adam 优化器(adaptive moment estimate optimizer),迭代次数为100次,批量大 小为4,初始学习率为0.001,数据集按照3:1 划分为训练集和测试集,使用扩充后的数据集 进行训练,输入训练图像为4126张。

图 6 对比了 FCN、U-Net 以及本文模型的损失值情况,其中 FCN 模型稳定性差,U-Net 模型 收敛后的 Loss 值比本文模型更大,本文模型收 敛速度更快、更稳定,其精确度有一定的提升。



图 6 三种模型训练过程中的Loss值对比

本文选择了 FCN<sup>[9]</sup>、U-Net<sup>[10]</sup>、Attention U-Net<sup>[17]</sup>、U-Net++<sup>[18]</sup>这几种分割网络与本文的改进模型进行对比,探究改进模型的优越性。表3展示了各个模型使用不同激活函数时,在测试集中的精度指标与完成整个测试集所用的时间。本组实验使用的是改进的 Focal-Dice 损失函数。从表3结果可以看出,本文改进的模型相比于FCN和 U-Net,精确度分别提高了 34.2%和 13.7%,这是因为 FCN 与U-Net 的特征提取与特

٠	37	٠

世刊	油江云粉		PA		MD4	147.77	D 11	n-1 6-1 (
侯型	微伯图叙	细胞核	细胞质	一 mPA 背景	MPA	MIOU	Kecall	内][H]/S
FCN		0.7284	0.8627	0.8539	0.8150	0.6900	0.8139	24.0
U-Net		0.9145	0.9346	0.9244	0.9245	0.8144	0.9417	25.8
Attention U-Net <sup>[17</sup>	<sup>7]</sup> ReLu	0.9655	0.9385	0.9488	0.9509	0.8983	0.9502	30.7
U-Net++ <sup>[18]</sup>		0.9763	0.9475	0.9627	0.9621	0.9198	0.9721	37.6
IR U-Net		0.9801	0.9728	0.9639	0.9727	0.9258	0.9764	29.1
IR U-Net	Leaky-ReLU	0.9801	0.9736	0.9661	0.9732	0.9262	0.9766	29.1

表 3 多种模型分割指标对比

征融合能力不足,无法有效利用编码器提取到的特征。相比于另外两种流行的U-Net改进网络也提升了3.1%和0.6%。本文模型在测试时间上,虽相较FCN与U-Net模型多出了21.2%和12.7%,但其精确度却提高了34.1%和13.6%,具有更好的性能。

另外,细胞核的特征中往往包含着更多可用 信息<sup>[1]</sup>,能将细胞核准确地分割出来具有一定的 意义。本文在细胞核的分割上与近年来一些相关 研究进行横向比较(见表4),发现本文在细胞核 这一类像素的分割上也提升了0.1%~1.5%。

模型	PA	Recall
Attention U-Net <sup>[17]</sup>	0.9655	0.9487
U-Net++ <sup>[18]</sup>	0.9763	0.9694
LFANet <sup>[19]</sup>	0.9787	0.9835
ICPN <sup>[8]</sup>	0.9720	0.9730
IR U-Net	0.9801	0.9837

表 4 多种模型的细胞核分割指标对比

本文改进的损失函数中,权重因子的变化也 会导致分割结果的变化,通过设置不同的权重因 子来探究其对分割结果的影响。首先是类别权重, 图7展示了数据集中各个类别像素个数的均值,细 胞核、细胞质、背景的比值接近于1:2:3,所以 本文将类别权重设置为w<sub>0</sub>:w<sub>1</sub>:w<sub>2</sub>=1:2:3。

表5展示了不同 w 与β值下本文模型的指标。从结果来看,将w比值设置为1:2:3时,相比于均衡的权值,仅在β取0.25时精确度有所下降,β取0.5与0.75时,精确度分别提升了0.1%和0.6%。当β取0.5时模型效果稍好,相比于另外两个取值所对应的精确度提升了0.1%

# 和0.4%,所以本文将默认设置β为0.5。



图 7 数据集每个类别像素的平均数量

#### 表 5 不同 w 和β值对模型的影响

$w_0 : w_1 : w_2$	β	MPA	MIoU
	0.25	0.9709	0.9236
1 : 1 : 1	0.50	0.9715	0.9254
	0.75	0.9653	0.9097
	0.25	0.9686	0.9190
1 : 2 : 3	0.50	0.9732	0.9262
	0.75	0.9713	0.9292

为探究不同损失函数对模型性能的影响, 本文在不同模型分别使用交叉熵损失、Dice损 失、Focal损失以及本文改进的损失函数之间 进行对比试验。表6展示了各个模型在不同损 失函数下的分割效果,结果表明改进的Focal-Dice损失函数相比交叉熵损失精确率提升了 0.5%~3.5%,相比Dice损失精确率提升了 0.2%~5.6%,相比Focal损失精确率提升了 0.4%~1.4%。由于本文数据集的类别不平衡, 单一损失函数很难衡量训练时类别的重要性,

4# mil	交	又熵	]	Dice	F	ocal	Foc	al-Dice
侠型	MPA	MIoU	MPA	MIoU	MPA	MIoU	MPA	MIoU
FCN	0.8396	0.6562	0.8227	0.6328	0.8570	0.6933	0.8694	0.7100
U-Net	0.9146	0.8030	0.9173	0.8082	0.9397	0.8560	0.9471	0.8954
Attention U-Net <sup>[17]</sup>	0.9456	0.8611	0.9485	0.8680	0.9409	0.8742	0.9509	0.8983
U-Net++ <sup>[18]</sup>	0.9420	0.8525	0.9440	0.8573	0.9503	0.8725	0.9621	0.9198
IR U-Net	0.9629	0.9035	0.9650	0.9087	0.9691	0.9239	0.9732	0.9262

表 6 不同损失函数对多种模型的影响

当训练细胞核这类数量较少的类别时,很容易 被其他类别所影响,改进的损失函数针对这一 点进行优化,使模型能够更好地关注于类别少 且难分类的像素点。

### 3.4 消融实验分析

为了验证本文改进方法的可行性,将三个 改进模块进行单独实验,探究每个模块各自对 模型性能的影响,在相同数据集上进行相应的 消融实验。

表7中第一行实验数据为原U-Net模型下的分割精度,其效果达不到期望。第二行实验数据表明,在加入Inception-Res模块后,各项指标分别提升了3.8%、11.6%和1.0%,结果优于原模型,这表明该模块对模型性能提升有所帮助。第三、四行实验数据中,分别再加上Leaky-ReLU与Focal-Dice模块,两者使得模型的

MPA分别提升了1.8%和2.3%, MIoU分别提升 了1.3%和3.2%, 由此证明了这两个模块对模型 性能提升的有效性。

表 7 不同模块对网络性能的影响

Inception-	Leaky-	Focal-	MDA	MIaU	Pagall
Res	ReLU	Dice	MPA	MIOU	кесан
			0.9146	0.8030	0.9394
$\checkmark$			0.9502	0.8969	0.9497
$\checkmark$	$\checkmark$		0.9678	0.9091	0.9621
$\checkmark$		$\checkmark$	0.9727	0.9258	0.9764
$\checkmark$	$\checkmark$	$\checkmark$	0.9732	0.9262	0.9766

### 3.5 结果可视化

本文使用的所有网络分割结果的可视化如 图8所示,结果表明,FCN模型由于特征提取与 像素还原能力低,细胞边缘细节刻画效果不佳,



图 8 各模型分割可视化结果

细胞核尚未完全分割出来,整体效果非常粗糙, U-Net模型存在同样问题,虽然在细胞核分割上 优于FCN,但其边缘细节仍然未划分出来,受 背景影响严重。Attention U-Net与U-Net++模型 在细胞核刻画上效果与本文模型相差不大,但 细胞质边缘区域分割效果仍然不佳。相比之下, 本文的改进模型能够有效地分割出细胞核质区 域,对细胞主体的刻画优于其他几种模型,具 有较好的效果,分割能力明显提升,且在样本 不均衡,即细胞区域较小时也能达到较好的分 割效果。

### 4 结语

目前宫颈细胞分割领域存在一定的问题, 本文在U-Net结构的基础上,结合了多尺度卷积 与残差连接的Inception结构,加宽网络的同时 避免了梯度消失,并使用改进的损失函数获得 了较好的分割效果。结果表明改进的模型相比 于目前常用模型均有精度提升,改进的损失函 数相比原始损失函数在对模型精度提升方面更 为优越,模型整体的分割结果与专业人员标记 接近,具有一定的现实意义。

### 参考文献:

- ZHANG S, XU H, ZHANG L, et al. Cervical cancer: epidemiology, risk factors and screening [J]. Chinese Journal of Cancer Research, 2020, 32(6):720.
- PLISSITI M E, NIKOU C, CHARCHANTI A. Automated detection of cell nuclei in pap smear images using morphological reconstruction and clustering [J].
   IEEE Transactions on Information Technology in Biomedicine, 2010, 15(2):233-241.
- [3] CHALFOUN J, MAJURSKI M, PESKIN A, et al. Empirical gradient threshold technique for automated segmentation across image modalities and cell lines [J]. Journal of Microscopy, 2015, 260(1):86-99.
- [4] LIU Y, ZHANG P, SONG Q, et al. Automatic segmentation of cervical nuclei based on deep learning and a conditional random field [J]. IEEE Access, 2018,6:53709-53721.
- [5]ROY R M, AMEER P M. Segmentation of leukocyte by semantic segmentation model: a deep learning ap-

proach[J]. Biomedical Signal Processing and Control, 2021,65:102385.

- [6] SABEENA K, GOPAKUMAR C, THAMPI R. An improved deep convolutional model for segmentation of nucleus and cytoplasm from pap stained cell images [C]//Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems(ICACCS). IEEE, 2020; 1274-1278.
- [7] ZHAO J, LI Q, LI X, et al. Automated segmentation of cervical nuclei in pap smear images using deformable multi-path ensemble model [C] // Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging(ISBI 2019). IEEE, 2019:1514-1518.
- [8] YANG X, WU J, YIN Y. Interacting convolution with pyramid structure network for automated segmentation of cervical nuclei in pap smear images [C] //Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 2020: 499-502.
- [9] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.
- [10] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C] //Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015:234-241.
- [11] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2818-2826.
- [12] SANTURKAR S, TSIPRAS D, ILYAS A, et al. How does batch normalization help optimization?
   [C] //NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2018: 2488-2498.
- [13] DUBEY A K, JAIN V. Comparative study of convolution neural network's relu and leaky-relu activation functions [M] // Applications of Computing, Automation and Wireless Systems in Electrical Engineering.

Springer, Singapore, 2019:873-880.

- [14] JADON S. A survey of loss functions for semantic segmentation [C] //Proceedings of the 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). IEEE,2020:1-7.
- [15] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C] //Proceedings of the IEEE International Conference on Computer Vision, 2017:2980-2988.
- [16] MARTIN E. Pap-smear classification [D]. Technical University of Denmark, 2003.
- [17] OKTAY O, SCHLEMPER J, FOLGOC L L, et al.

Attention U-Net: learning where to look for the pancreas[C]//MIDL, In Proc. MIDL, 2018, 1:1-10.

- [18] ZHOU Z, SIDDIQUEE M M R, TAJBAKHSH N, et al. Unet++: a nested U-Net architecture for medical image segmentation [M] //Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, Cham, 2018: 3-11.
- [19] ZHAO Y, FU C, XU S, et al. LFANet: lightweight feature attention network for abnormal cell segmentation in cervical cytology images[J]. Computers in Biology and Medicine, 2022, 145:105500.

# End-to-end cervical cell segmentation with multi-scale convolution

Wang Wentao<sup>1,2</sup>, Wang Jiaxing<sup>1,2</sup>, Zhang Gen<sup>1,2</sup>, Chen Dajiang<sup>1,2</sup>

(1. College of Computer Science, South-Central Minzu University, Wuhan 430074;

2. Hubei Provincical Engineering Research Center for Intelligent Management of Manufacturing Enterprise, Wuhan 430074)

Abstract: Cervical cancer is the only gynecological malignant tumor with clear etiology. Cervical cell screening is usually used for early detection and treatment. In the process of cell screening, it is necessary to segment the nucleus and cytoplasm accurately from the microscopic cell images in order to find early abnormal cervical cells. The existing methods of cervical cell segmentation have some problems, such as large amount of calculation, low accuracy and learning difficulties caused by data imbalance. In this paper, the idea of multi-scale convolution is introduced, and an end-to-end cell segmentation algorithm IR U-Net is designed by using a U-shaped encoder-decoder model to improve the performance of cervical cell segmentation algorithm. It mainly includes: ①Using multi-scale convolution structure to broaden the network, avoid manual selection of convolution kernel, and can extract multi-scale features. At the same time, residual connection is added to the convolution structure to reduce the phenomenon of vanishing gradient; ②Leaky-RelU can be used to solve the problem of network has many sparse features and is difficult to converge due to "neuron death"; ③Focal Dice Loss, as an improved Loss function was used to alleviate the problem of data imbalance. The simulation results show that the improved model has better accuracy and better segmentation performance than the control algorithm.

Keywords: cervical cells; image segmentation; U-Net; multiscale convolution; loss function

文章编号:1007-1423(2023)02-0041-09

DOI: 10.3969/j.issn.1007-1423.2023.02.005

# 一种基于YOLOv5算法的布匹瑕疵检测系统

邓 景,李成海,丁兆栋,杜光辉,陆 可\*

(安徽工业大学管理科学与工程学院,马鞍山 243000)

摘要:在纺织行业中,布匹瑕疵的出现将严重影响其经济价值,而目前纺织工厂的质检环节严重依赖 人工质检。为了解决人工质检过程中出现的错检、漏检以及检测效率随时间下降的问题,设计了一种基于 YOLOv5深度学习算法的布匹瑕疵检测方法,通过在真实产线上收集布匹样本并自建相关瑕疵的数据集, 使用YOLOv5训练出一组双面的布匹瑕疵模型。实验结果表明,该模型在检测速度与准确度方面与传统质 检方式相比表现优异,可以满足工业生产的实际需要。

关键词:布匹瑕疵检测;质量检测;深度学习;YOLOv5

### 0 引言

中国作为全球最大的纺织品生产、出口国, 纺织业已经成为我国一项重要的经济型支柱产 业。受限于技术、成本等多方因素,纺织品生产 过程中产生瑕疵是不可避免的。生产过程中的布 匹瑕疵会使布匹的价格降低45%~65%<sup>[1]</sup>。因此, 现代化的纺织工厂通常会安排一定数量的员工对 纺织设备与纺织产品进行定期、定点的监测。然 而,人工检验存在错检、漏检、检测准确率随时 间下降和检测标准不统一等问题,同时人工检验 的成本较高,不利于工厂的精益生产。

布匹瑕疵检测作为目标检测问题的一种, 许多研究人员提出了自动化的布匹瑕疵检测方 法,来代替人工检验<sup>[2]</sup>。传统的布匹瑕疵检测 算法根据类型分为特征统计法、滤波法、图像 切割法和数据结构法等。但这几类方法各自存 在不抗噪点、计算量较大、泛化、迁移能力较 差等问题。随着硬件水平的提高,深度学习逐 渐进入研究人员的视野,基于深度学习的目标 检测技术,也逐渐应用于各个行业的检测领域。

深度学习相较于传统的目标检测算法,其 优势在于可以自动提取目标图像中的特征信息。 从最初的R-CNN<sup>[3]</sup>、OverFeat<sup>[4]</sup>一出现便占据了 目标检测的主流地位,到后来的Fast RCNN<sup>[5]</sup>、 Faster RCNN<sup>[6]</sup>进一步提升了检测速度与精度; 之后, SSD<sup>[7]</sup>以及现在的 YOLO 系列实现了两阶 段到单阶段的革新。与两阶段算法相比,单阶 段算法不需要区域建议阶段,可以生成目标的 真实坐标和类别概率,以获得最终检测结果。 因此,单阶段算法通常具有更快的速度。YOLO 系列是2015年提出的算法,此算法的最大优势 在于不需要提前筛选候选区域,而是直接检测 整个图片,获取某个类别的置信度与位置信息, 所以检测速度更快。在过去几年中,研究人员 提出了一些 YOLO 的后续版本,如 YOLOv2、 YOLOv3、YOLOv4。而在2020年被提出的YO-LOv5, 其在模型大小、检测速度与精度上都有 了较大的优化。由此,本文提出了基于 YOLOv5 算法的布匹瑕疵检测系统。

本文的贡献如下:

(1)将YOLOv5算法应用于布匹瑕疵检测领

收稿日期: 2022-09-22 修稿日期: 2022-10-20

作者简介: 邓景(1997—),男,安徽芜湖人,硕士,研究方向为机器视觉、深度学习;李成海(1998—),男,安徽六安 人,硕士,研究方向为机器视觉、深度学习;丁兆栋(2000—),男,安徽亳州人,本科,研究方向为计算机视觉、深度学习; 杜光辉(1998—),男,安徽池州人,专科,研究方向为机器视觉、连锁经营管理;\*通信作者:陆可(1983—),男,广西南宁人, 博士,硕士生导师,从事领域为人工智能、机器视觉,E-mail:looker@ahut.edu.cn

域,其检测速度与准确度均满足实际工业生产 需要。

(2)提供了一套来自实际工业生产流水线的布匹瑕疵数据集,为布匹瑕疵研究领域提供了一些数据支持。

### 1 相关工作研究

为了解决自动化布匹瑕疵检测的问题,研 究人员提出了许多系统和算法用于解决自动化 布匹瑕疵检测问题。Tong等<sup>[8]</sup>利用复合差分进 化优化了几个Gabor滤波器,成功地从布匹图像 中分割出瑕疵。Mak 等<sup>[9]</sup>还通过遗传算法优化了 Gabor 滤波器。然而,优化的过滤器仅适用于某 些类型的布匹。Kumar 等<sup>[10]</sup>提出了具有不同尺 度和方向的 Gabor 滤波器组, Gabor 滤波器组覆 盖多分辨率,可以通过融合Gabor滤波的输出来 检测缺陷的特征。尽管Gabor滤波器组可以处理 不同类型的布匹,但简单地融合滤波器输出可 能会加剧正常布匹图像中的噪声。传统目标检 测算法尽管取得了一定的成果,但仍有相当的 优化空间。随着深度学习的广泛应用,研究人 员开始尝试利用深度学习算法解决布匹瑕疵的 检测问题。

目前主流的布匹瑕疵检测所采用的深度学 习算法均为使用卷积神经网络提取特征并分类, 例如,文献[11]将传统目标检测算法中的图像 增强技术与卷积神经网络算法相结合,对图像 样本进行图像增强后进行模型训练,实现了对 部分类型布匹瑕疵的有效检测,但输出的样本 类型较为有限;文献[12]提出了基于改进的 CNN的素色布匹瑕疵检测算法,通过设计浅层 的卷积神经网络结构,降低了网络的运算规模, 同时采用了双模型的训练方式,提高了模型的 训练效果,在单色布匹瑕疵分类方面具有良好 的性能。但仍然存在对复杂瑕疵样本不敏感的 问题。因此,本文将重点关注深度学习中的 YOLOv5算法。

文献[13]介绍了YOLO在目标检测场景中的应用。YOLO定义为单个神经网络,可以预测 多个锚框和锚框的类别概率。与其他传统深度 学习算法,如Faster R-CNN(regions with CNN)<sup>[14]</sup> 相比,其背景错误不足一半。当然,也有一些缺 点:第一,它虽然可以更快地检测图像,但缺乏 精度;第二,模型从给定的数据中学习预测边界 框,因此,当一般图像或具有不同纵横比的图像 或图像的配置改变时,预测将变得困难。

为了克服如本地化中的错误分析、相对较 低的召回率等问题, 文献[15]提出了另一个模 型一YOLO9000(YOLOv2),引入了批量归一化、 带锚框的高分辨率卷积、维度聚类、直接位置 预测、多尺度训练和细粒度特征,这使YOLOv2 相比于 YOLO 性能更加优异。文献 [16] 提出了 另一个优于 YOLOv2的 YOLO 模型。与传统方法 相比, YOLOv3的精度是传统网络的三倍。文献 [17]提出了其后的版本,这个新版本提高了目标 检测的速度,还将YOLOv3平均精度和每秒帧数 分别提高10%和12%。作者表示, YOLOv4适用 于单 GPU 训练,这是通过引入新的数据增强方 法、优化超参数和遗传算法实现的。文献[18]采 用 Darknet53 + CSP 作为主干网络的 YOLOv4 模 型,提升了模型对瑕疵图像的学习能力,在 Neck 层采用 SPP 模块和 FPN+PAN 结构, 进一步 提取深层网络信息。该方法对布匹瑕疵的检测 准确率和检测速度均较好,与其他同类深度学 习算法相比体现出了一定的优势,但仍然存在 网络模型较大等问题。在YOLOv4出版后的几个 月内, Ultralytics LLC团队推出了YOLOv5<sup>[19]</sup>。

在将传统的目标检测算法应用到工业场景的过程中,由于其参数密度大、网络模型复杂、 难以嵌入低配设备等问题,导致在应用过程中 因对硬件性能要求过高,成本过大而无法投入 实际生产。为了解决以上问题,首先在作业现 场采集了多种复杂场景的实际应用数据,将这 些数据输入模型进行学习以适应作业现场环境 的复杂多样性。其次,为了解决现场硬件配置 的限制导致的大规模目标检测算法难以嵌入的 问题,本文采用了YOLOv5算法。由于其轻量级 的网络结构可以部署在配置相对较低的设备上, 同时YOLOv5可以保持相对较高的准确度。实验 结果表明,YOLOv5算法在适应不同复杂度场景 的情况下,鲁棒性和准确度均显著优异。

### 2 模型训练及实验

### 2.1 算法原理

YOLO 目标检测算法的核心在于小尺寸模型 与快速计算。YOLO 的结构原理是将全局图像作 为神经网络的输入,最终在输出层直接输出回 归目标框的类别和位置信息<sup>[20]</sup>。值得指出的是, YOLO具有很强的泛化能力,因为YOLO可以研 究相当广义的特征,以转移到其他文件<sup>[21]</sup>。此 外,YOLO在训练和测试过程中使用全局图像, 可以对全局信息进行编码,并减少检查背景作为 目标的错误。本文使用YOLOv5的深度学习方法 检测毛巾中的布匹瑕疵,并开发了基于YOLOv5 的自动布匹瑕疵检测系统来检测毛巾瑕疵。

YOLOv5是在之前版本的 YOLO 基础上形成 的。在保持运行速度的前提下, YOLOv5不仅获 得了更好的预测精度,同时对小目标的检测能 力更强<sup>[22]</sup>。因此,它可以用于工业缺陷的实时 检测。而且YOLOv5的权重文件,其大小比上一 代的 YOLOv4小90%<sup>[23]</sup>。YOLOv5结合 CNN,利 用特征金字塔和非最大抑制算法预测多标签的 边界和位置。边界盒由独立的逻辑回归分类器 进行分类,而不是软最大值分类<sup>[24]</sup>。在检测 布匹瑕疵的情况下,准确性和效率非常重要, YOLOv5架构有助于提高布匹瑕疵的效果。该 架构由四种不同的模型组成,如YOLOv5s、 YOLOv5m、YOLOv5l和YOLOv5。四种架构的主 要区别在于特征提取模块和网络的卷积核;另 一个差异是架构,模型的大小和模型参数的数 量是不同的。例如,本文使用 YOLOv5s,则其 模型文件的大小仅为13.7 MB。

### 2.2 模型训练

在本节中,我们将介绍模型的训练和测试 过程。模型的训练和测试流程如图1所示。



#### 图 1 模型的训练和测试流程图

### 2.2.1 自建数据集与数据标注

本项目落地于江苏省某纺织企业,数据的 采集全部取自真实工厂的流水线上,图2是采集 的数据集中的常见瑕疵。首先,选用2D单目彩 色相机以及适合现场条件的光源组成CCD组, 对生产线上的布匹样本进行大量采集。然而, 当前的数据集数量过于庞大且种类繁多,并不 适合直接进行人工数据标注,需要对数据集中 的数据进行筛选和聚类预处理。预处理完成后, 使用LabelImg的数据标注工具进行人工数据标 注<sup>[25-26]</sup>。对处理完毕的数据进行成功标注后,输 出文本文件和类文件。为了提高速度和准确性, 我们对毛巾的正反面分别建立了模型,只将其 分为两类: OK和NG。文本文件包含5个十进制 数字,第一个数字表示边界框的类别,第二个 是中心x,然后是中心y,最后是宽度和高度。 中心 x 和中心 y 被描述为边界框的中心点[27]。通 过将数字除以图像的宽度和高度,来对数字进 行归一化, 使其值介于0和1之间。文本文件中 的行数取决于检测到的对象数。标注的标签用 于训练和测试数据集。表1显示了每个类别的训 练和测试拍摄的图像数量。表2显示本文项目的 类,即类文件。



图 2 数据集中的瑕疵

表 1 训练与测试样本数

	类别	训练样本	测试样本	总计样本
	0K	598	93	691
毛巾正面	NC	135	82	517
	NG	433	02	517
毛巾反面	ОК	535	87	622
	NG	613	97	710

类编号	类
0	OK
1	NG

表 2 类文件名

### 2.2.2 使用 YOLOv5 训练模型

在使用YOLOv5训练模型之前,首先要满足 训练YOLOv5模型的条件,表3为本项目所使用 的硬件配置条件。

表 3	硬件配置表

设备	参数
操作系统	Windows 11.7
GPU	GeForce RTX 3070
CPU	Intel Core i9-10980XE @ 3.00 GHz
硬件加速	CUDA10.1

然后配置YOLOv5环境、数据和目录结构, 重点是设置YAML配置文件。训练YOLOv5模型 需要两个YAML文件。第一个YAML指出测试 数据的位置、训练数据的位置和类的数量,即 检测对象的类型与该类型所对应的名称;第二 个YAML文件包括锚框、参数、YOLOv5主干和 YOLOv5头部。为了训练YOLOv5模型,需要执 行train.py命令。此时,可以指定超参数,如输 出图像大小、epoch数和batch大小。完成训练 过程后,YOLOv5的权重文件将存储在子文件夹 中。可以通过使用测试的 detect.py 命令来对训 练完毕模型进行测试。图 3 所示为自测试数据集 的测试结果。



图 3 测试集结果

### 2.3 实验结果

在目标检测领域,对训练模型的精确度 (P)、召回率(R)和平均精度(mAP)进行了评估<sup>[28]</sup>。作为性能和可靠性的通用规范,本文也 使用上述评估指标来评估缺陷检测模型的性能。 图4所示为mAP、精确度和召回率。

TP=正确识别的毛巾OK和NG部分的实例数。

FP=毛巾OK和NG部分的错误识别实例数。 FN=未识别的OK和NG部件的数量。



精确度(P)是检测到的正确对象数量与检测 到的对象数量之和之间的比率,是检测模型精 确度的度量。其计算公式如下:

$$P = \frac{TP}{TP + FP} \tag{1}$$

召回率是检测到的正确对象的数量与手动标记的真实对象的数量之和之间的比率,并且 是检测模型的整个错误率的度量。其计算公式如下:

$$R = \frac{FP}{TP + FN} \tag{2}$$

图 5 中的准确率-召回率(PR)曲线是根据相 应的准确率和召回率绘制而成。平均准确度 (AP)是精确召回率(PR)曲线下的面积值,通过 对精确召回函数积分来获得。

$$AP = \int_{0}^{1} P(R) dR$$
 (3)

*mAP*是多个测试集上测试的正确目标的平均值,通过平均精度(*AP*)获得。*mAP*的计算公式如下:

$$mAP = \frac{\sum AP}{N} \tag{4}$$

探索性数据分析最好借助相关图进行处理。 整个数据集的关系可以一目了然地可视化。图6 显示了置信度与准确度的P曲线,图7显示了置 信度与召回率的R曲线,图8显示了置信度与 F1曲线。

$$F1 = \frac{2}{\frac{1}{R} + \frac{1}{P}}$$
 (5)





图 6 P曲线







图 8 F1曲线

对于检测算法模型来说,检测速度是衡量 算法效果的一个重要指标。本文建立的布匹瑕 疵检测系统是实现对目标的实时检测,在应用 于实际工业场景的情况下,其检测速度为每秒 10帧,已充分满足实际需求。

# 3 布匹瑕疵检测系统

### 3.1 系统框架

建立一套自动化布匹瑕疵检测系统的难点 在于,除了实时检测对深度学习算法的性能有 一定的要求以外,对于本系统内各部分的耦合 度、鲁棒性与上下游工位之间的通讯稳定性均 有一定的要求。本文设计了基于深度学习的布 匹瑕疵检测系统,其物理图如图9—图11所示。 该系统由三部分组成:控制子系统、机器视觉 子系统与PLC电气控制系统,实现了YOLOv5算 法对布匹瑕疵的实时检测。

(1) 控制系统由控制计算机来担任。负责传 递来自PLC系统的工作电信号,确定物料的各 个工作位点;收集的数据传输到控制计算机, 由YOLOv5处理,以快速检测和定位布匹瑕疵。

(2) PLC 电气控制系统由 PLC、机械吸盘和 红外激光传感器组成。红外激光传感器在物料 到位后受遮挡,发送电信号至 PLC 传达到位信 号;机械吸盘由 PLC 控制拖拽物料至各个工作 位点; PLC 负责接收来自传感器与控制计算机 的电信号,控制机械吸盘的升降与移动。

(3) 机器视觉子系统由三部分组成:图像采

集、标签检测和数据处理。图像采集由双CCD 摄像机完成。当机械吸盘拖拽指定编号物料至 指定工作位时,上CCD摄像机和下CCD摄像机 以不同角度同时采集物料的正反面图像。标签 检测由计算机完成,将采集的多标签图像通过 相机数据传输接口传输到计算机以处理和分析 图像。数据处理则是通过在YOLOv5模型参数中 预设的置信度阈值,结合物料的正反面处理结 果,得出OK或NG并传回相应的电信号至PLC 程序,同时在上位机中显示。



图 9 物理左视图 图 10

图 10 物理右视图



图11 物理正视图

理结构示意图如图12所示。

基于YOLOv5的自动布匹瑕疵检测系统的物

RC 光源 机械吸盘 部游方向 毛巾 载物平台 光源 相机 光源 北源 北海和 计算机

#### 图 12 物理结构示意图

### 3.2 系统优化

系统在项目实施过程中,受限于诸多客观 因素,导致布匹瑕疵检测效果达不到预期的目 标。因此,本文针对项目过程中出现的以下问 题,对系统的若干部分进行了优化。

(1)在模型训练的数据收集阶段,现场采集的毛巾样本图像种类、数量庞大,人工分类效率过低,不便于进行数据标注的工作。由此,在数据的预处理阶段,使用了聚类的算法,自动对数据进行筛选、分类和去重。大大减少了准备工作阶段的时间消耗。

(2)由于工作现场环境的客观因素影响,在 图像采集过程中,CCD摄相机时常拍入毛巾样 本以外的干扰因素,例如:图像内的机械结构 和光源的倒影。由此,本文对YOLOv5输出端的 图像进行了预处理,根据待检测毛巾的最大尺 寸设计了最大预处理尺寸的锚框。加入该限值 后,消除了上述干扰因素的影响,检测效果明 显提高。图13表示最大预处理大小的锚框。



图 13 筛选锚框

#### 4 结语

本文提出了一种基于YOLOv5算法的自动化 布匹瑕疵检测系统,并将其成功应用到实际的 工业生产当中。相对于传统的人工检测,通过 应用YOLOv5算法,取得了良好的检测效果,有 效地降低了生产成本,显著提高了纺织行业的 经济效益。由于YOLOv5的权重文件小且运行速 度快,使得系统在硬件配置条件有限的前提下, 仍然可以成功检测出流水线上的布匹瑕疵,已 充分满足实际生产中的检测需求。

此后的工作将进一步关注更多复杂场景的 布匹瑕疵,为自建数据集加入更多的复杂布匹 瑕疵类型,进一步扩展模型可以检测的瑕疵类 别。同时继续对现有的YOLOv5算法进行改进, 提升算法在实际应用中的效果,同时对系统可 以输出的内容进行进一步的扩展,不仅仅是输 出布匹的良品与瑕疵,更进一步输出瑕疵的类 别与位置。

### 参考文献:

- [1] OUYANG W, XU B, HOU J, et al. Fabric defect detection using activation layer embedded convolutional neural network [J]. IEEE Access, 2019, 7: 70130 – 70140.
- [2] HUANGPENG Q, ZHANG H, ZENG X, et al. Automatic visual defect detection using texture prior and low - rank representation [J]. IEEE Access, 2018, 6: 37965-37976.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] //Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2014:580-587.
- [4] SERMANET P, EIGEN D, ZHANG X, et al. Overfeat: integrated recognition, localization and detection using convolutional networks [EB/OL]. arXiv Preprint arXiv:1312.6229,2013.
- [5] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6):1137 -1149.
- [6] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [C] //Proceedings of the European Conference on Computer Vision. Springer, Cham, 2016.
- [8] TONG L, WONG W K, KWONG C K. Differential evolution - based optimal Gabor filter model for fabric inspection [J]. Neurocomputing, 2016, 173: 1386-1401.

- [9] MAK K L, PENG P, YIU K F C. Fabric defect detection using multi-level tuned matched Gabor filters
   [J]. Journal of Industrial and Management Optimization, 2012, 8(2): 325-341.
- [10] KUMAR A, PANG G K H. Defect detection in textured materials using Gabor filters [J]. IEEE Transactions on Industry Applications, 2002, 38(2):425-440.
- [11] 安静,唐英杰,马鑫然.基于深度神经网络的素色布 匹瑕疵检测算法研究[J]. 包装工程,2021,42(3): 246-251.
- [12] 吴志洋,卓勇,李军,等.基于卷积神经网络的单色 布匹瑕疵快速检测算法[J]. 计算机辅助设计与图 形学学报,2018,30(12):2262-2270.
- [13] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C] // Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016:779-788.
- [14] GIRSHICK R. Fast R CNN [C] //Proceedings of the IEEE International Conference on Computer Vision, 2015:1440-1448.
- [15] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:7263-7271.
- [16] REDMON J, FARHADI A. YOLOv3: an incremental improvement [EB/OL]. arXiv Preprint arXiv: 1804.02767, 2018.
- [17] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. arXiv Preprint arXiv: 2004.10934, 2020.
- [18] SARANYA K C, THANGAVELU A, CHIDAM -BARAM A, et al. Cyclist detection using tiny YOLO v2 [M] //Soft Computing for Problem Solving. Springer, Singapore, 2020:969-979.
- [19] YANG G, FENG W, JIN J, et al. Face mask recognition system with YOLOv5 based on image recognition [C]//Proceedings of the 2020 IEEE 6th International Conference on Computer and Communications (ICCC). IEEE, 2020:1398-1404.
- [20] WAGEEH Y, MOHAMED H E D, FADL A, et al. YOLO fish detection with Euclidean tracking in fish farms [J]. Journal of Ambient Intelligence and Humanized Computing, 2021, 12(1):5-12.
- [21] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detec-

tion [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016:779 -788.

- [22] JIANG P, ERGU D, LIU F, et al. A review of YOLO algorithm developments [J]. Procedia Computer Science, 2022, 199:1066-1073.
- [23] MATHEW M P, MAHESH T Y. Leaf-based disease detection in bell pepper plant using YOLO v5 [J].
   Signal, Image and Video Processing, 2022, 16(3):841 -847.
- [24] LIN G, LIU K, XIA X, et al. An efficient and intelligent detection method for fabric defects based on improved YOLOv5[J]. Sensors, 2022, 23(1):97.

- [25] YAN B, FAN P, LEI X, et al. A real-time apple targets detection method for picking robot based on improved YOLOV5 [J]. Remote Sensing, 2021, 13, (9):1619.
- [26] 钟铭.基于卷积神经网络的纹理布匹瑕疵检测研究 [D].广州:华南理工大学,2021.
- [27] CHETVERIKOV D, HANBURY A. Finding defects in texture using regularity and local orientation [J]. Pattern Recognition, 2002, 35(10):2165-2180.
- [28] 许玉格,钟铭,吴宗泽,等.基于深度学习的纹理布 匹瑕疵检测方法[J/OL].自动化学报:1-16[2022-09-22].DOI:10.16383/j.aas.c200148.

# A fabric defect detection system based on YOLOv5 algorithm

Deng Jing, Li Chenghai, Ding Zhaodong, Du Guanghui, Lu Ke\*

(School of Management Science and Engineering, Anhui University of Technology, Ma'anshan 243000)

Abstract: In the fabric manufactory plant industry, the defects of fabric emerging randomly, which leading to the incalculable economic loss. However, the quality inspection of textile factories relies heavily on manual quality inspection. In order to solve the problems of false detection, missing detection and the decline of detection efficiency over time in the process of manual quality inspection, a cloth defect detection method based on YOLOv5 deep learning algorithm is designed. By collecting cloth samples on the real production line and building a data set of related defects, a group of double-sided cloth defect models are trained using YOLOv5. The experimental results show that the model performs better than traditional quality inspection methods in terms of detection speed and accuracy, and can meet the actual needs of industrial production.

Keywords: fabric defect detection; quality detection; deep learning; YOLOv5

文章编号:1007-1423(2023)02-0050-06

DOI: 10.3969/j.issn.1007-1423.2023.02.006

# 基于混合遗传算法的成品油二次配送优化

孙厚举

(江苏建筑职业技术学院信电工程学院,徐州 221000)

摘要:成品油公路运输多周期规划问题是成品油供应链的核心研究课题,属于多周期库存路径问题。 该研究使用松弛下届模型加快求解速度,形成合法解的集合,将数学规划模型与改进的启发式算法相结合,在 集合中快速迭代,在求解配送量时使用贪心机制得到最大收益方案。最终,在多周期循环配送算例集上测试, 结果表明,可以在多数小、中规模算例中获得理论最优解,在大规模算例中平均运输成本降低12.3%。

关键词:成品油公路运输;多周期库存路径问题;混合遗传算法;松弛模型

### 0 引言

多周期库存路径问题<sup>[1]</sup> (inventory routing problem with multi period, IRP-MP)属于非确定 性多项式(non-deterministic polynomial, NP)困难 问题<sup>[2]</sup>的一种,是通过供应商管理库存策略 (vendor managed inventory, VMI)<sup>[3]</sup>,在无限长的 时间内周期性地进行一对多补货配送方案的规 划和执行。成品油公路运输问题是典型的多周 期库存路径问题,石油公司负责对各加油站油 品库存的监控,统筹制定油品配送方案,确保 配送高效且可靠。通常情况下,石油公司将制 定总运距尽可能短的运送计划,以实现高质量 的物流运输目标。

学术界对成品油公路运输的研究经历了从 精确算法到智能化算法的过程。田立新等<sup>[4]</sup>采 用0-1规划模型解决单周期成品油运输问题。 宋杰鲲等<sup>[5]</sup>采用最短路径和最小费用模型制定 成品油运输模型。张立峰等<sup>[6]</sup>提出了两阶段路 径规划模型,首先使用聚类算法将配送需求合 并分载,然后改进遗传算法计算最终配送路线。 李珍萍等<sup>[7]</sup>考虑了多种约束条件,提出了一种 混合整数规划模型,继而使用遗传算法模型规 划运送路线。张雄等<sup>[8]</sup>使用改进蚁群算法求解 了带时间窗口的车辆路径问题。Xu等<sup>[9]</sup>提出了 多目标粒子群优化算法制定配送方案。文献 [10-11]研究了启发式算法求解多周期库存路径 问题,文献[12]介绍了运输成本问题,本文在 设计混合遗传算法的评估函数时参考了其中的 部分因素。

## 1 成品油公路运输问题描述

成品油公路运输研究的是成品油运输二次 配送阶段问题<sup>[13]</sup>。一次配送阶段主要是从炼油 厂将成品油运送至各地级市的油库,一个省往 往只有一个炼油厂,一个地级市只有一个油库 且运送方式多为管道运输,几乎不需要复杂的 规划模型。二次配送<sup>[14]</sup>是从油库将成品油运送 至加油站,目前石油公司统一采用主动配送方 式,即石油公司根据液位仪系统时时监控所有 加油站多种油品的存量,根据各加油站油品的 消耗情况生成配送需求,然后使用算法规划配 送时间和配送量。成品油公路运输多周期规划 问题中,通常每12小时为一个周期,将全天运 输计划分为早班和晚班,少数地区可能存在24 小时为一个周期的情况。

问题模型与假设条件如下:

(1) 允许油罐车跨地级市选取就近油库进

收稿日期: 2022-12-09 修稿日期: 2023-01-10

作者简介:孙厚举(1991—),男,江苏徐州人,硕士,助教,研究方向为启发式算法、人工智能

行提油,各油库油品存量充足,不存在油品不 足的情况。

(2)油库可设定当前可用油品,并可设定 营业时段。

(3)每个地级市拥有大量加油站,且每个加油站都安装了液位仪,拥有近三年液位仪数据。

(4)油罐车分为汽油车和柴油车,汽油和 柴油不可混装在一辆车上,不同的汽油油品不 可以混装,但可以同时装在同一辆油罐车的不 同隔仓里。

(5)油罐车可以分载,一辆油罐车可以一次给多个加油站配送油品。

(6)拥有油库到各加油站的距离,加油站与加油站之间的距离。如果运距不存在,则用9999公里代替。

(7)油罐车下班后停车点为油库。

(8)拥有各种约束数据,例如加油站限制 拖车,限制挂车,限制配送时段等数据。

### 2 建立数学模型

数学符号定义如下:

 $P = \{p_0, p_1, \dots, p_n\}$ :整个规划时间段,其中 元素 $p_0$ 和 $p_1$ 代表第一个规划周期的开始和结束 时间, $p_2$ 和 $p_3$ 代表第二个规划周期的开始和结束 时间,以此类推。

 $Q = \{q_1, q_2, \dots, q_n\}:$ 一辆油罐车的隔仓容量, 其中n表示隔仓数量,  $q_n$ 表示第n个隔仓的容量。

 $K = \{Q_1, Q_2, \dots, Q_n\}$ :油罐车辆,其中 $Q_n$ 表示第*n* 辆油罐车的隔仓容量集合。

 $G = \{g_1, g_2, \dots, g_n\}$ :油品列表,其中 $g_n$ 表示 第n种油品的油品编号。

 $V = \{v_1, v_2, \dots, v_n\}$ : 油品体积, 其中 $v_n$ 表示 第*n*种油品的体积。

 $D = \{(G_1, V_1), (G_2, V_2), \dots, (G_n, V_n)\}$ :所有油 库信息,其中 $(G_n, V_n)$ 表示第n个油库的油品种 类以及可用体积。

 $S = \{s_1, s_2, \dots, s_n\}$ :所有加油站。

 $D = \{(G_1, V_1), (G_2, V_2), \dots, (G_n, V_n)\}: 所有加$ 油站的各种油品的当前体积。  $SV = \{\{S_1, V_1\}, \{S_2, V_2\}, \dots, \{S_n, V_m\}\}$ :各加油 站实际配送量。

 $VM = \{v_1, v_2, \dots, v_n\}$ :加油站油品的最大容纳体积,其中 $v_n$ 表示第n种油品的最大可容纳体积。

 $SM = \{(G_1, VM_1), (G_2, VM_2), \dots, (G_n, VM_n)\}$ : 所有加油站每种油品的最大可容纳体积。

 $T = \{t_1, t_2, \dots, t_n\}: 加油站油品的预计断货时间。$ 

 $O = \{(G_1, T_1), (G_2, T_2), \dots, (G_n, T_n)\}$ :所有加 油站的所有油品的预计断货时间。

 $DS = \{ ds_{1,1}, ds_{1,2}, \dots, ds_{i,j}, \dots, ds_{n-1,n} \}$ : 各运输 节点间的距离,其中 $ds_{i,j}$ 表示节点i到节点j之间 的距离,其中节点i和节点j可以是油库也可以 是加油站或者停车场。

 $R = \{r_1, r_2, \dots, r_n\}$ : 各加油站预计需求量。

v: 车辆行驶速度。

SL<sub>i</sub>:油罐车在油库提油时间。

ST<sub>i</sub>:加油站卸油时间。

c: 断货时长惩罚。

p: 不容纳体积惩罚。

根据以上描述,成品油公路运输多周期规 划目标定义如下:

$$TC = \sum_{j=1}^{m} \sum_{i=1}^{n} P_i \times K_i \times DS_{i,j}$$
(1)

$$OP = c \sum_{j=1}^{m} \sum_{i=1}^{n} \left( SL_i + ST_i + \frac{DS_{i,j}}{v} - T_i \right)$$
(2)

$$NP = p \sum_{j=1}^{m} \sum_{i=1}^{n} \left( SV_{ij} - SM_{ij} \right)$$
(3)

$$\min(TC + OP + NP) \tag{4}$$

式(1)表示规划方案的总运输成本;式(2) 表示规划方案的总断货惩罚;式(3)表示规划方 案的总不容纳惩罚;式(4)表示该问题的总目标 是实现所有周期中每辆油罐车的总运输成本、 断货惩罚和不容纳惩罚最小。

在实际运输方案中存在如下约束:

$$\sum_{i=1}^{n} Y_{vlji} = \sum_{i=1}^{n} X_{vlj} = 1$$
 (5)

$$\forall Q, 0.9 < P_a \tag{6}$$

式(5)表示限制每个加油站节点的入度和出 度都必须为1;式(6)表示限制油罐车的每个隔 仓满载率必须大于90%。此外还需要满足部分 其他约束条件,比如加油站可容纳约束,因为 油罐车的每一个隔仓在运输途中满载率要么为 0,要么必须大于90%,所以隔仓卸油时必须一 次卸完,不允许出现加油站油罐已满而隔仓未卸 完的情况。部分城市对油罐车有交通管制,要求 凌晨0点到6点通行,其他时段不允许通行。

### 3 混合遗传算法模型

### 3.1 算法编码方案

本文提出一种混合遗传算法(RMH),算法 采用一位编码方式表示配送方案,但因为配送 方案较为复杂,所以在编码时将每辆油罐车的 配送路线进行了封装,所有油罐车的配送路线 组合在一起就形成了完整的配送方案。其中, 每辆油罐车的配送路线都再次被细化为三部分, 即时间节点、配送路线和装载方案。时间节点 包括油罐车到达油库时间、等待油库营业时长、 装油时长、到达加油站时间、卸油等待时长、 卸油时间,如果存在分载情况,则提供到达下 一加油站时间以及卸油等待时长和卸油时间, 最后加上到达下一次提油油库时间。配送路线 包括油库和加油站。装载方案包括每个隔仓装 载的油品种类和装载量。如图1所示,该编码表 示某油罐车一次配送路线的编码。



图 1 油罐车一次配送路线编码

编码中将时间节点、配送路线、装载方案、 油罐车配送方案分别封装为类,则通过这些类 的对象即可表示多周期配送路线方案。

### 3.2 算法过程

求解成品油公路运输多周期规划问题的算 法流程如下。

```
initialSolution S

Sbest \leftarrow S

iter \leftarrow 0

while iter < stopNum do

if iter > swqpTh then

find better solution S'

end if

iter++

S \leftarrow S'

if iter >= maxIter then

V \leftarrow visitedNodeSet(S)

S \leftarrow OptimizedInventory(V)
```

```
Sbest = S + LKH(V)
end if
end while
```

return Sbest

#### 3.3 构造初始解

首先利用松弛下界模型将部分约束条件放 开,快速构造符合松弛模型的解,然后利用子 回路消除方法<sup>[15]</sup>将配送方案中的子回路消除, 再使用精确计算方法找到较为优秀的解放入初 始解集合。假设某配送规划问题中只包含1个油 库D,和16个加油站,记为 $\{s_1, s_2, \dots, s_{16}\}$ ,在松 弛下界模型下构造的解可能并不是一条连续路 径,而是多条子路径,其中多条子路径并不经过 油库D。每条路径用 $R\{s_i, s_j, \dots, s_n\}$ 表示,经过子 回路消除操作之后,得到一条经过所有节点的回 路,这条回路是开销最小的。如表1所示。

· 53 ·

表1 子回路消除过程

	规划路径
消除前:	$\begin{split} & R_1 \big\{ D, s_1, s_4, s_3 \big\}, R_2 \big\{ s_5, s_8, s_9 \big\}, R_3 \big\{ s_6, s_{10} \big\}, \\ & R_4 \big\{ s_{12}, s_{13}, s_{14} \big\}, R_5 \big\{ s_{15}, s_{16} \big\} \end{split}$
消除后:	$R\{D, s_1, s_4, s_5, s_9, s_{12}, s_{13}, s_{14}, s_{16}, s_{15}, s_{10}, s_6, s_3\}$

其中, s<sub>2</sub>、s<sub>7</sub>、s<sub>11</sub>三个节点在构造时就没有包括 在子回路中,子回路消除是通过依次计算子回 路之间的距离将子回路串联在一起,形成一条 回路,所以消除子回路后依然包含 s<sub>2</sub>、s<sub>7</sub>、s<sub>11</sub>三 个节点。通过松弛下界模型构造的规划路径 R<sub>1</sub> 至 R<sub>5</sub>,只有 R<sub>1</sub>是包含了油库的,其他规划路径 都没有取油油库,经过子回路消除之后合并成 为一条规划路径,然而一辆油罐车的油罐数量 一般为4~5个,也就是说,一辆油罐车到油库 提一次油,最多只能配送5个加油站,如果子回 路中加油站数量小于等于三个,则设定一个油 库即可。所以调整后的子回路消除算法的消除 过程如表2所示。

表 2 改进后子回路消除过程

	规划路径
消除前:	$\begin{split} & R_1 \{ D, s_1, s_4, s_3 \}, R_2 \{ s_5, s_8, s_9 \}, R_3 \{ s_6, s_{10} \}, \\ & R_4 \{ s_{12}, s_{13}, s_{14} \}, R_5 \{ s_{15}, s_{16} \} \end{split}$
消除后:	$\begin{split} & R_1 \big\{ D, s_1, s_4, s_3 \big\}, R_2 \big\{ D, s_5, s_8, s_9 \big\}, R_3 \big\{ D, s_6, s_{10} \big\}, \\ & R_4 \big\{ D, s_{12}, s_{13}, s_{14} \big\}, R_5 \big\{ D, s_{15}, s_{16} \big\} \end{split}$

目标函数<sup>[16]</sup>是启发式算法中最为关键的设 计,对于已经构造好的规划方案,通过目标函 数进行评估,淘汰较差的解,保留优秀的解, 同时配合模拟退火策略,按照一定的比例保留 不够优秀的解,加强算法的全局搜索能力。

### 3.4 迭代搜索

合法的初始解生成之后,通过启发式算法 的迭代进行局部搜索<sup>[17]</sup>,在解空间中不停地探 索更优秀的解,然后用更优秀的解替换解集中 较差的解,然而迭代搜索需要既保持当前解的 优点,又要实现解的优化,这在实际问题中很 难实现,也是设计迭代搜索的目的。首先我们 认为整个问题的解空间在多维坐标系中是连续 的,因此迭代时需要在解的邻域做搜索,所以 邻域操作<sup>[18]</sup>是启发式算法迭代的核心操作。 本文提供了两种测量用于调整配送方案的 领域操作。第一种邻域操作为简单的删除或者添 加更为经济的新节点,或者删除某个节点;第二 种是在不同配送路径中交换节点。第一种操作方 式较为简单,可以根据目标函数直接评估,决定 新增节点和删除节点。第二种操作方式不需要考 虑某个节点被多次配送或者未被配送。第二种邻 域操作方式的调整过程如表3所示。

表 3 交换子路径节点操作过程

	规划路径
交换前:	$\begin{split} & R_1 \Big\{ D, s_1, s_4, s_3 \Big\}, R_2 \Big\{ D, s_5, s_8, s_9 \Big\}, R_3 \Big\{ D, s_6, s_{10} \Big\}, \\ & R_4 \Big\{ D, s_{12}, s_{13}, s_{14} \Big\}, R_5 \Big\{ D, s_{15}, s_{16} \Big\} \end{split}$
交换后:	$\begin{split} & R_1 \Big\{ D, s_1, s_8, s_3 \Big\}, R_2 \Big\{ D, s_5, s_4, s_9 \Big\}, R_3 \Big\{ D, s_6, s_{16} \Big\}, \\ & R_4 \Big\{ D, s_{12}, s_{13}, s_{14} \Big\}, R_5 \Big\{ D, s_{15}, s_{10} \Big\} \end{split}$

为防止一次邻域操作导致解在解空间中偏移过多,建议一次邻域操作只交换1~2次节点。

## 4 实验结果及分析

为了验证算法的有效性,本研究对3个周期、5~50个客户的小规模算例和6个周期5~ 30个客户的中规模算例,以及6个周期,客户 数分别为50、100、200的大规模算例进行测试。 本节将对分支切割算法(branch-and-cut, BC)算 法<sup>[19]</sup>、核启发式算法(kernel search heuristic, KSH)算法<sup>[20]</sup>和本文提出的RMH算法进行对比。

在小、中规模算例中,本文提出的RMH算 法在多数算例上取得了理论最优解。与BC算法 和KSH算法的对比结果如图2所示。



图 2 小、中规模算例的对比结果

图2给出的是小、中规模算例计算结果的平 均值。其中BC算法为精确算法,在同样取得较 优解的情况下,所用时间最短。从图2不难看 出,对于3周期的算例来说,三种算法得到的最 优解的平均值几乎没有太大区别。中等规模算 例下,KSH算法和RMH算法的平均值较BC算 法稍微有一些优势,但总体差别不大,但计算 用时明显比BC算法要高。

对于大规模算例,周期达到6个,客户节点 最多达到200个。随着客户节点数量的增加, BC算法的劣势越发明显。大规模算例的对比结 果如图3所示。



图 3 大规模算例的对比结果

从大规模算例的求解结果可以看出,该研 究提出的RMH算法和KSH算法的平均结果明显 比BC算法更优秀,而且客户节点数量越多,差 距越明显,证明了精确算法在求解大规模库存 路径问题上存在缺陷。而RMH算法相比于KSH 算法依旧有明显优势,总运送成本平均降低 12.3%,而且平均计算时间也相对更短,通过实 验对比,从求解质量和时间代价两个方面体现 了RMH算法的性能和效率。

### 5 结语

针对成品油公路运输多周期规划问题,该 研究提出一种混合遗传算法模型RMH,该模型 可在多数小、中规模算例中求解到理论最优解。 相较于BC算法和KSH算法,RMH算法在大规 模算例集获取的规划方案中优势明显。当然, 该方案也存在一些不足。实际操作中,若后续 需人工调整5%左右的规划路线,则会影响其它 路线正常运输,牵一发而动全身,最终导致规 划方案无法正常使用。因此,如何获得100%可 用的规划方案是后续的重要研究工作。

### 参考文献:

- [1] 威晏旻.求解多周期库存路由问题的混合启发式算 法研究[D].武汉:华中科技大学,2019.
- [2] FORTNOW L. Nondeterministic polynomial time versus nondeterministic logarithmic space: time-space tradeoffs for satisfiability [C] //Proceedings of the 12th Annual IEEE Conference on Computational Complexity (CCC'97). IEEE Computer Society, USA,1997:52-60.
- [3] SAXENA R. Vendor-managed inventory [J]. Computerworld, 2013, 33(7):20.
- [4] 田立新,唐焕超.成品油单周期库存与运输联合优化[J].系统管理学报,2009,18(5):588-590.
- [5] 宋杰鲲,张在旭,张宇.成品油配送路线优化问题[J].油气储运,2008(11):28-30,62,12.
- [6] 张立峰,易万里,刘晓兰.基于两阶段算法的大规模成品油二次配送优化[J].系统工程理论与实践, 2016,36(11):2951-2963.
- [7] 李珍萍,胡婉霞,吴凌云.基于遗传算法的成品油二次配送车辆路径问题研究[J].数学的实践与认识, 2018,48(9):189-198.
- [8] 张雄,潘大志.融合邻域搜索策略蚁群算法求解带时间窗口的车辆路径问题[J].计算机与现代化, 2022(3):98-102,110.
- [9] XU X, LIN Z, LI X, et al. Multi-objective robust optimisation model for MDVRPLS in refined oil distribution [J]. International Journal of Production Research, 2022, 60(22):6772-6792.
- [10]周舟,邱玉琢.基于改进蚁群优化算法求解应急物 资库存路径问题[J].粮食科技与经济,2021,46 (6):76-79.
- [11] 杨华龙,辛禹辰,李志敏,等.需求周期循环波动下 考虑车辆租赁的库存路径与定价问题[J]. 计算机 集成制造系统,2022,28(6):1903-1912.
- [12] 吴方方.成品油运输物流企业成本控制问题研究 [D].呼和浩特:内蒙古财经大学,2019.
- [13] 李珍萍,焦鹏博,姜崇宇.考虑多车型软时间窗的成品油二次配送库存—路径问题[J].科学技术与工程,2022,22(18):8043-8049.

- [14] 李珍萍,杨光,韩倩倩.考虑工作量均衡的成品油二次配送车辆路径问题[J].系统仿真学报,2022,34
   (2):221-233.
- [15] 胡贤满.考虑库存策略的电商共享物流配送节点选 址算法[J]. 广州航海学院学报,2022,30(3):56-60.
- [16] ZHU Z Q, LIU Z, WANG X L, et al. Construction of integral objective function/fitness function of multiobjective/multi-displinary optimization[J]. Computer Modeling in Engineering and Sciences, 2004, 6 (6) : 567-576.
- [17] OEZTAS T, TUS A. A hybrid metaheuristic algorithm based on iterated local search for vehicle routing problem with simultaneous pickup and delivery [J]. Expert Systems with Application, 2022, 202:

117401.

- [18] 孙鉴,刘淞佐,武晓晓.基于大规模邻域搜索的模拟 退火算法求解TSP[J/OL].计算机仿真:1-7
   [2023-01-09]. http://kns. cnki. net/kcms/detail/11.
   3724.TP.20221230.1429.004.html.
- [19] TAMKE F, BUSCHER U. A branch-and-cut algorithm for the vehicle routing problem with drones[J]. Transportation Research Part B Methodological, 2021, 144(4): 174-203.
- [20] ARCHETTI C, GUASTAROBA G, HUERTA-MUOZ D L, et al. A kernel search heuristic for the multivehicle inventory routing problem [J]. International Transactions in Operational Research, 2021, 28 (6):2984-3013.

# Optimization of vehicle routing problem of petroleum products distribution based on hybrid genetic algorithm

### Sun Houju

(School of Telecommunications Engineering, Jiangsu Vocational Institute of Architectural Technology, Xuzhou 221000)

Abstract: Multi period planning of product oil road transportation is the core research topic of product oil supply chain, it belongs to inventory routing problem with multi period. In this study, the relaxation model is used to speed up the solution, after getting the set of legal solutions, combining mathematical programming model with improved heuristic algorithm, fast iteration in collections, greed mechanism is used to obtain the maximum profit scheme when solving the distribution volume. Finally, the performance of the algorithm has been tested on the instance sets of different count of periods and customers, this algorithm can obtain optimal of small and medium scale instances with far better time behavior, while the average transportation cost is reduced by 12.3% in large-scale calculation.

Keywords: road transportation of product oil; inventory routing problem with multi period; hybrid heuristic algorithm; relaxation model 文章编号:1007-1423(2023)02-0056-06

DOI: 10.3969/j.issn.1007-1423.2023.02.007

# 基于时间跨度注意力机制的多变量时间序列预测方法

李文豪,严华\*

(四川大学电子信息学院,成都 610065)

摘要: 在深度学习领域中,多变量时间序列预测的关键是充分捕获其长、短期的依赖。通常的做法是 使用卷积神经网络(CNN)来提取多变量时间序列的短期依赖。但是目前的大多数方法都仅仅使用了一种尺 寸的卷积核来提取短期依赖,这样就只提取到了一种时间跨度的短期依赖,使得序列的短期依赖关系没有 被充分学习到。受多尺度特征提取和通道注意力机制的启发,提出了一种基于时间跨度注意力的神经网络 (TSANN)方法,通过充分利用短期依赖来提高多变量时间序列预测的准确度。首先,使用一系列不同大小 的卷积核来提取不同时间跨度的短期依赖。然后使用时间跨度注意力将这些特征融合起来并送入循环神经 网络(RNN),进一步提取长期特征。另外,与其他方法直接将非线性部分和线性部分的输出相加不同,采 用了一个权重学习模块来整合非线性部分和线性部分的输出。在四个基准数据集上进行了实验,结果表明 TSANN能有效提高多变量时间序列预测的准确度。

关键词:时间跨度注意力;短期模式;权重学习;多变量时间序列预测

### 0 引言

多变量时间序列预测已被广泛应用于交通 预测<sup>[1-2]</sup>、汇率预测<sup>[3]</sup>、用电消耗预测<sup>[4-5]</sup>、河流 水位预测<sup>[6]</sup>等领域。如果能实现精准的预测并 应用在这些领域,将会帮助人们进行重大决策。

多变量时间序列预测的关键是充分捕获 其长、短期的依赖。现实世界的应用通常需 要短期和长期重复模式的混合<sup>[7]</sup>。然而传统 的方法,例如线性移动平均模型<sup>[8]</sup>、支持向 量机线性回归<sup>[9]</sup>大多采用统计模型来研究时 间序列的演化,可能无法捕捉到复杂的潜在 非线性关系<sup>[10]</sup>。递归神经网络(RNN)<sup>[11]</sup>是一 种专为序列建模而设计的深度神经网络,因 其在捕捉非线性关系方面的灵活性而受到广 泛关注。DA-RNN<sup>[12]</sup>使用了RNN来捕捉时间序 列的长期时间依赖,但它没有使用专门的方法 来提取时间序列的短期模式。在计算机视觉领 域,卷积神经网络(CNN)<sup>[13-14]</sup>模型通过成功提取 局部和平移不变特征表现出了在输入图像的不 同粒度级别方面的出色性能。LSTNet<sup>[7]</sup>利用了 CNN的长处去发现多维输入变量之间的短期模 式。然而,与LSTNet相同,近年来提出的网络 MTNet<sup>[10]</sup>和TPA<sup>[15]</sup>都只使用了一种大小的卷积核 来提取多变量时间序列的短期模式,这使得时 间序列的短期模式并没有被充分提取。

仅仅将不同时间跨度的短期模式提取出来还不够,还要选择性地加以利用。因为在对不同的数据集和预测范围进行预测时,模型对不同时间跨度的短期模式的依赖是不同的。虽然MTGNN<sup>[16]</sup>使用了不同尺寸CNN卷积核来提取不同范围内的时间特征,但是没有对不同卷积输出进行选择,而是将它们不加区分地拼接在一起。

针对上述问题,本文提出了基于多尺度特征表示和时间跨度注意力的网络结构。多变量时间序列首先经过包含多尺寸卷积核的CNN模块来获得不同时间跨度的短期模式,然后通过时间跨度注意力模块来为短期模式分配权重。 之后,将时间跨度注意力模块的输出送入RNN

**收稿日期:** 2022-09-28 修稿日期: 2022-10-17

作者简介:李文豪(1998—),男,湖北荆州人,在读硕士,研究方向为多变量时间序列预测;\*通信作者:严华(1971—), 男,四川渠县人,博士,教授,研究方向为智能信息系统,E-mail:yanhua@scu.edu.cn

模块来进一步提取长期模式。与LSTNet相同, 本文也引入了线性回归模块来提高模型的鲁棒 性。但与LSTNet直接将非线性部分和线性部分 的输出相加不同,本文采用了一个权重学习模 块来整合非线性部分和线性部分的输出。在四 个公开数据集上的实验结果表明,本文提出的 模型能有效提高多变量时间序列预测的准确度。

### 1 TSANN

### 1.1 概述

本文提出的时间跨度注意力神经网络(a deep

neural network based time span attention, TSANN) 的结构如图1所示。其主体结构分为线性和非线 性两部分。在非线性部分,将多变量时间序列 数据输入到有多个分支的多尺度卷积模块来提 取不同时间跨度的短期模式。多个分支的输出 被裁剪成相同尺寸后在通道维度上拼接。之后, 多尺度卷积模块的输出被送入不同的短期模式 赋予权重的时间跨度注意力模块。图1中时间跨 度注意力模块的输出被赋予了不同的权重。然 后,时间跨度注意力模块的输出被送入迭代模 块进一步提取长期模式。最后,通过全连接层 得到非线性部分的输出。



图 1 时间跨度注意力神经网络

取非线性部分输入中近期的数据作为线性 部分的输入。线性部分的输出通过简单的AR模 块得到。最后,通过权重学习模块将两部分的 输出加起来得到最后的预测结果。

## 1.2 多尺度卷积模块

多尺度卷积模块的目标在于充分提取时间 维度上不同时间跨度的短期时间模式和变量间 的局部依赖。在该模块的多个 CNN 分支中,不 同分支间 CNN 卷积核宽度 w 不同,高度 n 相同。 其中 w 表示不同的时间跨度, n 表示变量个数。 这样,模型就能提取到多时间跨度的短期模式。

多尺度卷积模块中包含6个分支,其卷积

核宽度 w 分别为 {3, 6, 7, 12, 24, 36}。每个分 支中都设置 N 个相同的卷积核。下文通过实验 说明 w 设置的依据。第 k 个分支中的第 i 个卷积 核扫描输入矩阵 X,并产生如下关系:

$$\boldsymbol{h}_{ki} = \text{RELU}(\boldsymbol{W}_{ki} * \boldsymbol{X} + \boldsymbol{b}_{ki}) \tag{1}$$

其中: \*表示卷积操作, 输出 $h_{ki}$ 是一个向量, RELU函数为RELU(x) = max(0, x)。

### 1.3 时间跨度注意力模块

当预测任务不同时(数据集不同或预测时间 步长Q不同),网络模型对不同时间跨度的短期 模式的依赖是不同的,所以要对不同时间跨度 的短期模式进行选择性的利用。 多尺度卷积模块一共提取到了6种不同时间 跨度的短期模式,每种短期模式有4个通道。所 以卷积模块的输出一共有24个通道,表示为  $H \in R^{24 \times T \times 1}$ 。时间跨度注意力推断出一个一维 通道注意地图 $M_s \in R^{24 \times 1 \times 1}$ 。整体注意过程可以 概括为

$$H' = M_s(H) \otimes H \tag{2}$$

其中, ⊗表示元素级相乘。



#### 图 2 时间跨度注意力模块示意图

图2描述了时间跨度注意力图的计算过程。 时间跨度注意力计算公式为

$$M_{s}(H) = \sigma(MLP(AvgPool(H)) + MLP(MaxPool(H)))$$
(3)

(2

# 其中, σ表示 sigmoid 函数。 1.4 循环神经网络模块

在经过多尺度卷积模块和时间跨度注意力 模块之后,模型已经提取到了不同重要程度的 短期模式。但时间序列中还存在更为长期的依 赖关系,能帮助提高预测的准确度。本文使用 了擅长记忆历史信息的GRU<sup>[17]</sup>来组成循环神经 网络模块。

时间跨度注意力模块的输出被输入到循环 模块中。在时间为t时,循环单元的隐藏状态计 算公式为

$$r_{t} = \sigma \left( x_{t} W_{xr} + h_{t-1} W_{hr} + b_{r} \right)$$

$$u_{t} = \sigma \left( x_{t} W_{xu} + h_{t-1} W_{hu} + b_{u} \right)$$

$$c_{t} = \text{RELU} \left( x_{t} W_{xc} + r_{t} \otimes (h_{t-1} W_{hc}) + b_{c} \right)$$

$$h_{t} = (1 - u_{t}) \otimes h_{t-1} + u_{t} \otimes c_{t}$$
(4)

其中:  $\otimes$ 表示元素级相乘;  $\sigma$ 表示激活函数;  $x_t$  为模块在时间t时的输出。

循环模块的输出为每个时间步隐藏状态, 表示为*h<sup>R</sup>*。将输出送入一个全连接层,以此来 获得模型非线性部分的预测结果。全连接层的 输出为

$$h_{\iota}^{D} = W^{D} h_{\iota}^{R} + b$$
(5)  
其中,  $h_{\iota}^{D}$ 为非线性部分的预测结果。

### 1.5 权重学习模块

预测结果的准确度不仅依赖于长期数据中包 含的周期,也依赖于短期数据的线性关系。目前 使用线性和非线性两部分进行预测的网络,都是 直接将两部分的预测结果相加得到最终预测结 果,但是这样做忽略了预测结果对两者的依赖程 度。为了更加准确地表示预测结果对线性部分和 非线性部分的依赖关系并提高预测精度,本文设 计了权重学习模块。该模块包含两个可学习参 数,分别来学习对线性和非线性部分的依赖程 度。TSANN最终的预测结果计算如下:

$$\hat{Y}_{t} = \alpha h_{t}^{D} + \beta h_{t}^{L} \tag{6}$$

其中, α和β为两个可学习的参数。

### 2 实验

为了验证TSANN的有效性,本文将TSANN 在四个常用基准数据集上与八种方法进行了对 比,其中包括四种传统方法和四种近几年提出 来的先进方法。

## 2.1 数据集描述

LSTNet 发布了四个公开可用的多元基准数 据集。表1展示了数据集的相关统计信息。为了 便于验证,所有的数据集按照时间顺序分为训 练集(60%)、验证集(20%)和测试集(20%)。

表 1 四个基准数据集的描述信息

数据集	样本/个	变量/个	采样频率
Traffic	17544	862	1小时
Solar-Energy	52560	137	10分钟
Electricity	26304	321	1小时
Exchange-Rate	7588	8	1天

### 2.2 对比方法

参与比较的九种方法的描述如下:

AR: 一个标准的自回归模型。

LRidge:具有 L2 正则化的向量自回归模型,曾经是最流行的多元时间序列预测模型。

GP:用于时间序列预测的高斯过程模型<sup>[18-19]</sup>。

RNN-GRU:使用GRU单元的RNN模型。

LSTNet:使用 CNN 提取短期模式和使用 RNN 提取长期模式的模型<sup>[7]</sup>。

MTNet: 一种使用记忆网络捕捉极长期模式的模型<sup>[10]</sup>。

TPA:使用时间模式注意力来选择相关时间序列的模型,并使用其频域信息进行多元预测<sup>[15]</sup>。

MTGNN:使用图神经网络和图学习模块自动提取变量之间的单向关系的模型<sup>[16]</sup>。

TSANN:本文提出的模型。

对于前面提到的单值输出方法,例如GP、 LRidge和AR,本文为每个变量训练一个模型。 例如Solar-Energy数据集有137个变量,因此本 文总共训练了137个模型。

### 2.3 实验指标

由于在相同的数据集上对TSANN与LSTNet、 MTNet、TPA和MTGNN进行比较,因此继续使 用相同的评估指标: RSE和CORR。

根相对平方误差(RSE):

$$RSE = \frac{\sqrt{\sum_{(i,t) \in \Omega_{Test}} (Y_{it} - \hat{Y}_{it})^2}}{\sqrt{\sum_{(i,t) \in \Omega_{Test}} (Y_{it} - \text{mean}(Y))^2}}$$
(7)

经验相关系数 (CORR):

$$CORR = \frac{1}{n} \sum_{i=1}^{n} \frac{\sum_{i} (Y_{ii} - \text{mean}(Y_{i})) (\hat{Y}_{ii} - \text{mean}(\hat{Y}_{i}))}{\sqrt{\sum_{i} (Y_{ii} - \text{mean}(Y_{i}))^{2} (\hat{Y}_{ii} - \text{mean}(\hat{Y}_{i}))^{2}}}$$
(8)

其中: Y表示真实值; Ŷ表示模型的预测值。对 于RSE, 值越低越好; 而对于CORR, 值越高越好。

### 2.4 实验结果

表2列出了所有实验的结果。遵循以往的惯 常做法,设置 Q = {3,6,12,24},并使用 RSE 和 CORR 两个指标来评估模型。Q 越大,预测难度 越大。在表2中,每个数据集和每个指标的最优 结果被加粗突出显示。在所有的32个最优结果 中,TSANN占了19个,远多于其他方法,证明 了本文提出方法的优越性。在其他方法中, MTGNN的最优结果占了10个,TPA占了2个, LRidge占了1个。

TSANN之所以能取得优于其他方法的结果, 主要是因为使用了多尺度卷积模块来充分提取 多变量时间序列的短期模式,并使用时间跨度 注意模块进行选择性的利用。

文献[7]对这四个数据集进行了分析,发现 Solar-Energy、Traffic和 Electricity都具有明显的 周期性,而 Exchange-Rate则没有。LSTNet在 Exchange-Rate上的表现不如AR、LRidge等传统

表 2 九种方法在四个数据集上的对比实验结约
------------------------

		数据集															
<del>-)-</del> >+	***		Solar-l	Energy			Tra	ffic			Elec	tricity			Exchan	ge-Rate	
刀広 拍怀		Q			Q			Q			Q						
		3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
AR	RSE	0.2435	0.3790	0.5911	0.8699	0.5991	0.6218	0.6252	0.6293	0.0995	0.1035	0.1050	0.1054	0.0228	0.0279	0.0353	0.0445
(0)	CORR	0.9710	0.9263	0.8107	0.5314	0.7752	0.7568	0.7544	0.7519	0.8845	0.8632	0.8591	0.8595	0.9734	0.9656	0.9526	0.9357
LRidge	RSE	0.2019	0.2954	0.4832	0.7287	0.5833	0.5920	0.6148	0.6025	0.1467	0.1419	0.2129	0.1280	0.0184	0.0274	0.0419	0.0675
(1)	CORR	0.9807	0.9568	0.8765	0.6803	0.8038	0.8051	0.7879	0.7862	0.8890	0.8594	0.8003	0.8806	0.9788	0.9722	0.9543	0.9305
GP	RSE	0.2259	0.3286	0.5200	0.7973	0.6028	0.6772	0.6406	0.5995	0.1500	0.1907	0.1621	0.1273	0.0239	0.0272	0.0394	0.0580
(0)	CORR	0.9751	0.9448	0.8518	0.5971	0.7831	0.7406	0.7671	0.7909	0.8670	0.8334	0.8394	0.8818	0.8713	0.8193	0.8484	0.8278
RNN-GRU	RSE	0.1932	0.2628	0.4163	0.4852	0.5358	0.5522	0.5562	0.5633	0.1102	0.1144	0.1183	0.1295	0.0192	0.0264	0.0408	0.0626
(0)	CORR	0.9823	0.9675	0.9150	0.8823	0.8511	0.8405	0.8345	0.8300	0.8597	0.8623	0.8472	0.8651	0.9786	<u>0.9712</u>	0.9531	0.9223
LSTNet	RSE	0.1843	0.2559	0.3254	0.4643	0.4777	0.4893	0.4950	0.4973	0.0864	0.0931	0.1007	0.1007	0.0226	0.0280	0.0356	0.0449
(0)	CORR	0.9843	0.9690	0.9467	0.8870	0.8721	0.8690	0.8614	0.8588	0.9283	0.9135	0.9007	0.9119	0.9735	0.9658	0.9511	0.9354
MTNet	RSE	0.1847	0.2398	0.3251	<u>0.4285</u>	0.4764	0.4855	0.4877	0.5023	0.0840	0.0901	0.0934	0.0969	0.0212	0.0258	0.0347	<u>0.0442</u>
(0)	CORR	0.9840	0.9723	0.9462	0.9013	0.8728	0.8681	0.8644	0.8570	0.9319	0.9226	0.9165	0.9147	0.9767	0.9703	0.9561	0.9388
TPA	RSE	0.1803	0.2347	0.3234	0.4389	0.4487	0.4658	0.4641	0.4765	0.0823	0.0916	0.0964	0.1006	0.0174	0.0243	0.0345	0.0444
(2)	CORR	0.9850	<u>0.9742</u>	0.9487	<u>0.9081</u>	0.8812	0.8717	0.8717	0.8639	0.9429	0.9337	0.9250	0.9133	<u>0.9790</u>	0.9709	0.9564	0.9381
MTGNN	RSE	0.1778	0.2348	0.3109	0.4270	0.4162	0.4754	0.4461	0.4535	0.0745	0.0878	0.0916	0.0953	0.0194	0.0259	0.0349	0.0456
(10)	CORR	0.9852	0.9726	<u>0.9509</u>	0.9031	0.8963	0.8667	0.8794	0.8810	0.9474	0.9316	0.9278	<u>0.9234</u>	0.9786	0.9708	0.9551	0.9372
TSANN	RSE	0.1753	0.2339	0.3073	0.4380	<u>0.4453</u>	0.4639	0.4521	0.4814	0.0813	0.0896	0.0907	0.1000	0.0191	0.0263	0.0339	0.0440
(19)	CORR	0.9858	0.9749	0.9541	0.9086	0.8837	0.8752	<u>0.8739</u>	0.8648	0.9478	0.9339	0.9329	0.9272	0.9795	0.9698	0.9568	0.9396

方法,是因为线性模型比深度学习方法更适合 该数据集。本文提出的方法能在Exchange-Rate 上取得最优结果,是因为使用了权重学习模块, 能够根据不同数据集动态调整模型对线性模块 和非线性模块的依赖程度。

本文方法在Traffic数据集上的表现不如 MTGNN,可能是因为MTGNN使用了图神经网络,使其具有更强的提取变量间依赖关系的能力。在这四个数据集中,Traffic具有最多的变量数(862),而MTGNN中复杂的图神经网络在变量数较大时表现出了一定的优势。

# 2.5 消融实验

为了验证本文所提方法中关键模块的有效 性,在Solar-Energy数据集上进行了充分的消融 实验,每次移除TSANN中的一个模块,并将它 们命名如下。

w/o MC: TSANN 去除多尺度卷积模块。按照LSTNet 的设置,只使用宽度为6的CNN 卷积 核来代替多尺度卷积模块进行短期模式的提取。

w/o TSA: TSANN 去除时间跨度注意力模块。不同时间跨度的短期模式被不加选择地利用。

w/o WL: TSANN去除权重学习模块。将线性部分和非线性部分的输出直接相加。

w/o AR: TSANN 去除线性模块。预测结果 仅由非线性部分输出得到。

在进行消融实验时,Q取难度最大值24。 实验结果如表3所示。

化口 伯威天型和小	表 3	消融实验结果
-----------	-----	--------

	TSANN	w/o MC	w/o TSA	w/o WL	w/o AR
RSE	0.4380	0.4732	0.4506	0.4616	0.4565
CORR	0.9086	0.8867	0.8948	0.8895	0.8937

消融实验结果表明,TSANN引入的关键模 块都提升了网络结构的性能。多尺度卷积模块 的引入极大地提升了实验结果,因为它能充分 提取不同时间跨度的短期模式来帮助模型进行 预测。权重学习模块带来的效果也很明显,因 为它能调整最终预测结果对线性部分预测结果 和非线性部分预测结果的依赖关系,实现更准 确的预测。实验结果也进一步验证了LSTNet中 提出的AR模块的有效性。时间跨度注意力模块 在多尺度卷积模块的基础上进一步提升了预测 结果的准确度,因为它对不同时间跨度的短期 模式进行了选择性的利用。

### 3 结语

本文提出了一种基于多尺度卷积和时间跨 度注意力的深度学习框架(TSANN),来进行多 变量时间序列的预测。TSANN中的多尺度卷积 模块能充分提取时间序列的短期模式。时间跨 度注意力模块能对这些短期模式进行选择性的 利用,以应对不同的数据集和未来时间步Q。 另外,权重学习模块能动态调整最终预测结果 对线性和非线性输出的依赖程度,从而提高预 测精度。大量的实验在四个公开数据集上进行, 实验结果证明了本文方法的有效性。

本文中短期模式的时间跨度是通过实验确 定的。未来的工作将致力于找到一种能根据数 据集和未来时间步Q自动确定时间跨度的方法。

参考文献:

- [1] ORESHKIN B N, AMINI A, COYLE L, et al. FC-GAGA: fully connected gated graph architecture for spatio temporal traffic forecasting [C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2021,35(10):9233-9241.
- [2] BAI L, YAO L, LI C, et al. Adaptive graph convolutional recurrent network for traffic forecasting[J]. Advances in Neural Information Processing Systems, 2020, 33: 17804-17815.
- [3] MARKOVA M. Foreign exchange rate forecasting by artificial neural networks [C] // Proceedings of the AIP Conference Proceedings. AIP Publishing LLC, 2019,2164(1):060010.
- [4] CHAN S, OKTAVIANTI I, PUSPITA V. A deep learning CNN and AI-tuned SVM for electricity consumption forecasting: multivariate time series data [C]//Proceedings of the 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). IEEE, 2019: 0488-0494.
- [5] KIM T Y, CHO S B. Predicting residential energy consumption using CNN-LSTM neural networks[J]. Energy, 2019, 182:72-81.
- [6] ZHU S, HRNJICA B, PTAK M, et al. Forecasting of water level in multiple temperate lakes using machine

learning models[J]. Journal of Hydrology, 2020, 585: 124819.

- [7] LAI G, CHANG W C, YANG Y, et al. Modeling long-and short-term temporal patterns with deep neural networks [C] //Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018:95-104.
- [8] MAKRIDAKIS S, HIBON M. ARMA models and the Box-Jenkins methodology[J]. Journal of Forecasting, 1997, 16(3): 147-163.
- [9] CAO L J, TAY F E H. Support vector machine with adaptive parameters in financial time series forecasting
   [J]. IEEE Transactions on Neural Networks, 2003, 14
   (6):1506-1518.
- [10] CHANG Y Y, SUN F Y, WU Y H, et al. A memory-network based solution for multivariate time-series forecasting [EB/OL]. arXiv Preprint arXiv:1809.02105,2018.
- [11] ELMAN J L. Finding structure in time[J]. Cognitive Science, 1990, 14(2):179-211.
- [12] QIN Y, SONG D, CHEN H, et al. A dual-stage attention-based recurrent neural network for time series prediction [EB/OL]. arXiv Preprint arXiv: 1704.02971,2017.
- [13] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Communications of the ACM,

2017,60(6):84-90.

- [14] LECUN Y, BENGIO Y. Convolutional networks for images, speech, and time series [J]. The Handbook of Brain Theory and Neural Networks, 1995, 3361(10):1995.
- [15] SHIH S Y, SUN F K, LEE H. Temporal pattern attention for multivariate time series forecasting[J]. Machine Learning, 2019, 108(8):1421-1441.
- [16] WU Z, PAN S, LONG G, et al. Connecting the dots: multivariate time series forecasting with graph neural networks [C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020:753-763.
- [17] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [EB/OL]. arXiv Preprint arXiv:1412.3555,2014.
- [18] FRIGOLA R. Bayesian time series learning with Gaussian processes [D]. University of Cambridge, 2015.
- [19] ROBERTS S, OSBORNE M, EBDEN M, et al. Gaussian processes for time-series modelling [J]. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2013,371(1984):20110550.

# A time span attention based deep neural network for multivariate time series forecasting

### Li Wenhao, Yan Hua\*

#### (College of Electronics and Information, Sichuan University, Chengdu 610065)

Abstract: In the field of deep learning, the key to multivariate time series forecasting is to fully capture its short-term and long-term patterns. The usual approach is to use convolutional neural network (CNN) to extract the short-term patterns. However, most of the current methods only use CNN filters with one size, so that only one kind of short-term pattern of one time span is extracted, which makes the short-term patterns of the series not fully extracted. Inspired by multi-scale feature representation and channel attention, this paper proposes a deep neural network based time span attention(TSANN) to make full use of short-term patterns to improve the accuracy of multivariate time series forecasting and the universality of the model. First, this paper proposes to apply a set of CNN filters with multiple sizes to extract short-term patterns of various time spans. Then, this paper proposes time span attention to fuse these patterns and feed them to recurrent neural network (RNN) to further extract long-term patterns. In addition, unlike other methods which add the output of the linear and the non-linear part directly, the weight learning module is proposed to integrate them. Extensive experiments are conducted on 4 benchmark datasets, and experimental results show that our method can effectively improve the accuracy of multivariate time series forecasting.

Keywords: time span attention; short-term patterns; weight learning; multivariate time series forecasting

文章编号:1007-1423(2023)02-0062-06

DOI: 10.3969/j.issn.1007-1423.2023.02.008

# 基于Pareto支配的高维多目标优化算法的分析与研究

操心慧\*,许丽娟

(广州华商学院数据科学学院,广州 511300)

摘要:多目标进化算法(MOEAs)因其处理多目标优化问题的有效性和高效性而得到广泛的认可。近年来,大量研究表明,在处理三个以上目标的多目标优化问题(MaOPs),即高维多目标问题时,随着非支配解比例的增加,MOEAs的行为类似于搜索空间中的随机漫步。这种现象在大多数经典的基于Pareto优势的MOEAs(PDMOEAs)中都很常见,例如NSGA-II、SPEAII,这些算法由于缺乏选择压力,很难将搜索过程引导到最优Pareto前沿。因此,针对基于Pareto支配的高维多目标优化问题进行分析与研究。

关键词:高维多目标优化;非支配解; Pareto支配

# 0 引言

多目标优化问题(multi-objective optimization problems, MOPs)是指具有多个目标且在本质上 存在冲突,需要同时进行优化的问题。多目标优 化的概念在实际应用中经常被研究,如软件工程、 配电网络和工业调度[1]。多目标优化方案中各目标 的冲突行为,导致获得一组非支配解,称为Pareto 最优集,即目标空间<sup>[2]</sup>。多目标优化算法 (MOEAs)在求解多目标优化问题时是有效的,因 为它们能够在单次运行中获得 Pareto 最优集。 MOEAs 在求解 MOPs 时的主要目的是在收敛性 (得到的Pareto前沿与真实Pareto前沿的接近程度) 和多样性<sup>[3]</sup>(得到的Pareto前沿中种群的均匀分布) 之间取得平衡。MOEAs能否在收敛性和多样性之 间进行权衡,主要取决于所采用的选择策略。按 照机制将其大致分为: Pareto 支配型 MOEAs(PD-MOEAs)、分解型 MOEAs、指标型 MOEAs<sup>[4]</sup>,而 其中Pareto支配型MOEAs最为流行。

然而,大多数经典的基于 Pareto 支配的 MOEAs(PDMOEAs)在多目标优化问题上表现较

好,但在处理超过三个目标的问题,也就是高 维多目标优化问题(MaOPs)时,效果大大折扣。 PDMOEAs在解决多目标优化问题时,由于难以 在收敛性和多样性之间取得平衡,其性能急剧 下降。导致 PDMOEAs性能下降的主要原因是 Pareto支配关系的失效,后续多样性的维护操作 占据了主导位置,这种现象降低了选择压力, 不能引导搜索过程向收敛方向发展。也可以理 解为由于目标空间维度的增加,MOEAs的行为 采用了随机搜索,大多数个体变得无法比较, 而解集无法收敛到 Pareto 最优前沿<sup>[5]</sup>。

此外,随着目标数量的增加,在多样性和 收敛性之间取得平衡成为一项艰巨的任务。由 于计算复杂度的问题,MOEAs中使用的种群规 模有限,个体最终会在高维空间中彼此远离, 导致进化效率低下。因此,对于MaOPs来说, 设计一种算法使其稳定有效地找到收敛性和分 布性都不错的Pareto最优解集仍有一定难度。目 前进化计算的权威期刊《IEEE Transactions on Evolutionary Computation》和《Evolutionary Com-

**收稿日期:** 2022-09-21 修稿日期: 2022-10-17

基金项目: 2019年广东省普通高校特色创新类项目(2019KTSCX236):基于深度学习的楼宇巡查机器人视觉图像 识别分类研究

作者简介:\*通信作者:操心慧(1995—),女,安徽安庆人,硕士,助教,研究方向为数据可视化、多目标优化,E-mail: 945152118@qq.com;许丽娟(1980—),女,青海人,硕士,副教授,研究方向为数据挖掘算法、计算机视觉 putation》每年都发表一系列关于高维多目标优化的研究成果,并呈逐年递增趋势。在国内, 该领域的研究始于2010年,迄今为止不仅论文 发表数量不多而且研究成果远滞后于国外,属 于起步阶段,亟待加大研究投入,所以在国内 开展MaOEAs的研究具有十分迫切的需求。

### 1 高维多目标优化问题概述

### 1.1 高维多目标优化问题(MaOPs)

在解决多目标优化问题时,有些问题是难 以解决的。这是因为多目标优化问题(MOPs)不 同于单目标优化(SOPs),在最优解上多目标优 化问题具有多个最优解,而高维多目标问题是 普通多目标问题的复杂情况。

对于一个具有 k 个目标, n 维决策变量的 MOPs, 其数学模型<sup>[6]</sup>为

$$\begin{cases} \min \quad \mathbf{F}(x) = \left(f_1(x), f_2(x), \cdots, f_k(x)\right)^{\mathsf{T}} \\ \text{s.t.} \quad g_i \le 0 \, ; \, i = 1, 2, \cdots, p \\ h_i = 0 \, ; \, i = 1, 2, \cdots, q \end{cases} \tag{1}$$

其中,  $\mathbf{x} = (x_1, x_2, \dots, x_k) \in \Omega$ 被称为k维决策向 量,为一组个体。当 $k \ge 4$ 时为高维多目标优 化问题,在目标进化中优化函数向量 min  $\mathbf{F}(x) = (f_1(x), f_2(x), \dots, f_k(x))^T$ , k是目标函 数的个数。求解高维多目标优化问题的最终目 的是得到一组无限收敛于真实 Pareto 前沿,并且 均匀分布的 Pareto 解集。

### 1.2 解决高维多目标优化问题面临的问题

常规多目标优化算法能够有效解决两个或 三个目标的多目标优化问题,但将算法拓展到 MaOPs(高维多目标优化问题)时,其求解过程 将主要面临以下三个问题:

(1)随着目标维数的增加,对于给定的种 群,该种群中非支配解的数量占其种群规模的 比例接近于1。因此,Pareto支配关系不适合在 这类问题的解之间进行区分。MaOPs面临的另 一个问题是,近似整个Pareto前沿所需解的数 量,正常情况下会随着目标函数的数量呈指数 增长。因此,需要更大的种群规模来获得适当 的Pareto集近似,这又需要更长的执行时间。

(2) 随着目标维数的增加,如果 MaOPs 的

大多数目标高度相关,则其Pareto前沿不太可能 在高维目标空间的广泛范围内扩散。此外,如 果它们依赖于少数相互矛盾的目标,则目标空 间中Pareto沿的维数可能远远小于目标的数量。 除了目标的数量及其相互关系之外,实验表明 当每个目标的目标值范围不同时,一些专门设 计用于处理MaOPs的MOEAs可能无法获得预期 的性能。因此,需要进行目标归一化。然而, MaOPs缺乏一种有效的、鲁棒的归一化方法。 此外,观察表明,归一化在性能上的影响具有 很强的问题依赖性。

(3)多样性是另一个需要考虑的方面。虽然 非支配解会影响收敛性,但保留合理数量的解 可能有利于种群多样性。目标数量不断增加所带 来的巨大目标空间,为在MaOPs上保持解集均匀 分布带来了几个挑战。在这些挑战和研究机会 中,仍然需要开发适当的多样性保存机制、适当 的拥挤距离机制和最优解集的最终评估方法。

### 2 PDMOEAs的改进方法

针对PDMOEAs在处理高维多目标优化问题 时所出现的性能缺陷,专家学者们对其进行改进,改进的方向大致分为以下三类。

### 2.1 优势关系的松弛

这一类思想侧重于放松支配关系,即通过 引入新的支配关系来修改传统 Pareto支配关系的 定义。引入改良显性关系的主要目的是提高两 个体在 MaOPs上的可比性。

文献[7]提出了一种新的控制个体优势区域 (CDAS)的方法,以提高 MOEAs 的性能。CDAS 算法通过控制优势区域,诱导出合适的个体排 序,增强了选择过程。文献[10]提出了一种自 适应外部参数 S和自控制个体优势区(S-CDAS) 的 CDAS 方法来解决 MaOPs问题。在 S-CDAS算 法中提出的主要改进是,S-CDAS采用了细粒度 排序,将极限解始终包含在队列的最前面。文 献[8]提出了一个广义 Pareto 最优(GPO)来处理 PDMOEAs 的可扩展性问题。GPO 的主要思想 是,随着问题维度的增加,解决方案的主导区 域逐渐扩大。广义 Pareto 最优方法与控制优势区 域扩张程度的扩展角参数相关联。文献[9]提出 了 Pareto 支配关系的模糊化方法,并分析了该方 法在 MOEAs 设计中的适用性。同时, 描述了一 种通用排序方案,该方案以集合依赖、非对称 和尺度无关的方式分配具有支配度的任意向量 集。文献[10]引入了一种新的适应度评价机制, 将解连续划分为不同程度的最优解。在适应度 评价机制的基础上,基于模糊逻辑提出了一种 模糊 Pareto-dominance (FD),并将其整合到流行 的NSGA-II和SPEA2中, 增加解决MaOPs上的性 能。文献[11]提出了一种对传统 Pareto-dominance 的修正,称为g-dominance,可以被纳入到任何 MOEA 的设计中。g-dominance 利用了纳入参考 点的信息,没有任何缩放函数的帮助,逼近最优 区域周围的有效集。文献[12]提出了 $\varepsilon$ -MOEA, 该ε-MOEA采用了一种新的 Pareto 优势变体 ε-优势与存档和选择策略相结合,以提高对Pareto最优集的搜索。文献[13]提出了强化优势关 系(SDR), 它是对 Pareto 优势关系的一种新的加 强方式。然而,改良优势度关系的 PDMOEAs 处 理 MaOPs 的性能有所提高,但往往会出现局部 最优情况。

### 2.2 使用附加指标结合Pareto优势

这类思想集中于将附加的指标与Pareto优势 相结合,以促进PDMOEAs的融合。许多工作集 中在发展有效的额外指标,以提供适当的均衡 之间的收敛性和多样性。

文献[14]提出了一种 MOEA (NSGA-III), 该 MOEA 生成一组均匀分布的参考点,并优先 考虑与参考点垂直距离最小的非支配解,以保 持多样性。换句话说,在NSGA-III中,多样性 的保存机制借助于一组分布良好的参考点。在 NSGA-III的环境选择中,合并种群中的每个个 体都与一个参考点关联,与参考点距离最小的 解将被选择。文献[15]引入了一种多样性机制, 基于移位的密度估计(SDE)指标,旨在保持多 样性而不失去收敛性。SDE的基本思想是通过 相对于每个目标的收敛性比较,移动其他个体 的位置来估计候选解的密度。因此,收敛性能 较差的个体会有高密度值。文献[16]提出了基 于 Pareto-dominance 和排序方法的 MOEA, 在总 体中给每个候选解分配一个优先级排序。根据 ①非支配排序得到的解的 Pareto 排序给每个候 选解分配优先级排序; ②归一化目标和; ③改 变生态位半径。优先级最低的个体在交配和环 境选择中更受青睐,因为它们加速了向 Pareto 前沿的收敛,同时保持了多样性。文献[13]提 出了一种基于矢量角度的 MaOPs-EA,它采用 了 Pareto 优势与最大矢量角度优先原则相结合。 除了这些指标,还采用了劣解消除原则。文献 [17]基于有利收敛(FC)和方向多样性(DD)提出 了一种多目标进化算法(MOEA-DDFC)。在 MOEA-DDFC 算法中,引入了良好的收敛性和方 向多样性,使算法的收敛性和多样性同时提高。

### 2.3 以支配与分解相结合为基础

最后一类是将基于支配的方法与基于分解 的算法相结合。文献[18]提出了一个统一的范 式,采用基于优势的方法与基于分解的方法相 结合来解决 MaOPs(MOEA-DD)。在 MOEA-DD 中,利用一组权值将目标空间划分为不同的子 区域,每个权值向量定义一个适合度评估子问 题。文献[19]提出了一种双准则进化(BCE)框 架,将 Pareto 准则(PC)方法与非 Pareto 准则 (NPC)方法相结合,增强了收敛性和多样性。 双标准方法试图通过充分交换信息的协作方式 将 PC和NPC方法的优势结合起来,以促进彼此 的进化。另一方面,双标准进化试图弥补彼此 的弱点。

### 3 ad-MOEA

上节提到对于 PDMOEAs 算法的改进,最终的目的都是想要均衡算法的收敛性和多样性。 基于此,本节详细介绍一种具有自适应交配和 环境选择的多目标优化进化算法(ad-MOEA), 利用自适应交配和环境选择的策略自适应地平 衡算法的收敛性和多样性。

### 3.1 ad-MOEA的总体框架

该算法的总体框架类似于经典的NSGA-II。 在该方法中,随机生成一个大小为N的初始种 群,并对其初始化。初始化后,得到归一化目 标(SoNB)和拥挤距离(CWD)的Pareto最优解集。 根据这些准则计算每个候选解的加权级,然后 进行交配选择,随机选择两个亲本解,并根据 权重进行比较,选择一个权重最小的解作为后 代的亲本。在交配选择后,通过突变和重组操 作生成大小为N的后代种群。

将子代种群与父代种群合并,对合并种群中的每个个体,得到每个候选解的Pareto排序和归一化目标之和(SoNB)。环境选择过程中,在Pareto优势之后,个体的选择一直持续到临界前沿。

在临界前沿,一定概率下基于目标之和选择候选解,根据拥挤距离选择剩余的候选解。选择候选解的概率由一个自适应参数 w 决定。最初,参数"w"的值设置为1,因为在初始代中需要收敛。然后根据问题的特点,自适应参数 w 的值。但随着进化到需要增强多样性的最后阶段,"w"的值预计会下降。图1 描述了所提方法的框架,在 NSGA-II 框架的基础上结合归一化目标之和 SoNB 以及拥挤距离 CWD 得到ad-MOEA 算法的总体框架。



图1 ad-MOEA总体框架

### 3.2 交配选择

交配选择过程在 MOEAs 的进化过程中起着 重要作用。在产生过程中,筛选有前景子代进 化的解决方案将提高算法的性能。该算法在交 配选择中引入了加权的概念,为了得到加权排 序,首先对个体进行 Pareto 排序和归一化目标之 和的升序排序,并按照排序后的顺序给每个个 体分配一个排序,根据排序结果,选择有前景 的子代进化产生后代种群。算法1描述了交配选 择的过程。

```
算法1 交配选择(P, F, SoNB, CWD)
输入: P(个体)
1 计算每个个体的加权等级(WR)
2 S为非空集
3 |S| < N do</li>
4 从P中随机选取x和y
```

- 5 if WR(x) < WR(y) then 6  $S \leftarrow S \cup \{x\}$ 7 else if WR(x) > WR(y) then 8  $S \leftarrow S \cup \{y\}$ 9 else 10 if rand(1) < 0.5 then 11  $S \leftarrow S \cup \{x\}$ 12 else 13  $S \leftarrow S \cup \{y\}$
- 14 循环结束 15 return S

### 3.3 环境选择

环境选择的主要目的是为下一次迭代保留 作为父种群的精英。在环境选择中,首先将交 配选择后获得的子代种群与父种群进行组合。 然后对合并后的种群采用Pareto优势法,将个体 分配到不同的非优势Pareto前沿。根据Pareto-Dominance算法,得到了每个候选解的归一化目 标和拥挤距离之和。

首先,来自第一个非支配 Pareto 前沿  $F_1$ 和 判断 $|F_1| > N$ ,  $F_1$ 中的个体是基于考虑标准化目标总和(SoNB)和拥挤距离(CWD)的自适应方法 进行选择。如果前面个体解的大小( $F_1$ )小于N,则从第二个非支配 Pareto 前沿  $F_2$ 中选择解;如 果 $|F_1 \cup F_2| > N$ ,就将 Pareto 前沿  $F_2$ 视为非支配 Pareto 前沿,并采用自适应方法选择第二非支配 Pareto 前沿的解。对于下一代而言,在获得具有 精英解的N大小种群之前,遵循的相同过程选择非支配 Pareto 前沿的个体。本文采用了自适应 参数w,为了调整参数,利用关于组合总体中非 支配解数量的信息,自适应参数w采用如下公 式计算。

$$w_{i} = \left(0.99 \times w_{i-1}\right) + 0.01 \times \left(1 - \left(\frac{N_{F}}{N}\right)^{\frac{1}{M}}\right) (2)$$

其中: w<sub>t</sub>表示 t 代参数 w 的值; w<sub>t-1</sub>表示上一代 中w 的值; M 为目标个数; N<sub>F</sub>表示总体中非支 配解的数量, N 表示个体总数量。首先,基于 归一化目标与百分比(w × 100)所需的解决方案 是根据归一化目标的总和来选择的。然后根据 拥挤距离选择剩余解,因此参数 w 在环境选择 中起着关键作用。

# 4 结语

本文对解决高维多目标优化问题的算法进 行了研究与分析,重点介绍了一种具有自适应 交配和环境选择的多目标进化算法(ad-MOEA) 以及处理多目标和多目标优化问题的过程。该 方法将归一化目标之和的概念引入交配选择和 环境选择中,用以均衡收敛性和多样性,从而 得到最优结果。环境选择中,在临界前沿,首 先根据目标之和选择方案,然后根据拥挤距离 选择剩余方案。为了选择基于归一化目标和的 解,本文采用了由自适应参数决定的一定概率, 与现有的NSGA-II算法相比,该算法的性能有 所提高,与其他先进算法相比具有竞争力。尽 管如此,还无法完全解决PDMOEA在高维多目 标优化上遇到的问题。

### 参考文献:

- [1] DEB K, JAIN H. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints [J]. IEEE Transactions on Evolutionary Computation, 2013, 18(4):577-601.
- [2] 梁正平,刘程,王志强,等.基于存档和权值扩展的 大规模多目标优化算法[J]. 计算机学报,2022,45
   (5):951-972.
- [3] 郑金华,邹娟.多目标进化优化[M].北京:科学出版社,2017:90-94.
- [4] YU K, BUDHIRAJA A K, BUEBEL S, et al. Algorithms and experiments on routing of unmanned aerial vehicles with mobile recharging stations [J]. Journal of Robotic Systems, 2019, 36(3):602-616.
- [5] 孙浩,杨景明,刘醒,等.基于环境 Pareto 支配选择 策略的有约束多目标差分进化算法[J].控制与决 策,2016,31(1):45-51.
- [6] 刘建昌,李飞,王洪海,等.进化高维多目标优化算
   法研究综述[J].控制与决策,2018,33(5):879-887.
- [7] GU Z M, WANG G G. Improving NSGA-III algorithms with information feedback models for large scale many-objective optimization[J]. Future Generation Computer Systems, 2020, 107:49-69.
- [8] 陈振兴,严宣辉,吴坤安,等.融合张角拥挤控制策
   略的高维多目标优化[J].自动化学报,2015,41
   (6):1145-1158.

- [9] DEB K, PRATAP A, AGARWAL S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II[J]. IEEE Transactions on Evolutionary Computation, 2002,6(2):182-197.
- [10] GUO Y, ZHU X, DENG J, et al. Multi-objective planning for voltage sag compensation of sparse distribution networks with unified power quality conditioner using improved NSGA - III optimization [J]. Energy Reports, 2022, 8:8-17.
- [11] PALAKONDA V, MALLIPEDDI R. An evolutionary algorithm for multi and many-objective optimization with adaptive mating and environmental selection [J]. IEEE Access, 2020, 8:82781-82796.
- [12] 杨五四,陈莉,王毅,等.一种适应度排序的高维多 目标粒子群优化算法[J].西安电子科技大学学报, 2021,48(3):78-84.
- [13] TIAN Y, CHENG R, ZHANG X, et al. A strengthened dominance relation considering convergence and diversity for evolutionary many-objective optimization [J]. IEEE Transactions on Evolutionary Computation, 2018, 23(2): 331-345.
- [14] PAMULAPATI T, MALLIPEDDI R, SUGANTHAN P N. I<sub>SDE</sub>+: an indicator for multi and many-objective optimization [J]. IEEE Transactions on Evolutionary Computation, 2018, 23(2): 346-352.
- [15] MOLINA J, SANTANA L V, HERNÁNDEZ -DÍAZ A G, et al. g-dominance: reference point based dominance for multiobjective metaheuristics [J]. European Journal of Operational Research, 2009, 197 (2):685-692.
- [16] LAUMANNS M, THIELE L, DEB K, et al. Combining convergence and diversity in evolutionary multiobjective optimization [J]. Evolutionary Computation, 2002, 10(3):263-282.
- [17] LI M, YANG S, LIU X. Shift-based density estimation for Pareto-based algorithms in many-objective optimization [J]. IEEE Transactions on Evolutionary Computation, 2013, 18(3): 348-365.
- [18] XIANG Y, ZHOU Y, LI M, et al. A vector anglebased evolutionary algorithm for unconstrained many-objective optimization [J]. IEEE Transactions on Evolutionary Computation, 2016, 21(1):131-152.
- [19] CHENG J, YEN G G, ZHANG G. A many-objective evolutionary algorithm with enhanced mating and environmental selections [J]. IEEE Transactions on Evolutionary Computation, 2015, 19(4):592-605.

# Analysis and research of many-objective optimization algorithm based on Pareto-domination

Cao Xinhui\*, Xu Lijuan

(School of Data Science, Guangzhou Huashang College, Guangzhou 511300)

Abstract: Multi-objective evolutionary algorithms (MOEAs) have been widely recognized for their effectiveness and efficiency in handling multi-objective optimization problems. In recent years, numerous studies have shown that when dealing with multi-objective optimization problems (MaOPs) with more than three objectives, i.e. many-objective problems, as the proportion of non-dominated solutions increases, MOEAs behave like random walks in the search space. This phenomenon is common in most of the classic Pareto advantage-based MOEAs (PDMOEAs), such as NSGA–II, SPEAII, which are difficult to steer the search process to the optimal Pareto front due to the lack of selection pressure. Therefore, analyzes and studies the multi-objective optimization problem with more than three objectives.

Keywords: many-objective optimization; non-dominated solution; Pareto-domination

(上接第22页)

# Forecast study on carbon emissions of civil transport aircraft based on time series

Xiang Xiaojun<sup>1</sup>, Yang Zhihan<sup>2\*</sup>, Zhao Ganchao<sup>2</sup>

Scientific Research Department, Civil Aviation Flight University of China, Guanghan 618300;
 Flight Technology College, Civil Aviation Flight University of China, Guanghan 618300)

Abstract: With the rapid development of China's civil aviation industry, the carbon emission of transport aircraft has gradually attracted attention. The research uses the time series method to establish the traditional Autoregressive Integrated Moving Average(ARIMA) model and the optimized Long Short Term Memory (LSTM) model to predict the carbon emissions, carbon intensity and carbon emissions per ton kilometer of the aircraft. The learning rate and hidden node number in LSTM are optimized through the Whale Optimization Algorithm(WOA), avoiding the subjectivity and blindness of manually selecting parameters, it is helpful to improve the accuracy of model prediction. By comparing the Root Mean Squared Error(RMSE) and Mean Absolute Error (MAE) of the two models, ARIMA model performs well in aircraft carbon emission prediction, and WOA-LSTM model performs well in carbon intensity and ton kilometer carbon emission prediction.

Keywords: time series; ARIMA; WOA-LSTM; carbon emission

现代计算机 Modern Computer

文章编号:1007-1423(2023)02-0068-05

DOI: 10.3969/j.issn.1007-1423.2023.02.009

# 基于词向量与TextRank的政策文本关键词汇抽取方法研究

李 晨<sup>1</sup>, 赵燕清<sup>1</sup>, 于俊凤<sup>1</sup>, 张铭君<sup>1</sup>, DMYTRO LANDE<sup>1,2</sup>

(1.齐鲁工业大学(山东省科学院)情报研究所,济南 250014;2.乌克兰国立技术大学信息与计算机科学学院,基辅 03056)

摘要:通过对政策文本进行分析,设计了一种基于机器学习的关键词汇抽取方法,该方法可以自动从 政策中提取关键性词语或短语。首先,从互联网采集相关政策文本并与维基百科数据融合,利用 fastText构 建词向量;其次,综合考虑词语的位置信息和词语之间的语义相似度,共同构建转移矩阵;最后选择得分 最高的K个词语作为政策关键词。结果表明提出的方法抽取效果较好,实用性较高。

关键词:政策;词向量;关键词提取;TextRank

# 0 引言

政策通常是指政府、机构、组织为实现目标而订立的计划。政策文献是政策的物化载体, 是政府处理公共事务的真实反映和行为印迹, 是对政策系统与政策过程客观的、可获取的、 可追溯的文字记录<sup>[1]</sup>。关键词是对文本的高度 概括和抽象,能够帮助人们快速了解政策全文 信息。因为政策本身的特殊性,原文中并不会 设置关键词字段,如果可以采用自动的方式提 取出与主题相关的词语或短语,则可以更好地 辅助政策解读。

目前政策文本关键词提取算法大都以开源 分词工具为基础,结合词频统计和人工辅助来 实现。如吴宾等<sup>[2]</sup>利用开源工具人工提取海洋 工程装备制造业政策主题词;吴爱萍等<sup>[3]</sup>运用 扎根理论从政策样本中提取高频关键词。如果 可以利用机器学习的方式自动抽取政策关键词, 那么就可以进一步提高政策分析效率。现有的 基于机器学习的关键词提取算法大体可以分为 两类,分别是有监督提取方法和无监督提取方 法。有监督的提取方法是一种分类方法,需要 人工提前设置好训练集,然后训练出分类模型, 最后通过分类模型完成关键词提取。如果训练 集质量较高,用此方法可以得到比较好的结果, 但是这种方式需要人工的参与,总体来说效率 低、代价大。无监督的关键词抽取方式目前主 要有三种:基于统计方法(TF-IDF)的抽取方式、 基于主题模型(LDA<sup>[4]</sup>)的抽取方式和基于图模型 (TextRank)的抽取方式。

### 1 相关工作

基于TF-IDF的抽取算法是较为简单的一种 实现。该方法以词频统计为基础,按照某个词 在单篇文档中出现的次数和在所有文档中出现 的文档频率进行计算。该方法可以过滤一些常 见的无关紧要的词语,同时还能保留区分度较

基金项目:山东省重点研发计划(软科学项目)(2021RZA01017):山东省科技型企业发展现状及对策研究;齐鲁工业大学(山东省科学院)科教产融合创新试点工程项目(2022GH015):基于"双过程模型架构"的认知图谱关键技术及应用研究

**收稿日期:** 2022-09-14 修稿日期: 2022-10-21

作者简介: 李晨(1988—),男,山东济南人,硕士,图书馆员,研究方向为大数据与数据挖掘;赵燕清(1971—),女,山 东济南人,硕士,研究馆员,研究方向为智能决策与情报分析;于俊凤(1979—),女,山东济南人,硕士,副研究员,研究方向 为情报分析;张铭君(1990—),女,山东济南人,硕士,图书馆员,研究方向为情报分析; Dmytro Lande(1959—),男,乌克兰 基辅人,博士,教授,研究方向为智能信息系统
高的重要词语。张骁等<sup>[5]</sup>就利用TF-IDF算法结 合实际对科技服务业政策文本的关键词进行了 提取。方法虽然易于实现但是缺点也很明显, 单纯以词频衡量词的重要性,不够全面,同时 也不能反映词的位置信息。为了克服这些缺点, 很多人对TF-IDF算法进行了改进。张瑾<sup>[6]</sup>在原 有的算法基础上加入位置权值及词跨度权值, 避免了单纯采用TF-IDF算法产生的偏差。

LDA模型在自然语言领域被大规模应用, 该技术同样也适用于文本关键词抽取。基于 LDA的实现方式需要对数据集进行训练得到主 题模型,选取能够反映主题的词语作为候选关 键词,这种方式抽取的关键词很大程度上依赖 训练数据的主题分布情况。

基于图模型的关键词提取算法近年来研究 较多,借鉴 PageRank<sup>[7]</sup>算法思想进行改进与扩 展。该方法是将文本转化为相关词的词语网络 图,该图的节点是词,边是词语之间的共现关 系,该类算法无需引入外部语料进行训练,只 需对图进行随机游走即可实现词语排序和关键 词抽取。TextRank<sup>[8]</sup>借鉴了PageRank算法思想, 首次实现了对词图上的关键词评分并根据评分 结果完成关键词提取,成为了无监督关键词抽 取方法的典型代表。为了进一步改进TextRank 算法的提取效果,很多人对该算法进行了改进。 夏天<sup>[9]</sup>以TextRank为基础,引入词语位置信息 加权计算邻接词语的影响力转移矩阵,有效提 高了抽取效果;李航等<sup>[10]</sup>提出一种综合考虑词 性、词语位置信息、词语对文档集重要程度的 改进TextRank方法; 刘啸剑等<sup>[11]</sup>利用LDA构建 主题模型,计算词语相似度并以此相似度为权 重构建图的边,以短语作为图的节点,选择 top-k个词作为文章的关键词。

随着词向量技术的产生,越来越多的人开始研究将词向量与TextRank结合进行关键词提取。词向量技术可以挖掘出词与词之间的语义关系,然后将这种语义关系引入到TextRank算法的计算过程当中,从而解决TextRank只考虑词共现的缺陷。周锦章等<sup>[12]</sup>通过构建词向量,基于隐含主题分布思想和词汇的语义差异构建转移矩阵,将词向量与TextRank融合。

本文在已有研究基础之上以维基百科作为

外部知识库结合互联网获取的政策文本构建词向 量,根据《国务院公文主题词表》为词语初始化 权重,再利用词向量计算词语之间语义相似度, 结合政策文本位置权重共同构建TextRank转移矩 阵,最终选择K个关键词。

# 2 方法实现

# 2.1 基于fastText的词向量构建

词向量是指用来表示词语的向量,如比较 简单的One-hot representation。由Mikolov提出的 word2vec 是至今比较有名的词向量表示方式。 word2vec 的出现解决了传统词袋模型的缺点, 而word2vec 再生成词向量的时候会把每个词当 成原子,忽略词内部的形态特征。相对于 word2vec,fastText添加了subwords特性,使用 字符级的n-grams来表示单词,这样每个单词除 保留了本身外还被表示成多个n-grams。对于每 一个单词,fastText在拆分成n-grams表示的时 候,还在单词前后端加入"<"和">",用于区 分前缀和后缀,如单词hello采用3-grams可以表 示为: <he,hel,ell,llo,lo>和<hello>。在训 练模型的时候,当前词的词向量就是n-grams的 向量和:

$$\boldsymbol{v} = \sum_{g \in \Phi} z_g^{\mathrm{T}} \boldsymbol{v}_c \tag{1}$$

本文利用开源的fastText工具,将维基百科 和政策内容融合共同构建词向量。

#### 2.2 TextRank转移矩阵构建

利用TextRank 进行关键词抽取的思想比较简单:首先根据词共现关系构建无向带权图,然后利用PageRank循环迭代计算节点权值,排 序权值即可得到最终关键词。TextRank算法的核心计算公式如式(2)所示:

$$WS(V_{i}) = (1 - d) + d* \sum_{V_{j} \in \ln(V_{i})} \frac{w_{ji}}{\sum_{V_{k} \in Out(V_{j})} w_{jk}} WS(V_{j})$$

(2)

其中:  $WS(V_i)$ 表示节点 $V_i$ 的权值;  $\ln(V_i)$ 表示指 向 $V_i$ 的节点集合;  $Out(V_j)$ 表示节点 $V_j$ 指向的节点 集合;  $w_{ji}$ 表示两个节点之间边的权重;  $WS(V_j)$ 表 示节点 $V_j$ 的权值; d为阻尼系数, 一般取值0.85。 基于TextRank的关键词抽取步骤如下: (1) 文本预处理。包括按句子进行文本分 割、分词、词性标注、去停用词。

(2)构建词图。文本预处理之后的词语构成 节点集合,根据词语的共现关系构建边集。边 的构建采用滑动窗口机制,即当两个节点在长 度为K的窗口中共现,它们之间才会存在边。

(3)根据公式(2)迭代各节点的权重,直到结果收敛。

(4) 对结果进行排序,得到 top-k 关键词。

(5) 对所得到的关键词进行组合,如果组合 的词汇在政策全文中出现,则选择该组合词作 为一个关键短语。

本文通过引入词向量等方式对上述步骤进 行修改,从而达到面向政策文献的关键词抽取。

对于步骤(1),在进行分词的时候引入《国 务院公文主题词表》作为词库,同时剔除此表 中无区分度的词语,如:章程、条例、办法、 细则、规定、命令、决定、决议、公告、通告、 通知、通报、报告、请示、批复、函、会议纪 要、答复等。

当关键词出现在词表中的时候,在原有权 重的基础上再乘1.5。

对于步骤(2),在构建此图边集的时候综合 考虑词语的位置信息和词语之间的语义相似度, 共同构建转移矩阵。图中任意两个节点 $v_i$ 和 $v_j$ 之间的权重转移是通过边 $w_i$ 来完成的, $w_i$ 的构建如公式(3)所示:

$$w_{ii} = ft(i,j) + pos(i,j) \tag{3}$$

其中, ft(i, j)计算方式如下:

$$ft(i,j) = vsim(i,j)*coc(i,j)$$
(4)

*vsim*(*i*,*j*)表示*v<sub>i</sub>*和*v<sub>j</sub>*的的语义相似度,将两 者的词向量取出,按照余弦相似度进行计算, *coc*(*i*,*j*)表示二者共现次数。*pos*(*i*,*j*)表示政策文 本位置信息的重要影响程度,其计算方式如下 所示,*p<sub>i</sub>*表示词语出现的位置权重:

$$pos(i,j) = \frac{p_j}{\sum_{k \in \text{Out}(v_j)} p_k}$$
(5)

$$p_j = \begin{cases} 1.0, \ j \square 现在标题或段首 \\ 0.8, \ 其他位置 \end{cases}$$
 (6)

# 3 结果分析

本文以我国"双创"政策为例,验证提出 方法的有效性和实用性。首先从互联网相关网 站搜索以国务院以及各部委为发文主体,以 "双创"为内容的政策文件共计163篇(如表1所 示)。

表1 "双创"政策(部分)

编号	政策名称	发文时间	发文字号
1	国务院办公厅关于进一步支持大学生创新创业的指导意见	2021-10-12	国办发〔2021〕35号
2	国务院办公厅关于建设第三批大众创业万众创新示范基地的通知	2020-12-24	国办发〔2020〕51号
3	国务院办公厅关于支持多渠道灵活就业的意见	2020-07-31	国办发〔2020〕27号
4	国务院办公厅关于提升大众创业万众创新示范基地带动作用 进一步促改革稳就业强动能的实施意见	2020-07-30	国办发〔2020〕26号
5	国务院关于促进国家高新技术产业开发区高质量发展的若干意见	2020-07-17	国发〔2020〕7号
6	国务院办公厅关于应对新冠肺炎疫情影响强化稳就业举措的实施意见	2020-03-20	国办发〔2020〕6号
56	国务院关于进一步做好新形势下就业创业工作的意见	2015-05-01	国发〔2015〕23号
57	国务院办公厅关于创新投资管理方式建立协同监管机制的若干意见	2015-03-19	国办发〔2015〕12号
58	国务院关于取消和调整一批行政审批项目等事项的决定	2015-03-13	国发〔2015〕11号
59	国务院办公厅关于发展众创空间推进大众创新创业的指导意见	2015-03-11	国办发〔2015〕9号
161	关于支持和促进重点群体创业就业税收政策有关问题的补充通知	2015-01-27	财税〔2015〕18号
162	人力资源社会保障部办公厅关于做好留学回国人员自主创业工作有关问题的通知	2015-01-16	人社厅函〔2015〕19号
163	科技部关于进一步推动科技型中小企业创新发展的若干意见	2015-01-10	国科发高〔2015〕3号

根据本文提出的算法,对上述政策列表进 行关键词提取,剔除权重小于1.0的关键词,最 终得到关键词192个(见表2)。从关键词列表和 词云可以看出,国家以创新创业为核心制定多 项保障措施,如政府工作改革、提供支持政策、 加强科技支持、培养创新孵化企业、优化税收 政策、提供贷款资金支持等。

关键词	权重	关键词	权重	关键词	权重
创新创业	20.0	工作改革	12.22	科技 支持	9.43
就业创业	9.0	创业服务	9.0	孵化	9.0
发展改革	9.0	建设	8.28	创新 示范	6.84
税收优惠 政策	5.0	创业投资	4.26	贷款	4.10
创新发展	4.0	创业基地	4.0	众创 空间	4.0
					•••
技术人员 创新	1.0	投资引导	1.0	拉动 重点	1.0
农村产业 融合	1.0	就业服务 工作	1.0	奖励 股权	1.0

表 2 "双创"政策关键词列表(部分)

精确率方面,本文使用维基百科中文语料 和从互联网采集到的163篇"双创"政策作为词 向量训练文本(参数维度设置为100,窗口大小 为5),选择其中50条政策作为测试集,采用多 人交叉标注的方式为每篇政策选择10个关键词。 实验指标采用精准率P、召回率R和F1值进行 评测,其中N<sub>1</sub>表示人工标注合集,N<sub>2</sub>表示算法 抽取合集。

$$P = \frac{N_1 \cap N_2}{N_2} \tag{7}$$

$$R = \frac{N_1 \cap N_2}{N_1} \tag{8}$$

$$F1 = \frac{2*P*R}{P+R} \tag{9}$$

实验选择提取5、8、10个关键词对不同 的方法进行对比,结果如表3所示。具体包括 引入政策文本位置信息的TF-IDF方法(方法1)、 TextRank方法(方法2)、基于Word2Vector的关 键词抽取方法(方法3)。

表 3 不同提取方法的比较结果

		K=5			K=8			K=10	
	Р	R	F1	P	R	F1	P	R	F1
方法1	0.40	0.2	0.267	0.375	0.30	0.333	0.35	0.40	0.373
方法2	0.30	0.10	0.15	0.250	0.20	0.222	0.25	0.40	0.308
方法3	0.35	0.10	0.156	0.330	0.10	0.153	0.27	0.20	0.230
本文方法	0.45	0.25	0.321	0.420	0.37	0.393	0.40	0.43	0.414

为了能够更直观地查看对比结果,采用折 线图的方式将准确率P、召回率R和F1值进行 展示,如图1所示。



图 1 对比结果

从图1可以看出,实现方式最简单的是方法 1,其实验结果要优于方法2和方法3。对于方 法2,它的实现虽然不依赖于语料环境,但是在 没有任何改进的情况下也不能取得较好的结果。 在未加入外部语料和只考虑词语相似度的情况 下,方法3虽然引入了词向量技术,但是实验结 果却是最差。本文提出的方法在关键词抽取个 数不同的情况下相对其他几种算法效果都有明显的提升,准确率、召回率和F1值均高于其他三种方法,验证了提出方法的有效性和实用性。

# 4 结语

当对政策文本进行主题分析时,往往需要 提取政策文本的关键词汇,而由于其本身的特 殊性,并不会直接提供关键词字段,所以就要 对政策关键词进行提取。本文提出了一种将外 部知识库和政策库融合共同构建词向量,利用 《国务院公文主题词表》修改词语权重,综合考 虑位置信息和词语相似度构建TextRank转移矩 阵的政策文本关键词抽取方法。以"双创"政 策为例,提取政策关键词,结果表明本文提出 的方法具有较好的效果,可用于政策文本主题 分析,为政策研究人员提供辅助支持。

#### 参考文献:

- [1] 黄萃,任弢,张剑.政策文献量化研究:公共政策研究的新方向[J].公共管理学报,2015,12(2):
   129-137.
- [2] 吴宾,杨一民,娄成武.基于文献计量与内容分析的 政策文献综合量化研究:以中国海洋工程装备制造 业政策为例[J].情报杂志,2017,36(8):131-137.
- [3] 吴爱萍,董明,李华."互联网+"与"大众创业、万众 创新"政策结构分析:基于扎根理论和共词分析法
   [J].科技管理研究,2018,38(10):44-52.
- [4] 朱泽德,李森,张健,等.一种基于LDA模型的关键 词抽取方法[J].中南大学学报(自然科学版),

2015,46(6):2142-2148.

- [5] 张骁,周霞,王亚丹.中国科技服务业政策的量化与 演变:基于扎根理论和文本挖掘分析[J].中国科技 论坛,2018(6):6-13.
- [6] 张瑾.基于改进TF-IDF 算法的情报关键词提取方法[J].情报杂志,2014(4):153-155.
- [7] PAGE L, BRIN S, MOTWANI R, et al. The PageRank citation ranking: bringing order to the web[R]. Stanford InfoLab, 1999.
- [8] MIHALCEA R, TARAU P. Text rank: bringing order into texts [C] //Proceedings of Conference on Empirical Methods in Natural Language Processing. Barcelona, 2004:404-411.
- [9] 夏天. 词语位置加权 TextRank 的关键词抽取研究 [J]. 现代图书情报技术, 2013(9): 30-34.
- [10] 李航, 唐超兰, 杨贤, 等. 融合多特征的 TextRank 关 键词抽取方法[J]. 情报杂志, 2017, 36(8):183-187.
- [11] 刘啸剑,谢飞,吴信东.基于图和LDA主题模型的 关键词抽取算法[J].情报学报,2016,35(6): 664-672.
- [12] 周锦章,崔晓晖.基于词向量与TextRank的关键词 提取方法[J]. 计算机应用研究,2019,36(4): 1051-1054.

# Research on keyword extraction of policy using word vector and TextRank

# Li chen<sup>1</sup>, Zhao Yanqing<sup>1</sup>, Yu Junfeng<sup>1</sup>, Zhang Mingjun<sup>1</sup>, DMYTRO LANDE<sup>1,2</sup>

Information Research Institute, Qilu University of Technology (Shandong Academy of Sciences), Ji'nan 250014;
 Faculty of Information and Computer Science, National Technical University of Ukraine, Kyiv 03056)

Abstract: By analyzing policy texts, a key word extraction method based on machine learning is designed, which can extract key words or phrases automatically from policies. Firstly, it used fastText to construct word vectors for policies obtained from the Internet; then, integrated word vector and position information into TextRank transfer matrix; finally, selected the Top K words with the highest score as the policy keywords. The results show that the method proposed has good extraction effect and high practicability.

Keywords: policy; word vector; keywords extraction; TextRank

文章编号:1007-1423(2023)02-0073-05

DOI: 10.3969/j.issn.1007-1423.2023.02.010

# 基于 ResNet-18 网络的城市生活垃圾识别方法研究

金张根1,曹 杨2,于红绯2,孙才华1,刘 克2

(1. 辽宁石油化工大学人工智能与软件学院, 抚顺 113001; 2. 辽宁石油化工大学信息与控制工程学院, 抚顺 113001)

摘要:城市生活垃圾的自动化分类收集是解决城市生活垃圾精细化管理的有效途径。采用生活垃圾数据集 Kaggle 为图片样本,构建了基于 ResNet-18 卷积神经网络的生活垃圾识别模型,使用深度学习神经网络的方法对生活垃圾进行了分类识别研究。实验结果表明该方法识别的准确率可达 90.0% 以上,模型的平衡点在查全率和查准率分别为 0.8 和 0.89 处,为城市生活垃圾的自动化分类识别提供了依据。

关键词:城市生活垃圾;卷积神经网络;ResNet-18;图像分类

# 0 引言

我国人口基数大、地域发展不平衡,在一些较为发达的地区,多数垃圾会被无害化回收 处理。欠发达地区垃圾往往采用堆填埋的方法 处理,这导致大量土地被垃圾所占用,进而导 致土壤和地下水被污染<sup>[1]</sup>。垃圾分类作为一种 有效防止垃圾污染,实现垃圾资源化的手段, 既能实现资源的重复利用,又能给社会带来经 济效益。所以一个准确且高效的智能垃圾分类 系统对提高生态文明建设有着极其重要的作 用<sup>[2]</sup>。

当前对于垃圾分类识别的研究有许多种方法,但最主要的两种研究方法,一是基于手工特征的垃圾特征识别;二是基于神经网络的生活垃圾识别。其中,通过卷积神经网络来进行垃圾分类是较为高效的。卷积神经网络(CNN)在计算机视觉领域扮演着非常重要的角色,在图像分类、语义分割等方面取得了显著的成就<sup>[3]</sup>。Yann Lecun 在 1998 年提出来了基于 CNN

架构的结构简单的LeNet5模型<sup>[4]</sup>,该模型很好地 解释了CNN网络所包含的主要结构,如输入层、 卷积层、全连接层、输出层。2012年Krizhevsky 等<sup>[5]</sup>提出了AlexNet网络,网络中包含的ReLU 激活函数使模型收敛的速度更快,

并且提出了通过数据增强和Dropout来解决 训练过程中的过拟合问题。2014年,VGG模型 诞生,多个卷积层加一个最大池化层的模型使 得网络模型得到了更深的扩展<sup>[6]</sup>。2014年NiN 网络出现<sup>[7]</sup>,增加了模型的非线性拟合能力和 特征信息的提取能力。同年GoodLeNet模型被提 出,该模型主要提出了Inception结构的模块<sup>[8]</sup>, 提出了多尺度卷积变换的思想,减少了计算量, 并且提高了精准度。2015年He等<sup>[9]</sup>提出了残差 网络模型ResNet,在一定程度上解决了深度网 络在训练中准确率过低的问题,进一步提高了 特征提取的能力。

城市生活垃圾图像的信息丰富、目标尺寸 多变,同类目标之间和不同类目标之间尺寸差 别较大,并且目标形状不规则,这些因素影响

**收稿日期:** 2022-09-26 修稿日期: 2022-10-30

基金项目:辽宁省自然科学基金项目(2019-ZD-0052);辽宁省大学生创新创业训练项目(S202110148031)

作者简介:金张根(1999—),男,安徽安庆人,本科生,软件工程专业,研究方向为人工智能;曹杨(1982—),女,辽宁 抚顺人,博士,副教授,研究方向为机器博弈、计算机应用;于红绯(1982—),女,辽宁沈阳人,博士,副教授,研究方向为计 算机视觉、图像处理;孙才华(1999—),男,宁夏银川人,硕士研究生,研究方向为图像处理;刘克(2001—),男,安徽宿州 人,本科生,计算机科学与技术专业,研究方向为计算机应用

着不同目标的分类精度。

本文使用了深度学习的方法对城市生活垃圾进行识别研究。使用ResNet-18卷积神经网络搭建城市垃圾分类模型,采用经过预处理的Kaggle数据集作为识别对象,进行模型的优化和评估,最终探讨模型用于城市垃圾分类。

## 1 ResNet原理及网络结构

ResNet 残差网络是卷积神经网络中最为经 典的模型之一。该网络通过在输入层和输出层 之间添加跳跃链接层来有效解决卷积神经网络 随着层数增加出现的梯度消失或梯度爆炸问题, 能有效控制网络收敛飞速下降和模型的泛化能 力降低等一系列问题。ResNet 网络提升了训练 超深神经网络的速度,在深度增加的同时网络 性能也得以极大的提升。

ResNet 网络提出"恒等映射"原理的目的 是使网络在增加层数的同时训练误差 Loss 不会 增加。原始状态是网络输入x,得到输出为H(x),通过残差函数F(x) = H(x) - x,设置训练 目标将F(x)逼近于0,将其进行转化,得到 H(x) = F(x) + x,如图1所示。



#### 图 1 残差块结构原理

ResNet-8网络是最为经典的卷积神经网络, 网络结构如表1所示,假设输入图片的维度为 (224,224,3),经过第一层Conv\_1x输出图像为 (112,112,64),然后经过Conv\_2x的极大池化层 使得参数减少一半,此时图像的维度变为(56, 56,64),使用四层相同的卷积层处理图像大小 维度不变,依然为(56,56,64);到达第三层 Conv\_3x,卷积核的数量为128,最后输出的图 像大小为(28,28,128);到达第四层Conv\_4x, 卷积核数量为256,经过特征提取之后的图像大小维度变为(14,14,256);同样地,到达第五层 Conv\_5x时卷积核数量为512,经过处理后的图 像大小维度为(7,7,512),最后到达全连接层使 得图像的维度变为(1,1,512)。

表 1 ResNet18 网络结构表

名字	输出大小	网络结构/卷积核/步长
Conv_1x	112×112	7×7, 64, stride 2
		3×3, Maxpool, stride 2
		3×3, 64
Conv_2x	56×56	3×3, 64
		3×3, 64
		3×3, 64
		3×3, 128
G 2	28×28	3×3, 128
Conv_3x		3×3, 128
		3×3, 128
		3×3, 256
C 4	14×14	3×3, 256
Conv_4x	14×14	3×3, 256
		3×3, 256
		3×3, 512
Conv. 5v	7~7	3×3, 512
Conv_5x	1~1	3×3, 512
		3×3, 512
FC	1×1	Softmax, Avgpool
FLOPs		1.8×10 <sup>9</sup>

# 2 数据预处理和 ResNet 模型搭建

# 2.1 数据集获取

本文采用 2000 年发布的使用最为广泛的垃圾分类 Kaggle 数据集进行城市垃圾分类识别自动检测研究的数据集。Kaggle 数据集中主要包含城市中常见的生活垃圾,分别为可回收垃圾、 厨余垃圾、有害垃圾、其他垃圾四类。提取了 每种垃圾的一张图片作为范例,如图2所示。



#### 图 2 Kaggle 数据集样本图

实验使用监督学习的方法对神经网络模型

进行训练。为每类分类识别数据设定了对应的 标签,如表2所示。

表 2 数据集中不同垃圾对应的不同标签

垃圾	可回收垃圾	厨余垃圾	有害垃圾	其他垃圾
标签	0	1	2	3

# 2.2 ResNet-18模型搭建

本文采用 ResNet-18 的网络结构进行生活垃圾分类实验网络搭建,网络结构见图3。



图 3 ResNet-18卷积神经网络

此网络模型中包括卷积层和全连接层在内, 共有带有权重的18层,但不包括池化层和BN 层。首先, ResNet-18 网络中数据需要经过一个 7×7的卷积层, 然后经过四个残差块, 每个残差 块是由四个 3×3 的卷积层构成, 最后经过平均 池化层(Avg pool)后得到结果。

# 3 网络模型性能分析与评估

本研究硬件配置信息如下,实验电脑的基 本配置为: CPU为Intel(R) Core(TM) i7-9750H、 内存为16 GB DDR4、显卡为GTX 1660Ti、操作 系统为64 位 Windows 10,编程软件为PyCharm, 框架为Pytorch。

# 3.1 网络模型性能分析

选取 Kaggle 数据集作为模型的输入,该数 据集中包含 2940 张不同种类的生活垃圾图片, 图片的大小均为 256×256。将数据集按照 7:3 的比例分为训练集和验证集。通过多次实验确 定各个参数,其中 epoch 为 150,学习率 lr 为 0.001,训练的 Batch\_Size 为 32, Dropout 的断开 概率为 0.5。

图 4 是 ResNet-18 神经网络在训练集和验证 集上的准确率虽训练次数的变化。从图 4 可以看 出,训练过程中 ResNet-18 神经网络在城市垃圾 训练集和验证集上的准确率都是随着训练次数 的增加呈现先快速增加后缓慢增加的趋势,最 后收敛。多次实验表明,模型的准确率可达 90% 以上,且验证曲线基本上与训练的曲线一 致,说明优化参数之后的神经网络拟合能力较 好,识别各种生活垃圾的准确率达到了较为理 想的效果。



图 4 ResNet-18 在训练集和验证集上的准确率



图 5 ResNet-18 在训练集和验证集上的损失

图5为神经网络在训练集和测试集上的Loss 损失率随训练次数的变化。实验表明,优化参 数后的神经网络在训练集和验证集上的Loss损 失随着训练次数的增加先快速减小后缓慢减小, 最后接近收敛。训练集上的Loss损失降低至 0.25,验证集的Loss损失减低至0.43,验证曲线 基本上与训练的曲线趋势一致,拟合度较高, 说明优化参数之后的神经网络拟合能力亦有所 提升,识别垃圾分类的错误率明显降低。

# 3.2 模型评估

模型评估将预测正确的样本划分为一类, 预测错误的样本划归另外一类。即将多分类问 题转变成了二分类问题。根据预测结果可将所 有样本分为四大类,即真实情况的正例、真实 情况的反例、学习器预测出的正例、学习器预 测出的反例,如表3所示。通过以上四个参数并 不能直接判断模型的优劣,需进一步通过其组 合形成的评估指标中的查准率P和查全率R确定 模型的适用性。

表 3	分类结果矩阵	ł
-----	--------	---

古帝库阳	预测	结果
具头间仍	正例	反例
正例	TP(真正例)	FN(假反例)
反例	FP(假正例)	TN(真反例)

查准率 P(precision)表示预测正确的正样本 占所有预测为正例的比例,能反映完全正确的 样本的正确比例,是预测的底线。其计算公式 如式(1)所示:

$$P = \frac{TP}{TP + FP} \tag{1}$$

查全率*R*(recall)表示预测正确的正样本占 所有真实的结果为正例的比例,其计算公式如 式(2)所示:

$$R = \frac{TP}{TP + FN} \tag{2}$$

本文使用 ResNet-18 卷积神经网络识别城市 生活垃圾的查准率和查全率的 P-R关系如图 6 所 示, P-R 曲线反映出模型的平衡点出现在查全 率在 0.8、查准率在 0.89 的时候,此时模型的综 合性能最佳,同时具备较好的泛化能力,对种 类繁多、样式各异的城市生活垃圾有较好的识 别能力,结合本模型在验证集上的准确率达到 90.0% 以上,此模型适合于城市垃圾分类识别的 研究。



图 6 ResNet-18神经网络垃圾分类识别 P-R 曲线

# 4 结语

卷积神经网络可用于城市垃圾分类自动识 别并具有较高的准确度。本文将 Resnet-18 卷积 神经网络模型应用到城市生活垃圾的分类识别 研究中。实验结果表明,该模型在 Kaggle 生活 垃圾数据集上的训练准确率到达 90.0%, Loss 损 失可低至 0.43。查准率和查全率两项模型评价 的指标显示,模型的平衡点约在查全率为 0.8、 查准率为 0.89 的时刻,其查准率和查全率皆比 较理想。该模型的性能较好、泛化能力较强, 为进一步对各类城市垃圾的自动识别分类提供 了参考依据。

#### 参考文献:

- [1] 邓鹏.城市垃圾分类处理困境与建议[J].皮革制作 与环保科技,2022,3(10):143-145. DOI:10.20025/ j.cnki.CN10-1679.2022-10-49.
- [2] 战秋成,季龙华,赵际云,等.基于深度学习的智能 垃圾分类系统研究[J].机械工程师,2022(8):100-103.
- [3]] 季长清,高志勇,秦静,等.基于卷积神经网络的图 像分类算法综述[J]. 计算机应用,2022,42(4):1044 -1049.
- [4] WEI G, LI G, ZHAO J, et al. Development of a LeNet-5 gas identification CNN structure for electronic noses[J]. Sensors, 2019, 19(1):217.
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neu-

ral networks [J]. Communications of the ACM, 2017,60(6):84-90.

- [6] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large - scale image recognition [EB/OL]. arXiv Preprint arXiv:1409.1556,2014.
- [7] LIN M, CHEN Q, YAN S. Network in network [EB/OL]. arXiv Preprint arXiv:1312.4400,2013.
- [8] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015:1-9.
- [9] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016:770-778.

# Research on municipal solid waste identification method based on ResNet-18 network

Jin Zhanggen<sup>1</sup>, Cao Yang<sup>2</sup>, Yu Hongfei<sup>2</sup>, Sun Caihua<sup>1</sup>, Liu Ke<sup>2</sup>

(1. School of Artifical Intelligence and Software, Liaoning Shihua University, Fushun 113001;

2. School of Information and Control Engineering, Liaoning Shihua University, Fushun 113001)

Abstract: Automatic sorting and collecting of municipal solid waste is an effective way to solve the fine management of MSW. The domestic waste dataset Kaggle is used as the picture sample to construct the domestic waste identification model based on Resnet-18 convolutional neural network, and the method of deep learning neural network is used to carry out the classification and recognition research on domestic waste. The experimental results show that the recognition accuracy of this method can reach more than 90.0%, and the equilibrium point of the model is at 0.8 and 0.89 for recall and precision, respectively. This method provides a basis for the automatic classification and identification of MSW.

Keywords: municipal solid waste; convolutional neural network; ResNet-18; image classification

文章编号:1007-1423(2023)02-0078-04

DOI: 10.3969/j.issn.1007-1423.2023.02.011

# 基于光照估计滤波的太赫兹图像融合研究

张华忠\*,杜金花,潘曰凯

(中国民用航空飞行学院航空电子电气学院, 广汉 618300)

摘要:碳纤维增强复合材料广泛应用于航空航天等军民领域,在制备和使用过程中往往存在诸多缺陷, 太赫兹时域光谱成像技术为碳纤维增强复合材料无损检测提供了新方法。在使用太赫兹无损检测技术时, 可以通过在时域或频域中选择不同的参数来获取图像,同样的缺陷需要时域和频域不同的参数才可以有效 检测出来。采用基于光照估计滤波的算法将不同参数的太赫兹图像进行融合。首先,通过光照估计滤波分 辨初始图像序列中像素的曝光情况。然后隶属函数调用这些初始估计值。最终输出的融合图像是通过隶属 函数分配的权重在所有像素之间加权平均得到。实验结果表明,所提出的方法对图像边界信息和细节信息 都处理得比较好,适用于太赫兹多波段图像融合,具有更好的视觉效果。

关键词:图像融合;太赫兹多波段融合;光照估计滤波

# 0 引言

碳纤维增强复合材料作为高性能新型材料, 主要被用于航天、航空和国防等领域,在民用 飞机领域被广泛应用于飞机的尾翼、方向舵、 升降舵以及飞机机体等<sup>[1]</sup>。在飞机服役的过程 中,飞机蒙皮会出现裂纹、撞击和腐蚀等问题, 不仅降低飞机的使用寿命,而且还会对机上人 员的生命造成威胁。

人们期望借助无损探伤技术对碳纤维复合 材料的损伤类型和程度进行检测和分析<sup>[24]</sup>。因 具备对非金属材料穿透能力强、可进行非接触 检测等优点,太赫兹技术已经成为一种新型的 无损检测方法<sup>[5]</sup>。在太赫兹时域光谱成像过程 中对于不同的缺陷可以选取不同参数来成像<sup>[6]</sup>。

图像融合技术在图像处理领域发挥着重要 作用,将两张或多张图像包含的互补信息通过 某种图像融合方法进行组合可以得到具有更加 全面信息的图像<sup>[7]</sup>。图像融合能在像素级、特 征级和决策级三个层次上实现<sup>[8]</sup>。现在的研究 焦点是在像素级层面上,比如加权平均法、取 大法、区域方差等<sup>[9]</sup>。融合后的图像比未融合 图像对内容的描述更准确,可信性更高,可理 解性更好,更适合计算机检测、分类和识别等 处理<sup>[10]</sup>。

本文将基于光照滤波的图像融合算法首次 应用于太赫兹缺陷图像的融合,实现在一幅太 赫兹图像上展示出缺陷完整信息的目的。

# 1 图像融合算法实现

本文对民用飞机蒙皮内部缺陷的多个检测 参数获取的太赫兹图像进行融合研究,融合算 法的框图如图1所示。

**收稿日期:** 2022-09-18 修稿日期: 2022-10-17

基金项目:四川省科技厅苗子工程重点项目(2022JDRC0076):基于rPPG生理信号监测的民航人因安全研究及应用

作者简介:\*通信作者:张华忠(1989—),男,山东巨野人,硕士,副教授,硕导,研究方向为图像处理,E-mail:zhz\_ 233@yeah.net; 杜金花(1996—),女,甘肃庆阳人,硕士研究生,研究方向为图像处理; 潘曰凯(1999—),男,山东日照人, 硕士研究生,研究方向为计算机视觉



#### 图 1 图像融合算法框图

# 1.1 光照估计滤波

仅从一张图像上估计光照是一个已知的数 学不适定问题。使用一个鲁棒的递归包络<sup>[11]</sup>, 计算图像的包络,同时将计算的复杂性保持在 低水平。每次仅在一个维度上递归计算,并在 四个不同的空间方向上进行处理,以确保整个 图像中视觉信息的混合,基于递归滤波器生成 图像粗糙层和细节层的光照估计。

# 1.2 隶属度函数和像素权重

在对序列中的每个图像估计光照后,将隶 属度函数应用于每个光照图像,以确定每个像 素将参与最终融合图像的权重。

$$\boldsymbol{P}_1 = \boldsymbol{B}_1(\boldsymbol{E}_1) \tag{1}$$

$$\boldsymbol{P}_k = \boldsymbol{B}_2(\boldsymbol{E}_K) \tag{2}$$

$$\boldsymbol{P}_{N} = \boldsymbol{B}_{N}(\boldsymbol{E}_{N}) \,\forall k \in (2, N-1)$$
(3)

其中: *E*是光照估计图像, *P*是像素权重矩阵, *B*是隶属函数。

该算法这一阶段的主要目标是为那些曝光 良好的图像的像素分配更大的权重,为曝光不 足或者过度曝光的像素分配较低的权重<sup>[12]</sup>。

一个像素的曝光程度是指它与RGB通道的 距离<sup>[12]</sup>,因此曝光不足的像素,其值将在通道 下端附近,接近0;而过度曝光的像素,其值将 在通道上端附近,接近255;曝光良好的像素, 其值为128。因此,对于所有具有中间曝光的图 像选择三角形隶属度函数,其峰值为128,这意 味着对于这些图像,光照值在128左右的像素对 最终结果的影响最大。

曝光时间最长的图像的隶属函数不同,该

图像比序列图像中的任何其它图像都能更好地 捕捉场景的低亮度区域。低于128的像素值即使 被认为曝光不良,但它们在较长曝光图像中仍 然比序列中其它图像更好。因此,它的隶属函 数选择梯形。区间[0,128]上的所有像素值都被 指定为最大隶属值,即为1。类似地,具有最短 曝光的图像比序列中的任何其它图像更好地捕 获场景的高亮度区域,因此,虽然像素值高于 128不被认为曝光良好,但它们在最短曝光图像 中仍然比在任何其它序列中表现更好。因此, 隶属函数选择梯形,区间[128,255]内的所有像 素值被分配最大隶属值1。

# 1.3 混合函数和最终输出

所提出方法的混合函数由公式(4)描述。

$$W_{\text{out}} = \frac{\sum_{k=1}^{N} W_k \cdot P_k}{\sum_{k=1}^{N} P_k}$$
(4)

其中: N是场景中的像素总数,W是RGB颜色 空间的三个通道,公式(4)本质上是每次曝光的 R、G、B分量之间的加权平均,使用1.2节得出 的像素权重。结合隶属函数的形状,保证正确 曝光的像素对最终结果的参与度最大。

# 2 实验结果分析

实验环境为 Windows 10、Pycharm 2021,选 择三组不同频率下的太赫兹图像作为实验对象。 算法运行结果如图2~图4所示:本文算法保留 了图像边界信息, 增强了图像细节信息, 并有 效提高了图像的分辨率。图2为11.313 ps下的 时域图、0.897 THz下的频域图以及0.897 THz下 的反射图的融合结果,我们可以看到融合前的 时域图、频域图以及反射图的图像都表现出边 缘模糊的特点,不能很好地表示图像的显著特 征。经过本文算法处理后的图像对比度高,图 像边缘更锐利。图3为11.360 ps下的时域图、 0.842 THz下的频域图以及1.007 THz下的反射图 的融合结果,同样的,融合前的图像普遍存在 图像边缘模糊、细节丢失等问题。融合后的图 像从原图像中提取了图像的有用信息,目标和 背景更加清晰,局部特征也更加明显。图4为 11.367 ps下的时域图、0.916 THz下的频域图以











图 4 11.367 ps下的时域图、0.916 THz下的频域图以及0.916 THz下的反射图的融合结果

及0.916 THz下的反射图的融合结果,融合图像 保留了原图像的细节信息,且融合结果更适用 于人眼视觉观察。

# 3 结语

针对太赫兹图像的特点,本文将基于光照 估计滤波的图像融合算法首次应用于民用飞机 蒙皮内部缺陷检测的图像融合阶段,该算法可 以有效利用不同频段获取的太赫兹图像的信息, 使得最终的融合图像包含更多的缺陷信息,同 时改善了图像的分辨率。本文算法的有效性可 以在融合图像的视觉效果上得到验证。

#### 参考文献:

- [1] 李东升,杨应科,翟雨农,等.民用飞机复合材料机 身壁板装配协调形性调控技术研究[J].复合材料 学报,2022,39(9):4310-4318.
- [2] VLAHOVIĆ M M, JOVANIĆ, MARTINOVIC S, et al. Composites part B: engineering[J]. Composites Part B Engineering, 2009, 44(1): 458-466.
- [3] COSTA M F M. The importance of surface microtopographic inspection on quantitative NDT&E of composite coatings [C] // Proceedings of the 8th An-

nual International Conference on Composites Engineering, 2001:153-154.

- [4] NASSR A A, EL-DAKHAKHNI W W. Non-destructive evaluation of laminated composite plates using dielectrometry sensors [J]. Smart Materials & Structures, 2009, 18(5):055014.
- [5] 何普,赵纪元.采用太赫兹时域光谱技术的高精度 热障涂层测厚方法[J].西安交通大学学报,2022, 56(6):112-119.
- [6] 张瑾,王洁,沈雁,等.小波图像融合在太赫兹无损 检测中的应用[J].光谱学与光谱分析,2017,37 (12):3683-3688.
- [7] 胡钢,郑皎月,秦新强.结合局部邻域特性和C-BEMD的图像融合方法[J]. 计算机辅助设计与图 形学学报,2017,29(4):607-615.
- [8] YU L, CHEN X, RABAB K, et al. Medical image fu-

sion via convolutional sparsity based morphological component analysis [J]. IEEE Signal Processing Letters, 2019, 26(3):485-489.

- [9] 任全会,孙逸洁,黄灿胜.基于区域间相似度的红外 与可见光图像融合算法研究[J].红外技术,2022, 44(5):492-496.
- [10] 岳双燕.基于光度立体视觉的测量材质几何特征重 构[D].济南:山东大学,2016.
- [11] WU Y Q, CHEN S, LUO Z. Image fusion based on nonsubsampled contourlet transform [J]. Chinese Journal of Stereology and Image Analysis, 2008, 28 (4):1-4.
- [12] VONIKAKIS V, BOUZOS O, ANDREADIS I. Multi-exposure image fusion based on illumination estimation[C]//Proc. IASTED SIPA, 2011:135-142.

# Terahertz image fusion based on illumination estimation filtering

# Zhang Huazhong<sup>\*</sup>, Du Jinhua, Pan Yuekai

(Institute of Electronic and Electrical Engineering, Civil Aviation Flight University of China, Guanghan 618300)

Abstract: Carbon fiber reinforced composites are widely used in military and civil fields such as aerospace, and there are often many defects in the processes of preparation and application. Terahertz time-domain spectral imaging technology provides a new method for nondestructive testing of carbon fiber reinforced composites. Images can be obtained by selecting different parameters in time-domain or frequency-domain when terahertz nondestructive testing technology is applied. The same defect can be effectively detected only when different parameters in time-domain and frequency-domain are required. Terahertz images with different parameters are fused using an algorithm based on light estimation filtering. Firstly, the exposure of pixels in the initial image sequence is resolved by light estimation filtering. The membership function then calls these initial estimates. The final output fusion image is obtained by weighted averaging among all pixels through the weight assigned by the membership function. The experimental results show that the proposed method can deal with the edge information and details of the image well, and is suitable for terahertz multi band image fusion with better visual effect.

Keywords: image fusion; Terahertz Multiband Fusion; illumination estimation filtering

现代计算机 Modern Computer

# 实践与经验

文章编号:1007-1423(2023)02-0082-07

DOI: 10.3969/j.issn.1007-1423.2023.02.012

# 基于拟合的混响室莱斯*K*因子预测及信道重建

张雪莹,赵翔\*

(四川大学电子信息学院,成都 610000)

摘要:针对莱斯K因子的预测及给定莱斯K因子的信道重建问题,提出了使用少量不同配置下的K因子的测量值进行拟合,从而得到预测值或所需混响室配置的方法。在混响室中使用此方法对不同天线间距离处的K因子进行了拟合。结果表明,将Holloway表达式进行适当修正的函数拟合效果更优,且同时保留了Holloway表达式的物理意义,并对其进行了适当拓展。该方法对于如混响室品质因数等其它因素在不同取值处的K因子预测及信道重建问题也具有普适性。

关键词: 混响室; 莱斯K因子; 拟合; 拟合误差

# 0 引言

有别于有线通信系统,无线通信系统的信 号在开放的空间中传播,这种开放性导致无线 信道具有随机衰落及多径传输的特点。混响室 由于其四壁、天花板、地板及搅拌器可使信号 在腔室内进行多次反射而天然地具备模拟多径、 随机电磁环境的能力<sup>[1]</sup>。此外,相较于真实场 景,在混响室中进行无线信道的测量具有更高 的可靠性和可重复性。因此,混响室常被用于 对无线信道进行模拟及研究。例如,电磁兼容 (EMC)中使用混响室模拟了瑞利衰落信道,该 模型描述了发射机与接收机之间不存在直射信 号的无线信道。通过重新配置混响室可以模拟 更多随机电磁环境。

莱斯衰落信道是一种常见的无线传播环境, 它适用于收发信机之间存在直射信号的情况。 莱斯 K 因子是莱斯传播信道的关键参数之一, 它反映了接收信号的直射分量和散射分量的相 对强度。混响室中的莱斯 K 因子受收发天线间 的距离、混响室品质因数等因素影响,通过改 变混响室的配置,可以改变莱斯 K 因子的大小。 2006年,Holloway等<sup>[2]</sup>在混响室中对具有不同莱 斯*K*因子的衰落环境进行了模拟,同时对天线 和腔体特性对*K*因子的函数关系进行了研究; 2011年,Lemoine等<sup>[3]</sup>通过蒙特卡罗模拟和混响 室实验改进了*K*因子估计的解析表达式,并将 测量结果与Friis传输公式的计算结果进行了对 比;2018年,Marvin等<sup>[4]</sup>采用源搅拌和机械搅 拌的方式对混响室中的莱斯*K*因子进行了测量。

已有的文献多是对给定的不同配置情况下 的混响室莱斯*K*因子值的测量、计算及对比问 题进行的研究,而没有对莱斯*K*因子的预测及 如何由给定的莱斯*K*因子进行信道重建的问题 进行研究。当研究者需要获取不同配置下的混 响室莱斯*K*因子数值时,对莱斯*K*因子进行预测 可节省大量的研究时间及人力物力。此外,在 对具有特定莱斯*K*因子的无线信道进行研究时, 如何根据给定的莱斯*K*因子进行信道重建的问 题变得尤为重要。

针对*K*因子的快速预测及给定*K*因子的信道 重建问题,本文提出了基于拟合的混响室莱斯*K* 因子预测及信道重建的方法,以不同收发天线

**收稿日期:** 2022-11-23 修稿日期: 2023-01-09

作者简介:张雪莹(1996—),女,河北石家庄人,在读硕士研究生,研究方向为电磁兼容;\*通信作者:赵翔(1973—), 女,四川西昌人,教授,博士生导师,研究方向为电磁效应评估与电磁兼容、计算电磁学,E-mail:zhaoxiang@scu.edu.cn 间距离处的莱斯*K*因子值的预测及信道重建问题为例进行了研究。此方法不仅可以简单快速 地得到所需距离处的莱斯*K*因子,还可以为如 何配置给定莱斯*K*因子的混响室给出建议,为 后续的混响室研究提供了便利。该方法还可扩 展到对于如混响室品质因数等其它因素的不同 取值处的莱斯*K*因子预测及信道重建问题中, 具有普适性。此外,本文使用了几种不同的拟 合函数对实测数据进行了拟合,并从物理意义、 拟合误差等多个角度进行了讨论及比较。

# 1 无线信道莱斯 K 因子分析

2006年, Holloway 等<sup>[2]</sup>推导了混响室中的莱斯*K*因子表达式,如式(1)所示:

$$K = \frac{V}{\lambda Q} \frac{1}{r^2} D_r D_r \left( \overline{\rho}_r \cdot \overline{\rho}_r \right)^2 \tag{1}$$

式中: *V*为混响室体积;  $\lambda$ 为波长; *Q*为混响室 品质因数; *r*为收发天线间距离; *D*,为发射天线 的方向性; *D*,为接收天线的方向性, 文献[2]将 *D*, *D*, 假设为*D*, ( $\theta$ ,  $\varphi$ )、*D*, ( $\theta$ ,  $\varphi$ ),即远场方向 性。 $\bar{\rho}_{t}$ 和 $\bar{\rho}_{t}$ 为发射天线和接收天线的极化方向 单位矢量。此表达式体现了混响室特性、天线 方向性等因素与莱斯*K*因子之间的关系。通过 改变混响室的特性或腔内的天线配置,可以实 现具有所需莱斯*K*因子的信道。

在混响室中对收发天线间的散射参数进行 测量及后处理,可得到莱斯K因子的值。在*S*<sub>21</sub> 的散点图中可以直观地感受到直射分量及散射 分量对莱斯K因子的影响,如图1所示。当直射 分量不显著时,不同搅拌器位置的数据聚集在 一个以原点为中心的圆中,如图1(a)所示。当 直射分量明显占优时,数据点簇将远离原点, 如图1(b)所示。莱斯K因子的计算公式如式 (2)<sup>[2]</sup>所示。

$$K = \frac{d_R^2}{2\sigma_R^2} = \frac{\left(\left|\left\langle S_{21}\right\rangle\right|\right)^2}{\left\langle\left|S_{21} - \left\langle S_{21}\right\rangle\right|^2\right\rangle}$$
(2)

在 $S_{21}$ 的散点图中, $\sigma_R$ 是数据点簇的半径,  $d_R$ 是数据点簇的形心距原点的距离。〈•〉代表平 均值, $|\cdot|$ 代表复数的模。 $(|\langle S_{21} \rangle|)^2$ 是直射分量 的能量, $\langle |S_{21} - \langle S_{21} \rangle|^2 \rangle$ 是散射分量的能量。



(b)K = 2.9时的 $S_{21}$ 散点图

图 1 S<sub>21</sub>散点图

# 2 拟合原始数据获取

首先通过少量测量获取莱斯K因子拟合所 需的原始数据。混响室由腔室和搅拌器组成。 本文所使用混响室的尺寸为3.97m×2.8m× 1.91m, 混响室中使用非规则铝箔搅拌器。腔室 中放入3块吸波材料。发射及接收天线均使用喇 叭天线,将两个喇叭天线连接到矢量网络分析 仪,测量4.5~8.5 GHz范围内的S参数,选取 100个搅拌桨位置,使用式(2)对莱斯K因子进 行计算。实验过程中,两天线始终保持正对, 极化方向相同,实验系统示意图及照片分别如 图2、图3所示。



图 2 实验系统示意图



图 3 实验系统照片

改变收发天线间距离,对距离为0.2 m、0.4 m、 …、2.0 m时的莱斯*K*因子进行测量计算。测量结 果如图4所示,其中,图4(a)、图4(b)分别为 较短距离(0.2~1.0 m)及较长距离(1.2~2.0 m) 处的*K*因子随频率变化的曲线。



(b) 1.2~2.0 m处的莱斯K因子



由图4可以看出,随着收发天线间距离的增 大,莱斯K因子的值逐渐减小。当频率从4.5 GHz 上升到9.5 GHz时,莱斯K因子的数值也随之上 升。与文献[2]的结论一致。

# 3 莱斯 K 因子拟合及预测

# 3.1 莱斯 K因子拟合

为了减小测量不确定度对结果的影响,本 文在机械搅拌的同时加入了频率搅拌,研究表 明,当扫频带宽∆f与中心频率f的比值等于3% 时,使用少量的搅拌器位置和少量的独立频率 就可获得较高的估计精度[3]。因此,本文使用 △f/f=3%的扫频带宽,100个搅拌器位置对莱 斯K因子进行测量计算。将5 GHz、6 GHz、7 GHz、 8 GHz 频率处的莱斯 K 因子在 0.2~2.0 m 范围内 的测量结果使用 Matlab 进行拟合。文献[2]将  $D_{t}$ 、 $D_{t}$ 假设为 $D_{t}(\theta, \varphi)$ 、 $D_{t}(\theta, \varphi)$ ,即远场方向 性,因此使用式(1),即K =  $1/a_1r^2(r)$ 为收发天 线间的距离, a, 为参数)对天线近场处的莱斯 K因子进行拟合可能存在较大的误差。鉴于此, 本文除 $K = 1/a_1 r^2$ 外,还使用了对表达式(1)进 行适当修正的函数进行拟合。拟合函数如表1所 示。其中,函数2、函数3对函数1(未修正函 数)的分子进行了修正。在近场处,天线的方向 性随距离变化,且本文中两天线始终保持正对, 即 $\theta$ 、 $\varphi$ 保持不变,因此假设收发天线方向性分 别为 $D_t(r)$ 、 $D_t(r)$ 。由于天线近场方向性的表达 式较为繁琐,因此,在函数2、函数3中分别将 天线方向性简单地设为距离的一次及二次多项 式。此外,函数4对函数1的分母进行了修正, 在分母中增加了距离的一次项及常数项。表2以 (a1, a2, a3)的形式展示了使用 Matlab 拟合后各函 数的参数取值,其中, a1、a2、、a3为函数参数。 四个函数在0.2~2.0 m范围内不同频点处的拟合 结果如图5所示。

表 1 拟合函数

函数标号	函数形式
函数1	$K = \frac{1}{a_1 r^2}$
函数2	$K = \frac{a_2r + a_1}{r^2}$
函数3	$K = \frac{a_3 r^2 + a_2 r + a_1}{r^2}$
函数4	$K = \frac{1}{a_1 r^2 + a_2 r + a_3}$

观察图5可以发现,在小于0.3m时四个拟

表 2 各函数在不同频率点处的参数取值

函数标号	5 GHz	6 GHz	7 GHz	8 GHz
函数1	(0.3530)	(0.2681)	(0.2302)	(0.1966)
函数2	(1.5723, 5.6936)	(1.7982, 8.7249)	(2.9638, 6.2326)	(2.4259, 12.0242)
函数3	(1.5548, 5.7973, -0.0959)	(1.7574, 8.9673, -0.2241)	(2.7789, 7.3304, -1.0149)	(1.4019, 18.1053, -5.6215)
函数4	(0.0176, 0.1284, -0.0117)	(0.0131, 0.0877, -0.0068)	(0.0537, 0.0630, -0.0052)	(0.0545, 0.0236, 0.0015)



图 5 莱斯 K因子实测及拟合结果对比

合函数之间的差距较小,四条拟合曲线几乎重叠,当距离逐渐增大时,四条曲线间的差距逐渐增大, 四条曲线间的差距逐渐增大。与函数1相比,函数2、函数4与实测点间的差距较小,初步判断函数2和函数4的拟合效果较好。

# 3.2 莱斯 K因子预测

本文对拟合结果使用均方根误差进行检验<sup>[7]</sup>,其公式如式(3)所示。*E*表示均方根误差, $(x_i, y_i)$ 为实测数据,f(x)为拟合函数, $\delta_i$ 为拟合函数f(x)在 $x_i$ 处的误差,n为实测数据个数。将拟合函数在 $0.3 \text{ m}, 0.5 \text{ m}, \cdots, 1.5 \text{ m}$ 处的预测结果与测量结果进行对比,进一步检验拟合效果。

拟合误差和预测误差分别见表3和表4,图6直 观反映了四个函数的预测效果。

$$E = \left(\frac{1}{n}\sum_{i=1}^{n} |\delta_{i}|^{2}\right)^{\frac{1}{2}} = \left(\frac{1}{n}\sum_{i=1}^{n} |f(x_{i}) - y_{i}|^{2}\right)^{\frac{1}{2}}$$
(3)

表 3 拟合误差

函数标号	5 GHz	6 GHz	7 GHz	8 GHz
函数1	4.4380	6.6034	4.8263	9.6289
函数2	1.2284	0.9470	1.2248	3.4061
函数3	1.2279	0.9438	1.1726	2.7858
函数4	1.1402	0.9286	1.2293	2.3546



图 6 莱斯 K 因子实测及预测结果对比

表 4 预测误差

函数标号	$5~\mathrm{GHz}$	6 GHz	7 GHz	8 GHz
函数1	4.4489	7.0213	5.1216	12.3523
函数2	1.2256	1.8945	1.5496	4.2511
函数3	1.2463	1.8551	1.6009	2.6854
函数4	1.2289	2.0542	1.6011	1.4214

图6中,预测点的实测结果皆落在拟合曲线 附近,初步判断拟合曲线预测较准。

观察表3和表4可以发现,函数1在各频点处的拟合误差及预测误差均较大,函数2~函数4的拟合误差及预测误差均较小。其中,在5GHz、6GHz及8GHz处函数4的拟合误差最小,7GHz处函数3的拟合误差最小,该三个函数在同一频点处拟合的准确度的差距很小。在表4中,5GHz、7GHz处函数2的预测误差最小,6GHz处函数3的预测误差最小,8GHz处函数4的预测效果最好,且除8GHz外其它频点处三个函数的预测准确度差距较小。因此,函数2~函数4对莱斯

K因子关于距离r的拟合较函数1更为准确。

从函数形式上看,函数1、函数2的形式较 简单,函数3、函数4的形式较复杂。从物理意 义角度分析,函数4对函数1的分母进行了适当 修正,在分母中增加了距离的一次项及常数项, 使莱斯K因子随距离的二次项及一次项变化, 函数2、函数3对函数1的分子进行了修正,将 表达式(1)中的天线方向性简单地设为距离的一 次及二次多项式,它们不仅保留了表达式(1)的 物理意义,还将其进行了拓展,使其在近场及 远场情况下均适用。

因此,使用对Holloway表达式进行适当修 正后的函数进行拟合更为合适,根据拟合精度、 函数形式或物理意义等方面的需求可选择相应 的函数进行拟合,通过此拟合曲线,不仅可以 预测本混响室在0.2~2.0 m之间任意距离处的莱 斯*K*因子,还可根据所需的莱斯*K*因子确定其 所对应的收发天线间距离,从而对具有此莱斯 K因子的无线信道在混响室中进行重建。

# 4 结语

本文针对混响室莱斯K因子预测及信道重 建的问题提出了将少量不同配置下的莱斯K因 子测量数据进行拟合的方法,并以不同收发天 线间距离处的莱斯K因子值的预测及信道重建 问题为例进行了研究,对不同函数的拟合效果 进行了比较。拟合结果表明,将 Holloway 表达 式进行适当修正后的函数拟合效果更优,并同 时保留了 Holloway 表达式的物理意义,对其进 行了适当拓展。通过该方法不仅可预测不同距 离处的莱斯K因子数值,还为如何配置混响室 使其具有给定的莱斯K因子提供了指引。此方 法对于如混响室品质因数等其它因素在不同取 值处的混响室莱斯K因子预测及信道重建问题 也具有普适性。此外,当研究者需要获取大量 不同配置下的混响室莱斯K因子数值时,采用 本方法还可减少实验次数,节省人力物力,降 低测量工作量。

#### 参考文献:

- [1] 梁小亮. 电磁兼容混响室发展及应用综述[J]. 民用 飞机设计与研究,2010(2):60-61,73.
- [2] HOLLOWAY C L, HILL D A, LADBURY J M, et al. On the use of reverberation chambers to simulate a Rician radio environment for the testing of wireless devices [J]. IEEE Transactions on Antennas and Propagation, 2006, 54(11): 3167-3177.
- [3] LEMOINE C, AMADOR E, BESNIER P. On the K -factor estimation for Rician channel simulated in reverberation chamber [J]. IEEE Transactions on Antennas and Propagation, 2011, 59(3):1003-1012.
- [4] MARVIN A C, FRANKS L, FLINTOFT I, et al. Design and evaluation of a broadband source stirring antenna for use in a reverberation chamber[C]//Proceedings of the 2018 International Symposium on Electromagnetic Compatibility. Amsterdam, 2018: 666-670.
- [5] LEO A D, CERRI G, RUSSO P, et al. Experimental validation of an analytical model for the design of source - stirred chambers [J]. IEEE Transactions on Electromagnetic Compatibility, 2018, 60 (2) : 540 -543.

- [6] 李欢,刘晓东,赵翔,等.混响室随机多径衰落电磁 环境的实验研究[J].无线电工程,2019,49(8):715-719.
- [7] 任玉杰. 数值分析及其 Matlab 实现[M]. 北京:高等 教育出版社,2007.
- [8] SORRENTINO A, CAPPA S, GIFUNI A, et al. The rice K-factor distribution within a mode-stirred reverberating chamber [C] // Proceedings of the 2017 International Symposium on Electromagnetic Compatibility—EMC Europe. Angers, 2017:1-5.
- [9] SANCHEZ-HEREDIA J D, VALENZUELA-VAL-DES J F, MARTINEZ - GONZALEZ A M, et al. Emulation of MIMO Rician - fading environments with mode-stirred reverberation chambers [J]. IEEE Transactions on Antennas and Propagation, 2011, 59 (2):654-660.
- [10] SORRENTINO A, FERRARA G, CAPPA S. Experimental rice factor distribution within a reverberating chamber [J]. IEEE Antennas and Wireless Propagation Letters, 2017, 16:2911-2914.
- [11] 崔耀中,魏光辉,范丽思,等. 混响室关键配置对场 分布影响分析[J]. 电波科学学报,2011,26(6):1107-1112,1227.
- [12] 赵翔,茹梦圆,闫丽萍,等.电磁混响室搅拌方式研 究综述[J].强激光与粒子束,2020,32(6):41-50.
- [13] 刘逸飞,陈永光,程二威.基于平面波叠加的混响室场环境模拟与测试仿真[J].高电压技术,2017,43
   (9):3029-3035.
- [14] 王树峤,王庆国,贾锐.加载物对混响室参数的影响[J].科学技术与工程,2013,13(19):5621-5626.
- [15] LEMOINE C, AMADOR E, BESNIER P, et al. Antenna directivity measurement in reverberation chamber from Rician K-factor estimation [J]. IEEE Transactions on Antennas and Propagation, 2013, 61(10): 5307-5310.
- [16] ANDRIES M I, BESNIER P, LEMOINE C. Estimating K-factor and time spread parameters from a transient response of a pulse modulated sine wave in reverberation chamber[J]. IEEE Transactions on Antennas and Propagation, 2012, 61(1); 380-389.
- [17] 周鑫. 室内无线信道测量及基于混响室的信道模拟 研究[D]. 北京:北京交通大学,2018.
- [18] 苏政铭,刘强,赵远,等.基于柔性屏蔽材料混响室 的设计与应用[J].强激光与粒子束,2018,30(7): 91-96.

(下转第94页)

文章编号:1007-1423(2023)02-0088-07

DOI: 10.3969/j.issn.1007-1423.2023.02.013

# Web应用中间件性能测试系统设计与验证

刘 维<sup>1,2</sup>, 何冬辉<sup>1\*</sup>, 杨攀飞<sup>1,2</sup>

(1.工业和信息化部电子第五研究所,广州 511300;2.基础软硬件性能与可靠性测评工业和信息化部重点实验室,广州 511300)

摘要:作为三大基础软件之一,中间件是信息系统项目不可缺少的核心组件。近年来,国产中间件产 品虽有所发展,但在产品性能和稳定性方面与国际主流产品相比差距依然较大,软件质量依然是影响产品 市场接受度的重要因素。针对当前我国在中间件测评体系和测评工具方面还没有成熟的、标准的技术和产 品成果的现状,提出了一种基于汽车保险经纪基准模型的Web应用中间件测试系统,并在某国产品牌Web 应用中间件上进行了应用验证,测试结果表明该中间件通过所有测试项。

关键词: 国产Web应用服务器; 汽车保险经纪基准模型; 性能测试系统

# 0 引言

中间件是连接软件组件或企业应用程序的 软件<sup>[1]</sup>,其主要作用是屏蔽底层系统和通信的 异构性,进而支撑应用实现稳定、可靠和高并 发运行,并简化应用的开发流程<sup>[2]</sup>。它开始于 操作系统之上的一些附加功能,以促进复杂应 用程序的开发,然后发展了数据集成功能,后 来成为网络应用程序的促进者,最终成为每个 分布式环境、应用程序、系统和平台的重要组 成部分。到目前为止,任何类型的分布式系统 或应用程序都能找到中间件或一些涉及的中间 件功能。通常,它支持复杂的分布式业务软件 应用程序<sup>[1]</sup>。

作为三大基础软件之一,中间件是信息系 统建设项目不可缺少的核心组件,与操作系统、 数据库具备同样的战略高度。中间件实现的范 围是多种多样的,从面向通信的中间件到面向 组件的应用服务器<sup>[3]</sup>。目前国际通用的中间件 技术标准主要有由 SUN 公司提出的 Java EE、微 软的 COM+/.NET 框架以及公共对象请求代理体 系结构 (common object request broker architecture, CORBA)/CORBA组件模型(CCM)<sup>[4]</sup>。在政 策的指引下,国内涌现出了一批国产化中间件产 品,并广泛应用在各个行业与领域,现阶段的国 产中间件的核心诉求主要集中在应用中间件。

目前在中国市场,主要的中间件厂商(组 织)有8家,其中3家来自美国,5家为国内厂 商,国内厂商因受到政策驱动,市场保持稳定 发展态势<sup>[5]</sup>。在8家中间件厂商的主打中间件产 品中,7家为商业产品,1家为开源产品,8家 厂商均通过了Jakarta EE 8规范兼容测试认证。

获得系统性能特征的常用方法是通过基准 测试<sup>[3]</sup>,为保证国产Web应用中间件满足其服 务质量(QoS)要求,须使用基准测试来测量和验 证它们的性能。面对国产中间件无测评体系、 无配套测评工具的现状,通过对中间件测评技 术的研究,本文设计了一种Web应用中间件性 能测试系统,使用汽车保险经纪基准模型对 Web应用中间件进行性能评估,对通过了Java EE 8、Java EE 7、Java EE 6 国际标准兼容认证

**收稿日期**: 2022-10-17 修稿日期: 2022-10-31

基金项目:工业和信息化部电子第五研究所专项基金(20Z43)

作者简介:刘维(1986—),女,湖南岳阳人,硕士,高级工程师,研究方向为基础软硬件测评技术研究;\*通信作者: 何冬辉(1993—),男,湖南永州人,硕士,助理工程师,研究方向为基础软件性能、可靠性、安全性测评技术研究,E-mail: hedonghui@ceprei.com;杨攀飞(1989—),男,河南平顶山人,硕士,工程师,研究方向为基础软硬件测评、软件系统工程保障 的某国产品牌Web应用服务器软件进行评测。

# 1 系统设计

# 1.1 汽车保险经纪基准模型

汽车保险经纪基准模型旨在测量与Java EE 7.0 Web Profile 或更高版本规范兼容的应用服务 器的性能。它为 Web 应用程序的用户提供了一 个标准,用于比较从 Java EE Web Profile 应用服 务器、Java 运行时环境、网络和操作系统到数 据库服务器和存储的所有级别的应用程序栈 (application stack)的软件和硬件性能<sup>[6]</sup>。它由一 个保险应用程序组成,包含三种不同的服务: 车辆服务、保险供应商服务和保险服务,该应 用程序使用JSF和REST函数来驱动应用服务器 上的负载。

# 1.2 Web应用中间件测试系统的实现

整个测试系统可大致分为Driver, Application Server, Database Server 三个部分<sup>[7]</sup>,结构如 图1所示。

Driver的压力部分:首先Driver引用Faban 测试工具进行部署和测试,其工作原理是由主

节点创立多个从节点,之后用大量从节点来进 行压力负载的测试。基准测试系统使用Faban来 安排和自动化基准测试运行。Faban还提供了一 个嵌入式图形工具,可以绘制和比较运行结果。 该工具作为托管在Tomcat服务器上的Web应用 程序运行,它提供Web用户界面,并充当汽车 保险经纪基准测试的驱动程序组件的容器。线 束提供的Web用户界面用于调度运行,管理运 行,查看运行日志以及运行完成后查看结果(包 括详细图表)<sup>[7]</sup>。

Application Server部分:其主要运行着车辆 保险模型,该模型由三部分服务组成,分别是: 车辆服务、保险供应商服务和保险服务。该模 型支持集群化部署。

Database Server部分:提供数据服务,并且 能够及时响应,支持集群化部署。

通过 gradle 脚本程序获取到预先写好的 Application 的接口测试模型数据,之后通过创建主从线程来预热测试程序,预热之后按照接口测试模型数据针对性地对承载应用服务器 Application Server 的中间件进行测试。



图 1 Web应用中间件性能测试系统结构图

Web应用中间件性能测试系统的功能特点 见表1。

#### 表 1 Web 应用中间件性能测试系统的功能特点

功能项	功能特点
环境监测 及兼容	兼容主流硬件平台(Mips/LoongArch、x86、ARM 等)、操作系统(UOS、KylinOS、Centos等),支 持对测试环境软硬件关心信息的识别及提取
测试驱动 器	支持多线程框架,保证多个线程并发访问和取样 能力;提供对测试执行的访问结果回传测试管理 端的能力;支持多种协议的模拟访问
基准模型	依据相关基准规范,提供半自动化方式生成原数 据、创建表结构、部署基准应用程序,并提供运 行引导程序及操作指引
测试配置	提供参数的界面化设置,如:提供测试连接配 置、并发数、测试持续时间、预热时间等
测试结果 分析	提供测试结果的统计与图形化汇总展示,如:并 发数、吞吐量、事务平均响应时间及通过率等; 提供测试结果的图形化或图表显示,如:PDF/ JSON/TXT

# 2 应用验证

# 2.1 运行环境

测试系统软、硬件环境如表2所示。汽车保 险经纪基准模型测试的吞吐量由保险驱动程序 的活动驱动。注入率(injection rate, IR)衡量在 基准运行期间将加载多少数据,驱动器的吞吐 量与所选的注入率直接相关。为了提高吞吐量, 需要增加注入率。较高的注入率等同于更多的 客户端负载,因此相应地创建了更多的数据, 本文分别在注入率为500、1000、1500和2000 的条件下进行了测试,设置预热时长为600 s, 测试时长为3600 s,结束时长为300 s。

#### 表 2 测试系统运行环境

庙仲立持	CPU: kunpeng 990
硬针环境	内存: 96 GB
	中间件:某国产品牌Web应用中间件
	操作系统:银河麒麟V10
软件环境	数据库: PostgreSQL 11
	项目构建工具: Gradle 4.6

# 2.2 测试结果分析

本系统对Web应用中间件测试项分为四个

部分,分别为组合操作(Operation Mix)、响应时 长(Response Times(seconds))、逻辑执行时长 (Cycle/Think Times(seconds))及其他项(Miscellaneous Statistics),业务交易类型如表3所示。由 保险驱动程序运行的业务交易是保险服务的一系 列操作,业务交易序列由驱动程序根据表4中所 示的组合选择。驱动程序将85%的负载发送到 JAX-RS API,并将15%的负载发送到JSF生成的 HTML页面。表4中的每个序列都具有REST和 JSF变体。驱动负载的客户端模拟Web用户与保 险服务的应用程序进行交互,负载按照15%的 JSF请求和85%的REST请求的比例进行分配。

#### 表 3 应用中间件测试的业务交易操作

类型	业务交易	中文名
	Accept Quote	接受报价
	Accept Quote + WebSocket	_
	Add Vehicle	添加车辆
	Delete Vehicle	删除车辆
	Login	登录
	Logout	退出
ICE/DECT	Register	注册
JSF/RESI	Register Invalid	无效注册
	Unregister	注销
	Update User	更新用户数据
	View Insurance	查看保险
	View Quote	查看报价
	View User	查看用户信息
	View Vehicle	杳看车辆信息

#### 表 4 业务交易组合要求

业务交易顺序	百分比组合
无效注册	2
注册	8
登录/注销	5
登录/查看用户/退出	5
登录/查看用户/更新用户数据/退出	10
登录/增加车辆/查看报价/退出	25
登录/删除车辆/退出	10
登录/接受报价/查看车辆信息/退出	20
登录/Accept Quote + WebSocket /查看车辆信 息/退出	5
登录/查看车辆信息/查看保险/退出	10

# 2.2.1 组合操作

组合操作测试结果包括业务成功总数、失

败总数、实际组合、目标组合四项,对于每种 类型的业务交易操作,基准中实现的实际组合 须在目标组合的4%范围内(±2%),驱动程序检 查并报告是否满足组合混合要求。图2显示了 500、1000、1500和2000四种不同注入率 (txrate)下业务操作的成功总数,图3显示了实 际组合和目标组合之间的偏差在规定的范围内, 所有业务测试项通过。



图 2 不同注入率下业务的成功总数





## 2.2.2 响应时长

驱动程序测量并记录业务交易序列中每个 不同操作的响应时间,仅包括在测量间隔中成 功完成业务交易序列的操作。业务交易的每个 操作中至少有90%的响应时间(90%th)须小于 2 s(即90%目标响应时间为2 s)。每个业务交易 操作的平均响应时间不得超过记录的90%响应时 间的0.1 s,此要求可确保所有用户都能看到合理 的响应时间。例如,如果注册事务的90%响应时 间是1 s,那么平均值不能大于1.1 s。驱动程序检 查并报告是否满足响应时间要求。 该项测试,测试结果包括平均响应时长、 最大响应时长、标准差、90%实际响应时间、 90%目标响应时间五项,若每项业务平均响应 时间和90%实际响应时间符合以上要求,则该 测试项通过,反之则未通过。

图4表明不同注入率,业务交易的每个操作 中90%实际响应时间(90%th)均小于90%目标响 应时间(2s),平均响应时间(avg)符合规定,所有 测试项通过,通过计算验证也可证明以上结论。



图 4 业务交易操作的响应时间

### 2.2.3 逻辑执行时长

该部分测试结果分为目标平均(targeted avg)、实际平均(actual avg)、逻辑执行时长最 小值(min)和逻辑执行时长最大值(max)四项, 每一个业务交易操作的测试均生成上述四项结 果,当实际平均不超过目标平均时,该业务交 易操作项测试通过。测试结果表明在各注入率 下实际平均执行时长(actual avg)皆不高于目标 执行时长(targeted avg),该项测试通过,测试结 果如图5及图6所示。



图 5 逻辑执行时长的目标平均和实际平均



图6 逻辑执行时长的目标平均及实际平均之差

# 2.2.4 其他项

该部分测试项设测量基准设置,包括使用 安全连接、有效的运行属性设置;初步审计项, 该项是注入数据的初步数据量审计,包括车辆 描述、车辆、车辆保险、投保人、保险政策; 原子性测试,测试内容包括原子性测试一及原 子性测试二;持久性和正确性测试,包括JPA 缓存透写测试、幻读删除测试;最终审计项测 试,它对程序测试之后的数据审计;业务审计 项测试。具体描述见表5。

表 5 其他项测试

描述	中文名称	目的
Benchmark setup	测量基准设置	确保有效的运行属性和安全连接
Using secure connections	使用安全连接	
Valid run properties settings	有效的运行属性设置	
Initial auditing counts	初步审计项	确保所有数据库表都已加载了预期的数据量
Vehicle descriptions	车辆描述	
Vehicles	车辆	
Vehicle insurances	车辆保险	
Policy holders	投保人	
Policy coverages	保险政策	
Atomicity Tests	原子性测试	确保文件规定的原子性要求
Atomicity Test One	原子性测试一	
Atomicity Test Two	原子性测试二	
Persistence and Correctness Tests	持久性和正确性测试	确保实体被正确储存
JPA Cache Write Through Test	JPA缓存透写测试	
Phantom Delete Test	幻读删除测试	
Final auditing counts	最终审计项	确保数据库表包含驱动程序执行的所有新实 体、更新和删除
Vehicle descriptions	车辆描述	
Vehicles	车辆	
Vehicle insurances	车辆保险	
Policy holders	投保人	
Policy coverages	保险政策	
WebSocket Messages	WebSocket 消息成功率	
Operations audit	业务审计	确保只有少量请求失败
Accept Quote success rate	接受报价成功率	
Accept Quote + WebSocket success rate	_	
Add Vehicle success rate	添加车辆成功率	
Delete Vehicle success rate	删除车辆成功率	
Login success rate	登录成功率	
Logout success rate	退出成功率	

描述	中文名称	目的
Register rate success rate	注册成功率	
Register Invalid success rate	无效注册成功率	
Unregister success rate	未注册成功率	
Update User success rate	更新用户数据成功率	
View Insurance success rate	查看保险成功率	
View Quote success rate	查看报价成功率	
View User success rate	查看用户信息成功率	
View Vehicle success rate	查看车辆信息成功率	

续表5

在基准测试期间,业务交易序列的每个操 作有99%或以上须在稳定状态下取得成功,这 允许最多1%的操作类型失败。此外,数据库表 中的行数须不超过预期大小±1%的偏差。驱动 程序检查并报告是否满足这些要求。测试结果 表明,实际结果(results)符合目标结果(targeted results),被测Web应用中间件使用https安全连 接方式,与目标结果一致,有效的运行属性设 置符合目标结果,原子性测试通过。在误差允 许范围内,初步审计、最终审计及业务审计成 功率被测结果符合目标结果要求。图7在横轴上 三组不连续线段分别表示初步审计项、最终审 计项和业务审计的成功数。



图 7 初步审计、最终审计及业务审计成功数

# 3 结语

当前,我国信创产业快速发展,但在中间 件测评体系和测评工具方面还没有成熟的、标 准的技术和产品成果,本文设计了一个基于汽 车保险经纪基准模型的Web应用中间件性能测试系统,并在某国产品牌Web应用中间件上进行了应用验证,探索了在四组不同注入率下应用服务器的系统性能。实验结果表明,该测试系统能够实现对国产Web应用中间件性能的评估,在一定程度上有效牵引了国产Web应用中间件产品质量提升。

#### 参考文献:

- [1] Oracle. Enterprise manager getting started with oracle fusion middleware management [EB/OL]. (2011-04)
   [2022-08-07] https://docs.oracle.com/cd/E11857\_01/install.111/e17558.pdf.
- [2] 李利军. TongWeb 中间件实用教程[M]. 北京:人 民邮电出版社,2021.
- [3] BREBNER P, CECCHET E, MARGUERITE J, et al. Middleware benchmarking: approaches, results, experiences[J]. Concurrency and Computation Practice and Experience, 2005, 17(15):1799-1805.
- [4] LIU Y, FEKETE A, GORTON I. Predicting the performance of middleware-based applications at the design level [J]. ACM SIGSOFT Software Engineering Notes, 2004, 29(1):166-170.
- [5] 计世资讯. 2021—2022年软件基础设施(中间件)市 场发展趋势研究报告[R]. CCW research, 2022-03.
- [6] Standard Performance Evaluation Corporation. SPECj-Enterprise 2018 Web Profile[EB/OL]. (2020-08-25)
   [2022-08-07]. https://www.spec.org/jEnterprise -2018Web/.
- [7] Standard Performance Evaluation Corporation. SPECj-Enterprise 2018 Web Profile User's Guide [EB/OL].
   (2020-08-25) [2022-08-07]. https://www.spec.org/ jEnterprise2018Web/docs/UsersGuide.html.

# Performance measurement system of Web application servers: design and verification

Liu Wei<sup>1,2</sup>, He Donghui<sup>1\*</sup>, Yang Panfei<sup>1,2</sup>

 (1. The fifth Electronics Research Institute of the Ministry of Industry and Information Technology of P. R. CHINA, Guangzhou 511300;
 2. The Ministry of Industry and Information Technology Key Laboratory of Performance and Reliability Testing and Evaluation for Basic Software and Hardware, Guangzhou 511300)

Abstract: As one of the three basic software, middleware is an indispensable key product in information system projects. In recent years, although domestic middleware products have experienced rapid growth, compared with global market leaders in terms of performance and stability, it is still an important factor affecting the acceptance of the product market. Meanwhile, in terms of the middleware assessment system and evaluation tools, China has not yet developed a mature, standard technology and productive achievement. A performance measurement system of Web application server based on SPECjEnterprise®2018 Web Profile Benchmark is proposed, the application that is tested in this benchmark emulates an automotive insurance brokerage application, then the measurement system applied to a domestic Web application server, experimental results indicate that the Web application server achieved good results.

Keywords: domestic Web application server; SPECjEnterprise®2018 Web Profile Benchmark; performance measurement system

\_\_\_\_\_

(上接第87页)

# Rician *K*-factor prediction and channel reconstruction of reverberation chamber based on fitting

Zhang Xueying, Zhao Xiang\*

(College of Electronics and Information Engineering, Sichuan University, Chengdu 610065)

Abstract: For the prediction of Rician *K*-factor and the channel reconstruction of a given Rician *K*-factor, a method is proposed to obtain the predicted value or the desired reverberation chamber configuration by fitting the measured values of a few different configurations of Rician *K*-factor. This method is used to fit *K*-factors at different distances between antennas in a reverberation chamber. The results show that the function which replaces the antenna directivity in Holloway expression with polynomial has better fitting effect. This function improves the Holloway expression and makes it applicable to both near and far fields. This method is also applicable to the prediction of *K*-factor and channel reconstruction at different values of other factors such as the quality factor of reverberation chamber.

Keywords: reverberation chamber; Rician K-factor; fitting; error of fitting

DOI: 10.3969/j.issn.1007-1423.2023.02.014

# 基于SPSS的管制工作负荷与航班架次关系分析

郭东鑫\*, 李科扬

(中国民用航空飞行学院空中交通管理学院, 广汉 618307)

摘要:为了确定扇区的容量,对管制工作负荷与航班架次进行了相关性分析。在DORATASK方法的 框架下,将管制工作负荷分为三类。利用 SPSS 软件建立了以航班架次为因变量,以三类管制工作负荷为自 变量的多元线性回归模型。通过拟合优度等指标确定了模型的有效性,对于确定扇区最大容量有一定的指 导意义。

关键词:管制工作负荷;航班架次;相关性分析;多元线性回归

# 0 引言

近年来,随着国民经济的快速发展,我国 民航运输的业务量日益增加,而扇区的容量是 有限的,所以对扇区容量进行及时而准确的评 估是非常有必要的。目前,我国的容量评估体 系已比较成熟,依据中国民航局发布的《机场 时刻容量评估技术规范》,我国的容量评估方法 包含了基于历史统计数据的评估方法、基于数 学计算模型的评估方法、基于计算机仿真模型 的评估方法和基于管制员工作负荷的评估方法。 本文采用 SPSS (statistical package for the social sciences)软件建立的管制员工作负荷与航班架 次的回归模型本质上就是基于管制员工作负荷

目前,针对管制员工作负荷的研究,最常 采用的方法是 DORATASK 和波-布分析法 (MBB)。由于 DORATASK 应用广泛且比较成 熟,因此,国内的相关研究在测量管制员工作 负荷时,常常采用 DORATASK法<sup>[1-2]</sup>。本文借鉴 了 DORATASK法,将管制员的工作进行了分 类,分别是管制员进行陆空通话的时间;管制 员操作鼠标、键盘,填写进程单等工作的操作 时间以及管制员对航班进行排序, 配备航空器 间隔的思考时间<sup>[34]</sup>。

本文利用 SPSS 软件对实际测得的管制员工 作负荷和航班架次数据进行相关性分析,旨在 建立一个能够快速、高效测量管制扇区容量的 回归模型,以期能够为扇区容量评估工作提供 一定的参考。

# 1 数据分析

# 1.1 数据来源

为了更好地研究管制员工作负荷与航班架 次之间的关系,从而达到通过管制员工作负荷 测量管制扇区容量的目的,本研究采集了某管 制扇区8月份的管制扇区通话时长和小时航班架 次数据。通过问卷调查的方法统计了管制员的 操作负荷以及思考负荷。通过通话时长、操作 负荷、思考负荷三者加和的方法得到了管制工 作总负荷。部分数据见表1。

# 1.2 相关性分析

相关分析是研究随机变量之间相关性的一 种统计方法,通过该分析方法可以确定变量之 间的相互性和密切程度。一般来说,相关程度

**收稿日期:** 2022-12-04 修稿日期: 2023-01-12

**作者简介:** \*通信作者:郭东鑫(1997—),男,河南驻马店人,硕士研究生,研究方向为空中交通管理,E-mail:guodongxin8839@ qq.com;李科扬(1996—),男,河南洛阳人,硕士研究生,研究方向为空中交通流量管理

越高,回归分析的结果就越可靠,多数研究在 做回归分析之前先要做相关分析,这是判别回 归分析结果的一个重要依据。因此在对管制员 工作负荷与航班架次进行回归分析之前,首先 要进行相关性分析<sup>[5]</sup>。

在 SPSS 软件中,系统提供了皮尔逊、肯德 尔和斯皮尔曼三种分析方法。由于皮尔逊分析 方法操作简单,同时能避免评分等级膨胀问题, 因此本文选用皮尔逊分析方法,对管制扇区通 话时长、操作负荷、思考负荷、航空器架次和 管制工作总负荷5个变量进行相关性分析,软件 分析结果见表2。

统计学中把低于±0.30以下的相关系数称 为微相关,±0.30~±0.50之间为低度相关, ±0.5~±0.8之间为显著相关,绝对值在0.8以 上为高度相关。由表2可以看出,航空器架次与 思考负荷之间为显著相关;航空器架次与塔进 管制扇区通话时长、操作负荷和管制工作总负 荷之间均为高度相关。此外,所有变量的皮尔 逊相关性在0.01水平(双侧)上显著相关,说明 所有变量之间的相关性是显著的。

# 1.3 模型建立

在进行回归分析时,有曲线拟合和线性回 归两种方法可供选择。其中,曲线拟合只能进 行单因素的回归分析,而本文旨在建立一个包 含管制扇区通话时长、思考负荷、操作负荷等 多因素的回归模型,此外,多元线性回归还具 有以下优点<sup>[67]</sup>:

(1)建模速度快,不需要很复杂的计算,在 数据量大的情况下运行速度依然很快;

(2) 可根据系数给出每个变量的理解和解释;

(3) 对异常值很敏感,可用于异常数据判断。

因此本文采用多元线性回归分析的方法进 行回归分析。

对航班架次的影响因素进行分析,确定航 班架次为因变量,自变量为影响航班架次的相 关因素,包括管制扇区通话时长、思考负荷、 操作负荷等。假定因变量与自变量之间是线性

表 1 管制员工作负荷与航班架次统计

日期时段	塔进管制扇区通话时长/秒	操作负荷/秒	思考负荷/秒	航空器架次/架	管制工作总负荷/秒
8月1日00:00-00:59	928	83	114	11	1125
8月1日01:00-01:59	1261	70	138	9	1469
		•••••			
8月31日13:00-13:59	1278	138	46	13	1462
8月31日14:00—14:59	1856	141	47	13	2044

表 2 皮尔逊法相关性分析结果

		塔进管制扇区通话时长/秒	操作负荷/秒	思考负荷/秒	航空器架次/架	管制工作总负荷/秒
塔进管制扇区	皮尔逊相关性	1.000	0.846**	$0.487^{**}$	$0.878^{**}$	0.998**
通话时长	Sig.(双尾)		0.000	0.000	0.000	0.000
揭佐吾苔	皮尔逊相关性	0.846**	1.000	0.311**	$0.970^{**}$	0.872**
探作贝何	Sig.(双尾)	0.000		0.000	0.000	0.000
思考负荷	皮尔逊相关性	0.487**	0.311**	1.000	0.527**	0.499**
	Sig.(双尾)	0.000	0.000		0.000	0.000
脑会思加发	皮尔逊相关性	$0.878^{**}$	$0.970^{**}$	0.527**	1.000	0.904**
机全奋采伏	Sig.(双尾)	0.000	0.000	0.000		0.000
管制工作 总负荷	皮尔逊相关性	0.998**	0.872**	0.499**	0.904**	1.000
	Sig.(双尾)	0.000	0.000	0.000	0.000	

\*\*表示在 0.01 级别(双尾)水平下,相关性显著。

# 关系,如式(1)所示:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$
(1)

其中: Y是航班架次;  $\beta_0 \sim \beta_4$ 是回归系数;  $X_1 \sim X_4$ 分别是管制扇区通话时长、思考负荷、操作 负荷和管制工作总负荷;  $\varepsilon$ 是随机误差。将实测 数 据 带 入 公 式 (1), 可 获 得 N 组 观 察 数 据  $(X_{i1}, X_{i2}, \dots, X_{ip}, Y_i)$ (i=1, 2, …, n),则公式(1)可 以表示为:

$$Y_{1} = \beta_{0} + \beta_{1}X_{11} + \beta_{2}X_{21} + \beta_{3}X_{31} + \beta_{4}X_{41} + \varepsilon_{1}$$

$$Y_{2} = \beta_{0} + \beta_{1}X_{12} + \beta_{2}X_{22} + \beta_{3}X_{32} + \beta_{4}X_{42} + \varepsilon_{2}$$
.....
$$Y_{n} = \beta_{0} + \beta_{1}X_{1n} + \beta_{2}X_{2n} + \beta_{3}X_{3n} + \beta_{4}X_{4n} + \varepsilon_{n}$$

写成矩阵形式为:  $Y = X\beta + \varepsilon$ ,通过已知的 X、Y数据使用 SPSS 软件可以求解出 $\beta_0 \sim \beta_4$ ,最 后可得到模型的表达式。

由表1可以看出,塔进管制扇区通话时长在 几百秒至一二千秒之间,而操作负荷与思考负 荷的大小通常在几十秒至一二百秒之间,通话 时长的大小与操作负荷和思考负荷的大小不在 一个数量级上,因此在进行回归分析之前,首 先要对数据进行标准化处理,消除数量级对回 归分析的影响<sup>[8]</sup>。

对标准化后的数据进行回归分析, SPSS输出结果如表3所示。

模型	R	$R^2$	调整后R <sup>2</sup>	标准估算的错误
1	0.878	0.771	0.770	2.10256
2	0.885	0.784	0.783	2.04326
3	0.999	0.998	0.998	0.17957

表 3 模型摘要

其中,模型1代表以航班架次为因变量,以管制 扇区通话时长为自变量的模型;模型2代表以航 班架次为因变量,以管制扇区通话时长和思考 负荷为自变量的模型;模型3代表以航班架次为 因变量,同时以管制扇区通话时长、思考负荷 和操作负荷作为自变量的模型。由表3可以看 出,模型3的拟合优度*R*<sup>2</sup>的大小达到了0.998, 说明模型3的拟合效果最好,回归方程见式(2):

 $Y = 11.986 + 0.014X_1 + 1.093X_2 + 3.905X_3$ 

(2)

## 2.1 多重共线性检查

在进行线性回归分析时,可能会出现自变 量之间彼此相关的现象,这种现象被称为多重 共线性<sup>[9]</sup>。适度的多重共线性并不会影响分析 结果,但如果共线性问题较为严重,会导致分 析结果不稳定,出现回归方程与实际情况完全 相反的情况,因此在建立模型后,有必要对模 型进行多重共线性检查。

SPSS分析结果如表4所示。

表 4 共线性检查

模型 -		共线性统计		
		Tolerance	VIF	
	常量	—	_	
3	塔进管制扇区通话时长	0.229	4.373	
	思考负荷	0.727	1.375	
	操作负荷	0.271	3.694	

表4显示了共线性诊断的两个统计量:容差 (Tolerance)和方差膨胀因子(VIF),一般认为如 果Tolerance < 0.2或者VIF > 10,则要考虑自变 量之间存在多重共线性的问题,而表4显示的 Tolerance均大于0.2,VIF均小于10,说明模型 自变量之间不存在共线性或者自变量之间的共 线性程度不足以影响分析结果。

# 2.2 方差分析

利用 SPSS 做回归分析,得出方差分析 (ANOVA)表,用以显示模型的方差和残差以及 F检验的结果<sup>[10]</sup>,结果如表5所示。其中,"回 归"代表所使用的方法;"回归平方和"代表反 应变量的变异中,在回归模式下所包含的自变 量能够解释的部分;"残差平方和"代表反应变 量的变异中没有被回归模型所包含的变量解释 的部分,"自由度"代表可以自由取值的变量的 个数;"均方"代表了方差与自由度相除的结 果;"F"代表F检验统计量,其值检验样本的 结果能够代表总体的真实程度。

由表5可以看出,模型1、2、3的"回归平 方和"的数值呈递增状态,表明反应变量的变 异中,在回归模式下所包含自变量可以解释的

ł	模型	平方和	自由度	均方	F	显著性
	回归	8612.123	1	8612.123	1948.113	0.000
1	残差	2564.036	580	4.421		
	总计	11176.159	581			
	回归	8758.886	2	4379.443	1048.991	0.000
2	残差	2417.273	579	4.175		
	总计	11176.159	581			
	回归	11157.522	3	3719.174	115345.445	0.000
3	残差	18.637	578	0.032		
	总计	11176.159	581			

表 5 ANOVA 表

部分也是依次递增的;而"残差平方和"的数 值依次递减则说明反应变量的变异中不能被回 归模型所包含的变量解释的部分是依次减少的, 这两项指标说明模型1、2、3能够解释模型变异 的能力呈递增趋势,从这点来看,模型3是最优 的。"F"是用来检验回归方程是否具有统计学 意义,一般来说,F值越大,回归方程的统计学 意义越显著,显然模型3的F值最大,所以模型3 在所有模型中最有统计学意义。三个模型的显著 性值均<0.01,说明三个模型的显著性均很强。

综合以上分析,模型3的各项指标均满足要求,可以运用在实际工作环境中,通过设置管制员工作负荷的阈值确定出一个最大的航班架次,进而确定单位时间内扇区的最大容量。

## 3 结语

本文从多元线性回归模型出发,根据实测的管制员工作负荷与航班架次构建模型,通过 SPSS软件对数据进行标准化处理以消除量纲影 响,建立了以航班架次为因变量,以管制扇区 通话时长、操作负荷和思考负荷为自变量的多 元线性回归模型,通过拟合优度、共线性检查、 方差分析等指标验证了模型的有效性,模型对 于确定扇区容量有一定的参考价值。

本文建立了一个多元线性回归模型,综合 考虑了三类管制工作负荷对于航班架次的影响, 但模型没有考虑管制工作负荷对于管制员的工 作状态的影响。在实际工作中,管制员的工作 负荷不可能始终如一地保持稳定,而不同的管 制负荷下,管制员的工作状态可能不同,所使 用的工作时间可能并不服从线性变化。今后的 研究中,可以考虑将管制工作负荷划分为不同 的等级,分段建模,得到更加准确的管制负荷 与航班架次的对应关系。

#### 参考文献:

- [1] 袁乐平,孙瑞山,刘露.基于DORATASK的管制员 工作负荷测量方法研究[J].安全与环境学报, 2014,14(3):76-79.
- [2] 吕庆文,樊树海,徐文浩,等.基于DORATASK法的标准作业负荷评估模型[J].中国安全科学学报, 2020,30(8):183-188.
- [3] 王硕.空中交通管制员扇区工作负荷浅析[J].电子 制作,2014(22):231.
- [4] 李钱.基于改进DORATASK法的空中交通管制员 配备与工作负荷评估[J].计算机与数字工程, 2019,47(8):1996-1998,2038.
- [5] 周静. SPSS 在数学建模中的应用实例[J]. 天津职业 院校联合学报,2012,14(11):93-96.
- [6] 陈坤,陈天之,覃金寿,等.基于多元线性回归的大
   学英语四级成绩分析[J].科技创新与应用,2018
   (10):37-39.
- [7] 秦帅.高校社会实践安全风险影响因素研究[J].中 国安全科学学报,2021,31(1):18-23.
- [8] 夏丽华,谢金玲,等. SPSS数据统计与分析标准教程[M].北京:清华大学出版社,2014:173-175, 192-198,209-211.
- [9] 黄利花. SPSS统计方法及适用性分析[J]. 延安职业 技术学院学报,2014,28(4):83-84,91.
- [10] 刘天成,凌书平,许景生.基于 SPSS 软件回归分析 的基站用电建模与应用[J].软件,2021,42(4): 57-61.

(下转第102页)

DOI: 10.3969/j.issn.1007-1423.2023.02.015

# 基于胶囊模型的短文本细粒度情感分类

邵辉

(广东科学技术职业学院计算机工程技术学院,珠海 519090)

摘要:针对短文本的细粒度情感分类,提出一种基于胶囊网络的模型,在词嵌入层使用BERT预训练 模型对上下文序列、目标词序列编码;然后采用LSTM对上下文序列、目标词序列提取相应的特征;在注 意力编码层均采用多头注意力完成注意力编码;最后在胶囊层完成情感分类。实验采用三个情感极性,分 别对应积极、中性和消极情感。结果表明,该方法不仅提高了分类精度,而且F1值也得到了不小的提升, 说明提出的模型对短文本的细粒度情感分类是有效的。

关键词:胶囊模型;短文本;BERT;情感分类

# 0 引言

情感分析(sentiment analysis)是指人们对各种服务、产品,包括其属性的情感、评价和观 点等的分析研究,又被称为观点挖掘<sup>[1]</sup>。传统 的文本情感分析<sup>[24]</sup>重在对句子、文档级别进行 分析,得出文本中主要观点所表达的情感倾向, 通常分为积极、中立以及消极这三种。但互联 网普及之前,情感分析的相关研究成果很少, 一是因为收集情感文本语料存在困难,很难获 取所需要的文本语料。二是文本处理、分析方 法达不到应用要求。进入新世纪后,互联网迎 来爆发式增长,基于此,各种海量的文本信息 在互联网上不断出现,特别是社交领域和电商 平台,促使情感分析研究快速发展。

在社交领域和电商平台等,篇幅短小的短 文本是互联网用户经常使用的。短文本文字有 限,但往往包含明确的观点,特别是在社交软 件中,更是经常出现带有用户直接偏好的短文 本,这些都非常有利于情感分析领域的研究。 当前, 情感分析研究已经发展到涉及社会的方 方面面, 互联网上每天都有无数的短文本产生, 对短文本的情感分析工作可以让政府掌握舆情, 及时掌握社会大众的诉求, 以便于保证社会的 稳定和谐; 可以让企业和单位了解用户对服务 和产品的意见, 以便于做出更好的服务和产品。 由此可见, 短文本情感分析工作是十分有意义 的。本文以短文本为研究对象, 从目标级的情 感分类角度提出了一种基于胶囊模型的方法。

## 1 相关研究

目标级的情感分类(aspect-based sentiment classification, ABSC)与情感分析(SA)不同<sup>[5]</sup>,其 重在发现文本中实体方面有关的情感。例如对 于评论文本,就不同于以往对整句做出情感分 析,而是根据句中不同的实体分别做出不同的 情感分析,如此一个短文本评论就可能有多个 情感极性。比如,"今天一位朋友请我们吃饭, 饭菜的味道还可以,但是人太多了,我们等了 很久才吃上,而且价格也不便宜。"从这个例子

收稿日期: 2022-09-28 修稿日期: 2022-11-24

基金项目:2021年广东省高职教育教学改革研究与实践项目(GDJG2021153);广东科学技术职业学院 2022年度校级科研项目(XJPY202209):基于胶囊模型的细粒度短文本情感分析技术研究

作者简介: 邵辉(1977—),男,湖北黄冈人,硕士,副高,CCF会员,研究方向为深度学习

可以看出来,此评价带有三个实体方面的情感 倾向:一是餐厅的味道还可以,这是积极的情 感倾向;二是等待的时间太久,这是负面的情 感倾向;三是价格太贵,这也是消极的情感倾 向。如果是情感分析(SA)任务,那就不会有这 么多情感极性分析,可能就这个文档或整句给 出一个情感极性,那就会忽略其它包含在文档 或整句中的情感,只有目标级的情感分类才能 做到更细粒度的要求。

目前 ABSC 相关研究中, 卷积神经网络<sup>[6]</sup>、 循环神经网络[7]和循环自编码模型[8]等神经网络 模型已经取得了很好的效果,但是目前存在数 据集标注成本高昂、有时需要附加的语言知识 辅助等问题。胶囊网络出现后,对解决上述问 题起到了很好的作用。胶囊网络由一组神经元 构成,是基于动态路由的结构<sup>[9]</sup>。胶囊利用动 态路由算法完成参数互相传递,每个类别的语 义由高层胶囊的输出向量来表示;每个实体的 实例化参数由激活向量来表示:每个情感极性 的预测概率则由向量长度来表示。文本分类中 应用胶囊网络是 Zhao 等<sup>[10]</sup>第一次实现,发现胶 囊网络不仅能保持灵活的表达能力,同时提高 了编码的有效性; Chen等<sup>[11]</sup>提出了一种迁移胶 囊网络模型,用于将文档级别的知识迁移到面 向目标的情感分类。

在前面工作的基础上,本文设计的胶囊模型第一利用BERT预训练模型充分挖掘文本蕴含的情感语义信息,使模型具有更加丰富的情感语义表达;第二利用多头注意力机制让各类特征进行交互,抽象更深层次的上下文内部语义关联;第三采用胶囊网络生成最终的文本表征,从而实现了更好的短文本细粒度情感分类。

## 2 胶囊网络模型

给定上下文嵌入<sup>[12]</sup>,其由n个词构成上下 文序列 $s = \{w_1, w_2, \dots, w_n\}$ 。再给定目标嵌入,其 由k个目标构成目标序列, $t = \{a_1, a_2, \dots, a_n\}$ , 很显然, $a \ge s$ 的一个子序列。由此本文的目标 可以表示为pol =  $f_{pol}(s, a_i)$ ,其中 $f_{pol}$ 是非线性变 换函数。本文设计的胶囊模型如图1所示,分为 词嵌入层、特征提取层、注意力编码层和胶囊 层,共四层。



#### 图 1 胶囊模型各层

在词嵌入层使用 BERT 预训练模型,将包含 n个词的上下文序列转换成 $s = \{v_1, v_2, \dots, v_n\}$ ,其  $v_i$ 表示上下文序列第i个词的d维向量,句子 的输入词向量矩阵则是S;同理,目标实例则转 换为 $T = \{v_{\alpha}, v_{\alpha+1}, \dots, v_{\alpha+m-1}\}$ ,其包含m个词, 形成目标词嵌入序列。其中 $v_i$ 表示目标实例第i个词的d维向量。

在特征提取层,对于上下文序列形成的词 向量矩阵,其中的每个词的依赖关系利用 LSTM进行建模,充分在BERT预训练模型的基 础上对隐含语义进行挖掘,得到隐藏状态序列  $L_h = \{l_1, l_2, \dots, l_k\}, 形成上下文序列的高阶特征;$ 对于目标词序列,其中属性实例的各词依赖关系也利用LSTM进行建模,同理充分对隐含语义进 $行挖掘,得到隐藏状态序列<math>T_h = \{t_1, t_2, \dots, t_m\},$ 最后也形成目标词序列的高阶特征。

在注意力编码层,上下文序列、目标词序 列均采用多头注意力完成注意力编码。多头注 意力机制可以简单有效地对上下文依赖关系进 行抽象,并捕获句法和语义特征。本层继续对 上层输入表示做进一步挖掘,并生成两类输出 特征。

最后在胶囊层,对两个多头注意力的输出 O<sub>k</sub>和O<sub>m</sub>进行封装加工,最后转换为矢量胶囊集 合。胶囊网络确定彼此关系是依据动态路由协 议,利用反复迭代的方式直到收敛成功。在注 意力编码层,因其输出只能表达局部特性,无 法对句子级别进行全局语义表示,因此本层在 水平方向对注意力编码层的输入采用全局最大 池化进行压缩,使得输出特征在各子空间内聚 合。同时利用 squash 函数将胶囊向量的模长压 缩到1以内,用来表示该特征存在的概率。最后 最终情感分类输出层由多类胶囊构成,完成每 个情感极性分类。

# 3 实验

本文评测标准数据集采用 semeval 2014 的餐 厅评论数据集,对于不同的目标实体,分成三 类情感极性:积极,中性和消极。此数据集中有 极少量数据标记为冲突,将其予以删除。数据 集的情况如表1所示。

表 1 餐厅评论数据集 单位:个

餐厅评论数据集	积极	中性	消极	冲突
训练集	2164	807	637	91
测试集	728	196	196	14

本模型采用预训练 BERT,其维度设置为 768。为保证性能,学习率设为2e-5,动态路由 迭代次数设置为7,多头注意力头数设置为8。 模型最后利用分类精度和F<sub>1</sub>值来评价模型性能, 并同两个基线模型 RAM、TransCap进行比较, 结果如表2所示。从表2可知,胶囊模型的分类 精度和F<sub>1</sub>值均高于 RAM和TransCap。

表 2 对比结果

模型	分类精度	$F_1$ 值
RAM	0.8093	0.7068
TransCap	0.7831	0.7078
胶囊模型	0.8301	0.7859

# 4 结语

从实验结果看,本文采用的胶囊模型,其 网络深度的有效增加提升了模型性能;采用预 训练模型BERT也提高了分类精度,并且对参数 微调能可继续提高模型性能;采用胶囊模型则 不仅有效提高了分类精度,而且F<sub>1</sub>值也得到了 不小的提升,这说明本文模型对短文本的细粒 度情感分类是有效的。

#### 参考文献:

 PANG B, LEE L. Opinion mining and sentiment analysis [J]. Foundations and Trends in Information Retrieval, 2008, 2(1/2):1-135.

- [2] TABOADA M, BROOKE J, TOFILOSKI M, et al. Lexicon-based methods for sentiment analysis[J]. Computational Linguistics, 2011, 37(2):267-307.
- [3] WILSON T, WIEBE J, HOFFMANN P. Recognizing contextual polarity: an exploration of features for phrase-level sentiment analysis[J]. Computational Linguistics, 2009, 35(3): 399-433.
- WAWRE S V, DESHMUKH S N. Sentiment classification using machine learning techniques [J]. International Journal of Science and Research (IJSR), 2016, 5 (4):819-821.
- [5] 滕磊,严馨,徐广义,等.使用胶囊网络的细粒度情感分析方法[J].小型微型计算机系统,2020,41 (12):2550-2556.
- [6] LEI T, BARZILAY R, JAAKKOLA T. Molding CNNs for text: non-linear, non-consecutive convolutions [EB/OL]. arXiv Preprint arXiv: 1508.04112, 2015.
- [7] TAI K S, SOCHER R, MANNING C D. Improved semantic representations from tree - structured long short - term memory networks [C] //Proceedings of the Association for Computational Linguistics, 2015: 1556-1566.
- [8] BALY R, HAJJ H, HABASH N, et al. A sentiment treebank and morphologically enriched recursive deep models for effective sentiment analysis in Arabic [J]. ACM Transactions on Asian and Low-Resource Language Information Processing(TALLIP), 2017, 16 (4):1-21.
- [9] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules [C] //Proceedings of the Advances in Neural Information Processing Systems, 2017:3856-3866.
- [10] ZHAO W, YE J, YANG M, et al. Investigating capsule networks with dynamic routing for text classification [EB/OL]. arXiv Preprint arXiv: 1804.00538, 2018.
- [11] CHEN Z, QIAN T. Transfer capsule network for aspect level sentiment classification [C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), 2019:547-556.
- [12] 邓钰. 面向短文本的情感分析关键技术研究[D]. 成都:电子科技大学,2021.

# Fine-grained sentiment classification of short text based on capsule model

# Shao Hui

(Computer Engineering Technology College, Guangdong Polytechnic Science and Technology, Zhuhai 519090)

Abstract: For fine-grained sentiment classification of short texts, a model based on capsule network is proposed. The BERT pre-training model is used in the word embedding layer to encode the context sequence and target word sequence. Then LSTM is applied to extract the corresponding features on the context sequence and target word sequence. In the attention encoding layer, multi-head attention is used to complete the attention encoding. Finally, the emotion classification is completed in the capsule layer. Experimental adopts three affective polarities, corresponding to positive, neutral and negative affect, respectively. The results show that this method not only improves the classification accuracy, but also improves the F1 value, which shows that the model is effective for fine-grained sentiment classification of short texts.

Keywords: capsule model; short text; BERT; sentiment classification

(上接第98页)

# Analysis of the relationship between controller workload and flight sorties based on SPSS

Guo Dongxin<sup>\*</sup>, Li Keyang

(School of Air Traffic Management, Civil Aviation Flight University of China, Guanghan 618307)

Abstract: In order to determine the capacity of the sector, the correlation analysis between the control work load and the number of flights was carried out. In the framework of DORATASK method, the control workload is divided into three categories. Using SPSS software, a multiple linear regression model was established with flight sorties as the dependent variable and three types of control workload as the independent variable. The validity of the model is determined by goodness-of-fit and other indicators, which has certain guiding significance for determining the maximum capacity of the sector.

Keywords: ATC controller workload; sortie; correlation analysis; multiple linear regression

# 开发案例

文章编号:1007-1423(2023)02-0103-07

DOI: 10.3969/j.issn.1007-1423.2023.02.016

# 一种面向铁路领域在线客服内容违规和风险的应急管理方法

皮尔达伟斯·巴吐尔<sup>1</sup>,刘 捷<sup>1,2\*</sup>

(1.西南交通大学计算机与人工智能学院,成都 611756;2.可持续城市交通智能化教育部工程研究中心,成都 611756)

摘要:基于情景-应对模型和知识图谱,提出一种面向铁路领域在线客服内容违规和风险的应急管理方法。首先采用 Protégé 构建情景模型和应对模型,分别包含事件、形式、风险识别本体,及措施、方案本体。基于预设的情景模型,结合客服内容违规和风险事件数据,对事件进行风险识别,得到事件的形式实体及形式实体对应的风险识别实体;基于预设的应对模型,结合事件的形式实体,得到客服内容违规和风险事件的对应措施和方案实体。然后将知识导入 Neo4j,进行存储、检索和应用。最后,以客服平台图片侵权的案例对该方法进行了验证。通过建立情景-应对模型,考虑事件的情景和应对之间的全局关联性,使得制定的应对措施和方案准确可靠。

关键词: 应急响应; 情景-应对; 在线客服内容违规和风险; 情景模型; 应对模型

## 0 引言

网络信息内容治理已成为国家战略的重要 部分,"网络信息内容生态治理规定"明确了网 络信息内容治理在治理目标、主体、客体、过 程、工具5个方面的核心要义<sup>[1]</sup>。近年来,在国 铁企业信息化治理过程中,借助人工智能的AI+ 内容安全技术能够高效发现问题,但在发现违 规和风险问题之后,如何及时识别事件类别、 判断风险级别、定位响应部门、决策应急预案 仍然缺乏研究。刘铁民<sup>[2]</sup>指出突发事件情景构 建是当前公共安全领域最前沿科学问题之一, 采用"情景-应对"的理论与方法,是国内外公 共安全领域近年来提出的一种行之有效的科学 手段。本文基于"情景-应对"模型和知识图 谱,提出了一种面向铁路领域在线客服内容违 规和风险的应急管理系统和方法。

# 1 违规和风险应急管理研究现状

朱世强等<sup>[3]</sup>指出知识图谱等技术的发展, 能够加速将内容安全治理向自动化、智能化、 高效化、精准化方向推进,并建议建立健全合 法有效的监管机制和人工智能内容安全标准体 系。王建亚等<sup>[4]</sup>对网络信息内容安全风险进行 研判和特征分析,将网络信息内容安全分为9 类,分析了每种风险的内涵和表现,并对各类 型风险进行了场景化解构。白文琳等<sup>[1]</sup>按照戴 明环PDCA循环思路,同时将治理目标、主体、 客体、过程、工具等政策要素全部纳入,设计 了"PDCA-五维要素"治理行动逻辑。范维澄 等<sup>[5]</sup>提出了公共安全体系的三角形模型,突发 事件及其应对中存在三个主体:突发事件、承 灾载体、应急管理。张辉等<sup>[6]</sup>在面向国家应急 平台体系的基础科学问题和集成平台研究中突

基金项目: 国家铁路集团有限公司科技研究开发重点课题(N2021S010)

作者简介:皮尔达伟斯·巴吐尔(1998—),男,新疆喀什人,学士,学生,研究方向为知识图谱;\*通信作者:刘捷(1979—),男,四川资阳人,硕士,讲师,研究方向为软件与理论、知识图谱等,E-mail:ludgee@swjtu.edu.cn

**收稿日期:** 2022-11-22 修稿日期: 2022-12-10

出了"情景-应对"型决策的重要性。郭世刚 等<sup>[7]</sup>研究了突发事件应急决策方法等演变,将 现有决策模型分为经验直觉法、"预测-应对" 模型、"情景-应对"模式三类。仲秋雁等<sup>[8]</sup>基 于知识元构建了非常规突发事件情景模型。王 宁等<sup>[9]</sup>提出了一种基于应急案例的情景决策支 持方法。吴艳华等<sup>[10-11]</sup>研究了"情景-应对"模 型在铁路领域的应用。

# 2 情景模型设计

在本文创建铁路领域在线客服违规和风险 管理知识图谱过程中,在客服平台和相关网站 和公众号采集近年来在客服违规方面发生过的案 例,再通过网上检索到的案例对各个实体的属性 进行保存,通过整合来构建铁路在线客服违规的 事件库。本文创建的情景模型如图1所示。



图 1 情景模型

其中,事件(Event)实体是用来存储所有事件的实体,在识别出铁路在线客服内容违规和风险后,都存储在这里,是整个情景模型的主体。用此实体来创建情景表的主体,通过各种属性来充实事件实体的各个知识内容,此实体的实体父类为情景,与形式为不相交的类,并且没有添加其他的公理等描述。事件的实例是整个事件的承载者。

将铁路客服过程中产生的事件实例采集到 Protégé中,如图2所示。通过对事件的属性: 事件id、事件类型名称、事件发生原因、对象、 事件发生时间来对应一个唯一的事件,如图 3 所示。

Data properties	Individuals by class × DL C	Query × OntoGraf			
Active ontology	× Entities × Classes ×	Object properties			
Class hierarchy (inferre	Annotations Usage				
Class hierarchy	Annotations: 事件	20800			
Class hierarc 🛙 💷 🗖 🛙	Annotations				
😫 🕵 🐹 erted					
• owl:Thing	Description <sup>,</sup> 事件	20000			
●措施					
──● 方案 ✓─● 案例	◆ 12306客服态度	<b>7</b> @×			
	◆ app奔溃	?@X			
	◆ app异常	<b>?@</b> ×			
- 11-4	上海客服态度问题	<b>?@</b> ×			
	◆ 图片侵权	?@×			
	◆客服工作缓慢 Ass	Asserted in: http://www.113.com			
	◆客服繁忙	<b>?@</b> ×			
	◆ 客服违规	?@X			
	文章内容错误	?@×			
	文章字体侵权	?@×			
	◆ 有用户受到信息骚扰	7@X			
	◆ 线上功能失效	7 @ ×			
	◆虚假网站	<b>?@</b> X			

图 2 Protégé中的事件实例

Data properties	Individuals by	class	× DL (	Query	× OntoGraf	×
Active ontology	× Entities	× Classes	×	Object	properties	×
Datatypes Individuals Annotation properties Data properties	■ ◆ 图片侵 Annotations	权 — http://w Usage	/ww.113	.com#图	片侵权	
Object properties	Annotation	13. 国力 1973				
Classes	Annotations	Ð				
Individuals:	1					
◆ 12306客服态度 ◆ app奔溃	Property a	cortions: 图	上/3 北7		mc	
◆ app崩溃	Floperty as	sertions. 🔄	11217			
<ul> <li>▲ app开常</li> <li>▲ 上海客服态度问题</li> <li>▲ 侵权方案</li> </ul>	Object propert	ty assertions 🕂 形式 图片侵权相	ж		7@X	0
<ul> <li>◆ 公众号文章错误</li> <li>◆ 公告消息违规</li> </ul>	Data property	assertions 🕂	5-14"		000	
● 取消票的效益开要求3		. 王问词 2017-	-3-14			× I
● 图片信权	一方家	度55 先度88 1188 上去	cto / 2 407 Dal/			ХI
● 图片侵权相关		生原因 图片内	谷反伏剂	874/149871		21
◆ 外部网站发布	● 事件尖	空石林 图方使	ήX.			21
<ul> <li>◆ 外部网站发文方案</li> <li>◆ 字体侵权</li> </ul>	● 事件ロ				7@×	9
● 32116 〒作編幅	Negative obje	ct property asserti	ons 😈			

图 3 Protégé 中事件的对象属性和数据属性示例

形式(Form)实体是铁路在线客服违规的形 式,铁路客服违规通常有多个形式,比如一个 事件是文章中所采用的图片是其他公司所注册 的,此事件的形式就被定义为图片侵权。事件 和形式实体是构建情景表的主体成分,形式实 体与事件实体同属于情景实体父类。此实体的 不相交实体为事件。此实体包含的实例包括公 众号文章错误,公告消息违规等形式的实例, 这些实例构成了事件的形式的集合,将具有相 同形式性质的事件打包成一个形式,方便应对 措施采取的统一性。

在应急事件发生的过程中,将事件中的相 似类型抽象归纳于一起,形成客服违规事件对
应的形式,例如,图片侵权相关、公告消息违规、隐私信息采集告知、公众号文章错误、外部网站发布、客服无法联系、投诉无门、无法线上退票、字体侵权、客服态度恶劣、非实名购票、app崩溃、客服误导乘客等。在Protégé中创建形式的实例如图4所示,图5展现了主要的数据属性(形式id、形式名称)和对象属性(形式\_措施、形式\_风险识别)。

Data properties	× II	ndividuals by	class	× DL Q	uery	× OntoGraf	>
Active ontology	×	Entities	× Classes	×	Object p	properties	>
Class hierarchy (inf	erred)	Annotations	Usage				
Class hierarchy		Annotatio	ns: 形式			208	
Class hierarc 🛛 🛛		Annotations	0				
😫 🕵 🐹 erte	ed 📀	Annotations					
∽		Descriptio	n: 形式			208	
<ul> <li>措施</li> <li>方案</li> </ul>		Instances 🕂					
		<b>◆</b> app崩溃				700	3
● 风险识别	别	◆公众-	号文章错误				3
● 形式		◆ 公告注	肖息违规			<b>?@</b> ?	3
		◆ 图片(	曼权相关			702	3 1
		◆ 外部	网站发布			?@?	3
		◆ 字体(	曼权			700	3
		◆客服:	态度恶劣			200	
		◆ 客服无法联系				?@	3
		<ul> <li>◆ 客服误导乘客</li> <li>◆ 投诉无门</li> </ul>	吴导乘客			?@	
			モ门			700	
		◆ 无法线上退票		A	sserted i	n: http://www.113	.com
		● 隐私	言息采集告知			200	a.
		●非实:	名购票			000	5

图 4 Protégé中的形式实例



图 5 Protégé中形式的对象属性和数据属性示例

将唯一的事件id和形式id以事件\_形式的关 系形成联系,将事件归合为一个形式,通过形 式再来确定所应该采取的应急解决方案。

风险识别(Risk Identification)实体是铁路在 线客服违规后,通过识别风险的等级来确定事 件所带来的影响,根据影响的程度来划分事件 的风险优先级,再根据事件的优先级来确定事 件的处理顺序。此方法能够让事件处理更加有 序,通过风险识别来大大缩小事件的影响程度。 风险识别的属性包含:风险类别名称、风险类 别id、优先级、响应速度、影响程度,如图 6 所示。风险识别实体是应对计划中的重要一环, 确定事件的等级,其父类为情景,与事件、形 式实体为不相交的类。没有添加其他的公理等 内容。此实体包含的实例包括了风险级别1、风 险级别2、风险级别3。这些实例中保存了事件 的解决顺序和带来影响的数据信息,对事件出 现时,能快速、准确、有序地解决事件起到关 键作用。



图 6 Protégé 中风险识别实例的数据属性示例

在铁路领域客服违规事件发生的初期,能 充分地掌握事件(E)、形式(F)、风险识别(R) 等信息,一个事件(E1)通过各个属性来判断事 件形式(F1),通过事件的事件id(E1)属性对应 一个事件的形式id(F1),以记为(E1-F1)的事 件\_形式关系来表示具体事件E1的对应形式F1。

通过得到的事件所属形式,来确定事件所归 属的形式,进而确定一个形式所对应的风险识别 中的风险等级,通过确定的形式id(F1)对应风险 类别id(R1),从而从关系形式\_风险识别来对应, 记为(F1-R1)。通过事件、形式、风险识别的对 应关系可得知事件的明确信息为(E1-F1-R1),某 事件的详细情景通过这些关系构建出来。

## 3 应对模型设计

事件的事件集合表示为(事件、形式、风险 识别),也可以将事件集对应事件的情景集合, 再通过各个实体的数据属性来使知识的表示更 加详细。将事件的应对策略创建为事件的应对 集,即为(措施、方案),并以这些应对策略中 的数据属性将措施的具体行使方式和措施完成 后对后续存在的问题进行精化的解决方式。本 文创建的应对模型如图 7所示。



#### 图 7 应对模型设计

措施(Measure)是与事件、形式、风险识别 所构成的情景相对应的应对过程的主体内容, 措施拥有措施名称、措施id、部门等数据属性。 能够明确采取措施的响应部门,确定了措施在 行使的过程是由什么部门来进行,使得分工明 确且能够通过对事件的处理来评定部门的绩效。 此实体包含的实例包括了取消票的效益并要求 乘客实名购票、告知隐私信息采集风险等实例, 这些实例集合成了事件应对集的快速响应的解 决措施,是直接解决事件的第一步的手段集合。



图 8 Protégé中措施的对象属性和数据属性示例

确定事件的具体情景模型的详细信息后, 对情景模型创建应对的模型,通过形式与措施 的关系属性形式\_措施来确定事件的应对策略, 从措施的数据属性措施id来确定形式(F1)的对应 的措施(M1),将形式和措施形成的关系形式\_措 施关系记为(F1-M1),以此关系来确定形式为 F1的事件所应采取的措施为M1,反向的推理可 以推理出事件的完整关系,事件E1的形式为 F1.形式名称,此形式风险等级为R1.风险类别 名称,所应采取的措施为M1.措施名称的应对 策略。针对每一个不同事件、形式、风险识别 的过程有不同的应对策略进行匹配。

方案(Scheme)是指对事件措施的后续补充, 方案包括:方案名称、方案id、方案内容、完 成速度,如图9所示。方案内容是对事件采取 的措施的后续补充,措施完成后,对后续相关 的部门进行补救,完整地将事件完成,对后续 的影响进一步进行完整的措施;方案中的完成 速度是指方案在实行的过程中应该采取的执行 速度,执行的速度应不低于方案中规定完成速 度。此实体包含的实例包括外部网站发文方案、 客服违规方案等一系列方案的实例,这些实例 可以用来做措施完成后的后续安排,让事件的 影响逐步消除,直到影响完全消除为止。



图 9 Protégé 中的方案实例

在情景-应对模型构建基础完成后,对措施 进行进一步的扩张,因为单纯的措施无法完成 对事件的后续处置的完善,通过扩展的方案可 以完善对事件措施的支援,然后通过措施和方 案的关系,将措施 M1和方案 S1 联系起来,将 两者之间的联系记为 M1-S1。所以事件的整体联 系为 E1-F1-R1和 F1-M1-S1。

通过对事件、形式、风险识别、措施和方 案的详细分析,充分明确了事件的形式和风险 级别,通过这些明确的信息来确定采取措施, 和措施完成后的后续方案的响应主体。这为进 一步深入分析事件的应急事件处理打下了基础。

## 4 铁路领域在线客服违规和风险管理 知识图谱应用实例

本文的主题是铁路客服违规和风险管理知 识图谱,而且要运用到情景应对模型,所以在 此基础上创建知识图谱的实体,需根据知识图 谱一般构建方法,先构建并识别出铁路风险管 理知识图谱中的实体,再确定各个实体之间的 联系,定义实体之间的属性,然后根据实体之 间的联系再进行铁路领域知识图谱数据的抽取、 融合和存储。

## 4.1 Protégé构建知识图谱

本文构建铁路领域在线客服和风险管理知 识图谱采用了Protégé软件来进行设计,知识来 自于铁路在线客服平台的案例抽取,抽取案例 的平台如图 10所示。



#### 图 10 案例平台示例

本文创建的铁路领域在线客服违规内容和 风险管理知识图谱中一个完整的关系实例如图 11所示。



#### 图 11 知识图谱中完整的实例

Protégé还提供了图显示的功能,此功能可 将 Protégé 所构建的知识图谱本体所有的内容通 过图的形式显示出来,如图 12所示。

## 4.2 Neo4j导入和检索知识图谱

将 Protégé 里所创建的铁路领域客服违规与 应急风险管理的知识图谱本体存储到 Neo4j,就 要用到 Neo4j 的扩展性包 neosemantics。在经过 Neo4j 内置操作语句的修改后,将其显示为最后 的实体名称。导入结果如图 13 所示。



图 12 知识图谱可视化



图 13 Neo4j图数据展示内容

Neo4j中采用了基于路径的图索引,基于这种索引所创建的GraphGrep是一种典型的图索引。例如:match(n:'ns0\_事件')return n。此语句可以输出所有事件的实例,如图 14所示。



图 14 事件查询

在图数据库中还可以利用复杂语句,通过 各个实例间所形成的关系来对拥有关系的数据 进行查询,如图 15 所示。

\$ match data* (n:'ns@_非作'('ns@_非作表服名称':'app拼索'})-[r:'ns@_非作_服式']-(m:'n () () () () () () () () () () () () ()	s0 ±	Å
Image: State		
sept (1) ma##4_Bid(1) 目 icke		
Active		
34/10		
A		
A ext		
appear.		
ade *		
1		
apping		

#### 图 15 查询示例

match data= (n: 'ns0\_事件' { 'ns0\_事件类 型名称': 'app异常' } )-[r: 'ns0\_事件\_形式']-(m: 'ns0\_形式') return data。所表示的实际含 义为事件类别名称为app异常的事件所属的形式 是哪个,从图中可看出事件app异常的形式为 app崩溃的形式。

#### 4.3 Neo4j中违规事件的应用示例

为验证本文所提出的基于案例的应急事件 处理办法,对事件客服平台图片侵权的案例进 行情景-应对模型的验证。首先是此事件集(事 件id: 1,形式id: 1,风险类型id: 1),应对集 (形式id: 1,措施id: 1,方案id: 1)。此事件 的基本表述为E(客服平台图片侵权),事件所对 应的形式为F(形式名称:图片侵权相关),此类 形式F所对应的风险识别R1(风险类别名称:风 险等级1),R1的优先集为低,影响程度为有损 信誉,应该尽快解决,所采取的措施M(措施名 称:撤回信息),措施M所对应的方案为S(方案 名称:侵权方案),此方案所采取的方式为对侵 权的图片、字体等进行整改,对侵权的机构进 行道歉,此方案要求的完成速度为不多于1天。

## 5 结语

国铁企业信息化治理过程中,如何借助知 识图谱等人工智能技术将内容安全治理向自动 化、智能化、高效化、精准化方向推进,具有 重要的意义。本文结合知识图谱和"情景-应 对"模型,提出了一种面向铁路领域在线客服 内容违规和风险的应急管理方法。并采用 Protégé和Neo4j进行了知识抽取、知识存储、和 知识应用。并最终以客服平台图片侵权的案例 对该方法进行了验证,该方法能够较好地辅助 铁路领域在线客服内容违规和风险的应急管理。

## 参考文献:

- [1] 白文琳,周毅.网络信息内容生态的安全风险治理
   行动及其转型[J].图书情报工作,2022,66(5):24-32.
- [2] 刘铁民.重大突发事件情景规划与构建研究[J].中 国应急管理,2012(4):18-23.
- [3] 朱世强,王永恒.基于人工智能的内容安全发展战略研究[J].中国工程科学,2021,23(3):67-74.
- [4] 王建亚,马榕培,周毅.网络信息内容安全风险:特征、演变及场景要素解构[J].图书情报工作,2022,

66(5):13-23.

- [5] 范维澄,刘奕.城市公共安全体系架构分析[J].城 市管理与科技,2009,11(5):38-41.
- [6] 张辉,刘奕.基于"情景-应对"的国家应急平台体系 基础科学问题与集成平台[J].系统工程理论与实 践,2012,32(5):947-953.
- [7] 郭世刚,张鹏,吴立志,等.突发事件应急决策方法 演变研究[J].中国公共安全,2018(3):40-43.
- [8] 仲秋雁,郭艳敏,王宁,等.基于知识元的非常规突

发事件情景模型研究[J].情报科学,2012,30(1): 115-120.

- [9] 王宁,郭玮,路国粹.基于应急案例的情景决策支持 方法研究[J].运筹与管理,2017,26(1):68-75.
- [10] 吴艳华,王富章,李平.韩国铁路应急管理及启示 [J].中国铁路,2011(9):69-72.
- [11] 吴艳华,王富章,李平,等.基于资源共享的高速铁 路监测预警与应急平台研究[J].铁路计算机应用, 2015(9):17-21.

# An emergency management method for online customer service content violations and risks in the railway field

PiErDaWeiSi BaTuEr<sup>1</sup>, Liu Jie<sup>1,2\*</sup>

(1. School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756;2. Engineering Research Center of Sustainable Urban Intelligent Transportation, Ministry of Education, Chengdu 611756)

Abstract: An emergency management method for violations and risks of online customer service content in the railway field is proposed based on the Scenario-Response Model and Knowledge Graph. Firstly, Protégé is used to construct a Scenario Model and a Response Model, which respectively include Event, Form, and Risk Identification ontology, as well as Measure and Scheme ontology. Based on the preset Scenario Model, combined with customer service content violations and risk event data, the risk identification of the event is carried out, and the Form individual of the event and the Risk Identification individual corresponding to the Form individual are obtained; based on the preset Response Model, combined with the Form individual of the event, the corresponding measure and Scheme individuals for customer service content violations and risk event are obtained. Knowledge is then imported into Neo4j to storage, retrieval and application. Finally, the method is verified with a case of image infringement on the customer service platform. By establishing a Scenario-Response Model and considering the overall correlation between the event scenario and response, the prepared response measures and plans are accurate and reliable.

Keywords: emergency response; scenario-response; online customer service content violations and risks; scenario model; response model 文章编号:1007-1423(2023)02-0110-07

DOI: 10.3969/j.issn.1007-1423.2023.02.017

## 基于区块链和"时间银行"的互助式服务平台的设计和开发

向佳欣, 王宏杰, 梁桂萍, 赖沛鑫

(广州华商学院数据科学学院,广州 511300)

摘要:为缓解社会养老压力,以"时间银行"模式作为新型养老模式对传统养老模式进行补充,而在 现有的"时间银行"系统中,时间币存储在一个中心化的服务节点上,这种方式不仅存在数据丢失、篡改 的风险,还存在时间币的发行和流通都缺乏透明度的弊端。对此,基于区块链技术与"时间银行"概念, 对新型互助式服务平台进行设计,该平台以政府为主导,以"时间银行"为媒介,借助"时间币"进行价 值衡量,有效地运用区块链去中心化、不可篡改、可追溯等特点解决了传统"时间银行"系统中存在的问题,从而助推国家养老发展。

关键词:养老;时间银行;时间币;区块链;智能合约

## 0 引言

近年来我国社会老龄化加重,传统的社区养 老模式<sup>[11</sup>已不能满足我国的养老需要,需寻求新 的养老模式作为补充。"时间银行"顾名思义是 储蓄时间、提取时间的银行。一般表现为,需要 帮助的老年人在平台上发布需求,志愿者接单后 为其提供服务,以此获得一定的时间币,未来志 愿者有需要时可以通过支付时间币为自己或者亲 人换取养老服务,简单来说是用现在的服务时间 换取以后的服务时间。时间银行通过一种将服务 时间转换成时间币的方式来鼓励公众参与志愿者 活动,提高人们对志愿服务的积极性。

现今中国部分地区已开展试行时间银行的 志愿服务,然而在开展过程中存在一定的缺陷 和不足,以时间币为例,现有的时间银行在时 间币的发行和流通方面缺乏透明度,用户无法 了解时间币的发行量。此外,时间币交易数据 存储在一个中心化的数据库中,一旦数据库受 到攻击,就会造成数据库损坏,导致数据丢失、 篡改,甚至产生数据无法恢复的严重后果。区 块链是一种新型去中心化协议,由不同节点组 成,它不依赖第三方机构的审核,可以自动执 行智能合约,实现分布式存储时间币交易信息 或其它数据,保证数据的安全性<sup>[2]</sup>。利用区块 链技术可以使时间币的发行和流通公开透明, 时间币的结算不依赖某个中心化的节点。

针对上述情况,本文提出将区块链技术与 "时间银行"概念相结合的互助式服务平台。平 台以政府为主导,以"时间银行"为媒介,借 助"时间币"进行价值衡量。同时,应用区块 链去中心化管理数据实现时间币的结算流通, 促进服务供需双方完成服务交换。

平台分为用户端和管理端。用户端主要采用 uni-app框架完成前端开发,使用Web3.js和智能 合约作为接口连接以太坊区块链;而管理端主要 利用Thymeleaf模板完成前端开发。二者均利用 Spring Boot框架及Spring Security框架完成后端开 发,Solidity编写智能合约连接服务端和区块链,

基金项目:广东省科技创新战略专项资金("攀登计划"专项资金)(pdjh2022b0645)

作者简介:向佳欣(2002—),女,广东广州人,学士,研究方向为软件系统与技术、物联网;王宏杰(1982—),男,湖北 襄阳人,硕士,讲师,研究方向为人工智能、深度学习;梁桂萍(2001—),女,广东江门人,学士,研究方向为软件系统与技 术、大数据与数据科学;赖沛鑫(2002—),男,广东东莞人,学士,研究方向为软件系统与技术、软件工程服务与应用

收稿日期: 2022-09-09 修稿日期: 2022-10-19

进行数据交互,用MySQL完成数据存储。

## 1 技术基础

#### 1.1 uni-app和Thymeleaf

uni-app是一个基于 Vue.js开发所有前端应用的框架,实现一套代码可发布到 iOS、Android、Web 以及各种小程序(如微信、支付宝、百度等)、快应用等多个平台<sup>[3]</sup>。

Thymeleaf 是一个能够处理HTML、JS、XML 并服务于 Java 的模板引擎,它可用于 Web 与非 Web 环境中的应用开发。使用 Thymeleaf 可以提 高平台开发的速度,实现完美的页面渲染<sup>[4]</sup>。

1.2 Web3.js

Web3.js 是一个 JavaScript 库集合,它封装 了以太坊的 JSON RPC API,可以提供一系列 Javascript 对象和函数与以太坊区块链交互,包 括查看网络状态,检索用户账户,发送交易, 与智能合约交互等<sup>[5]</sup>。

## 1.3 Spring Boot和Spring Security

Spring Boot是Spring项目的一个子项目,它由 Pivotal团队提供,其样板化的配置可以精简Spring应用的初始搭建以及开发过程,使项目开发变得简单、方便和快捷<sup>[6]</sup>。

Spring Security 是基于企业级应用系统,为 应用程序开发提供声明式安全访问控制的框架。 它对安全性的支持依赖于 Servlet 过滤器。过滤 器可对进入的请求进行安全检测,检测通过后 应用程序才会对该请求进行处理<sup>[7]</sup>。

## 1.4 Solidity

Solidity是Ethereum的一种契约型编程语言,属静态类型,支持继承库和用户自定义类型以

及其他功能,旨在定位到以太坊虚拟机。

## 1.5 MySQL

MySQL由 MySQL AB公司开发,是一种关 系型数据库管理系统<sup>[8]</sup>,与其他数据库将所有 数据放在一个中央仓库不同,关系数据库将数 据保存在不同的表中,从而增加了数据存储速 度并提高了数据查询的灵活性。

#### 2 系统设计

## 2.1 总体设计

与传统意义上的社会互助不同,新型互助服 务平台是基于"时间银行"概念开发的,由政府 加强政策引导,提高统筹层次,服务交换双方通 过线上交流,以"时间银行"为媒介,借助"时 间币"进行价值衡量,同时应用区块链去中心化 管理数据并实现时间币的结算流通,促进服务供 需双方完成服务交换,最终实现全社会所有群体 共同帮助有需要的老年人的互助服务平台。因此 平台服务主要设计了用户端和管理端。

用户端,利用uni-app、Spring Boot以及Web3.js 构建移动端,它主要是为供需双方提供服务交 换的线上平台,入驻平台的社会组织发布志愿 服务信息和已认证老年人发布个人需求,志愿 者接收需求,完成服务并获得时间币。同时, 利用区块链去中心化协议,使时间币的发行和 流通公开透明,时间币的结算不依赖某个中心 化的节点。

管理端,利用Thymeleaf、Spring Boot和Spring Security 实施,它是一个 B/S 结构的在线 Web 系统,管理员通过浏览器实现对平台的管理。

系统架构如图1所示。



## 图 1 系统架构图

## 2.2 功能设计

平台主要为用户端和管理端,用户端和管 理端又分别包含若干个功能模块,如图2所示。



#### 图 2 系统功能模块

用户端按时间币交易流程分为发布需求、 接收服务、服务资讯、兑换商城以及家庭通讯 录等五个模块。

(1)发布需求。发布需求形式有两种:第 一种为政府主导,政府鼓励社会组织入驻平台 并在平台上发布志愿服务信息,同时发放时间 币;第二种为个人主导,已在平台认证的老年 人,发布个人需求。在需求者发布需求前,平 台会收集需求者的服务地点、服务时间、服务 时长、服务内容等具体信息,以便平台为志愿 者提供具体的服务信息。

(2)接收需求。"接服务"分为志愿服务队 发布的志愿服务和个人认证后允许发布的需求, 用户可以选择需求类型并报名,为了提高报名 的通过率,志愿者需要在申请过程中填写详细 的申请信息,通过审核后才算报名成功。报名 后按照规定进行服务,志愿者服务完成后经过 审核平台认证后获得相应时间币,如果未能按 照要求完成任务或超时完成,平台会扣除志愿 者的积分。

(3)服务资讯。平台主页设有关于公益宣 传的咨询,同时提供社区论坛功能,志愿者和 被服务群体均可在论坛里进行交友,分享心得 体会,丰富平台的娱乐功能。

(4)兑换商城。为鼓励志愿者进行志愿服务,本系统设置了兑换功能。在兑换商城中,用户可利用时间币进行物品兑换。

(5)家庭通讯录。在家庭通讯录功能中, 用户可以邀请自己的朋友和亲人加入,形成好 友链,好友链间用户可以共享时间币。

管理端根据时间币交易流程,分为账户管 理、存取管理、系统信息管理、查询管理以及 基本业务管理等五个模块。

(1)账户管理。管理员可以对用户的相关 信息进行管理,包括志愿服务队和个人的账户 开户、账户信息修改和挂失与销户等功能。

(2)存取管理。主要包括账户中时间币的 存与取,系统发放用户时间币,以及家人账户 之间的时间币共享和转发,商城兑换的时间币 交易等功能。

(3)系统信息管理。主要包括社区活动信息、用户信息、活动额度设置、商品兑换展示和数据备份等管理。也包括账户查询、导出、打印等功能。

(4)查询管理。主要包括查询社区居民的 账户信息、活动申请和完成信息、捐赠信息等 功能。

(5)基本业务管理。基本业务包括商品管理、需求管理、消息管理和资讯管理等。主要是对用户的数据进行读取、新增、修改、保存和删除等操作。

#### 2.3 数据库设计

本系统采用 MySQL 作为关系型数据库系统,按照数据库设计原则,在服务器的 MySQL 数据库中设计用户信息表(<u>用户id</u>、用户姓名、 用户密码、用户电话、用户类型、账户余额等)、 新闻资讯表(<u>新闻id</u>、新闻名、新闻作者名、发 表时间、新闻图片、新闻详情等)、兑换商城信 息表(<u>商品id</u>、商品名称、商品图片、商品价格、 商品数量等)、活动需求表(<u>需求id</u>,需求名称、 需求类型、需求时间、需求备注等)和时间币交 易记录表等多个数据表,如表1所示。

表 1 数据库表设计

数据表名称	说明
user	用户信息表
news	新闻资讯表
shop	兑换商城信息表
order	活动需求表
bill	时间币交易记录表

## 2.4 区块链设计

为保证用户时间币交易安全,用户端需要 把提交到服务端的数据分为个人信息、服务信 息和时间币交易信息等,服务端通过消息摘要 算法运算得到摘要信息,通过Web3.js与以太坊 节点进行连接,将摘要信息提交到区块链存储, 同时用户端将信息存储到本地数据库。管理端 定期检查数据并判断信息是否被篡改,管理端 首先发出请求,通过智能合约查询从而获取区 块链中存储的信息,然后与本地数据库中存储 的信息进行对比。信息一致则证明数据没有被 篡改,信息不相同就表明数据存在被篡改的可 能性。

#### 2.5 智能合约设计

智能合约以Solidity为开发语言,在以太坊 区块链环境下,主要完成客户端到区块链端摘 要信息的上传,以及管理端到区块链端摘要信 息的查询。智能合约中使用 mapping 存储数 组<sup>[9]</sup>,设计userInfo、serviceInfo和billInfo三个变 量。userInfo存储用户个人信息数组,其中user 为用户对象,包含用户名和用户的以太坊地址; serviceInfo存储用户服务信息的摘要;billInfo存储 用户的时间币交易信息的摘要,具体如表2所示。



#### 图 3 区块链设计方案

#### 表 2 智能合约变量设计表

变量名称	数据类型	描述
userInfo	struct user{ string username; address Raddess;} mapping(address=>user) private userInfo	包含用户名、以太坊地址。userInfo为存储 用户简历信息数组
serviceInfo	<pre>mapping(address=&gt;mapping(string=&gt; string)) private serviceInfo</pre>	tableInfo存储简历信息内容数组,分别为: 用户以太坊地址、时间币交易信息名
billInfo	mapping(address=>mapping(string=> string)) private billInfo	billInfo存储简历信息内容数组,分别为: 用户以太坊地址、用户服务信息名

## 3 系统开发

## 3.1 开发过程

平台的原型设计、功能设计以及针对各个 操作交互设计均基于"时间币"交易流程开发, 时间币开发以政府公信力为背书,同时需要建 立一个审核平台对时间币进行审核认证,整个 过程如下:

(1)时间币的发行。"时间币"由政府主导 发行,平台以各政府单位为全节点,以用户为 轻节点,构建公有链网络;以政府公信力为背 书,为志愿者发放时间币,主要表现在政府引 导各大志愿组织在平台上发布志愿服务内容, 志愿者参加服务,完成服务后由审核平台进行 审核,审核完成后向志愿者支付相应时间币<sup>[10]</sup>。

(2)时间币的交易。平台允许经过审核的个 人用户在平台上发布需求,主要表现在老年人 向平台提供相关证明材料,审核平台对证明材 料进行审核,审核完成后即可发布个人需求。 依据个人需求,志愿者向有服务需求的老年人 提供相应的服务。审核平台在服务完成后将会 对整个服务过程进行审核。服务通过审核后, 结合被服务者的满意程度以及服务类型等综合 因素,平台将服务时间进行服务资产数字化, 即将服务时间转换为等值的时间币。

(3)时间币的转让。平台用户可以邀请自己 的朋友和亲人加入,形成好友链,用户在好友 链中进行账号关联,关联后的账号平台可以实 现时间币共享。

(4)时间币的支付。平台设置线上商城,政 府与商家达成合作,商家通过平台宣传产品, 并为平台提供产品赞助,用户可以在商城页面 换取赞助商的产品。

## 3.2 用户端

(1) 用户登录。iOS 和 Android 系统中,点 击首页下方"我的"按钮进入用户个人界面, 点击头像或未登录进入登录界面,用户输入手 机号和密码,点击登录。小程序中,点击首页 下方"我的"按钮进入用户个人界面,点击头 像或未登录,平台跳转至微信授权界面,用户 点击允许,"我的"界面即可获取用户信息,如 图4所示。



图 4 登陆界面

(2)发布需求。用户在首页选择"发需求"进入页面填写表单,其中,用户可点击"服务类型选择"进行修改,如图5所示。

••••• WeChat?	13:46	50% 💷•
<	发布需求	••• •
服务类型选择		服务类型选择 >
内容 请描述	些一下你的大概需求	
图片上传		0/4
<u>i</u>		
请输入详细需求	多行文本输入框	
日期选择		2022-5-12 >
时间选择		12:01 >
	提交表单	

图 5 发布需求表单

(3)接收需求。用户在首页选择"接收服务"进入界面选择服务,其中用户可以在页面上方进行服务类型、时间币排序、时间排序的筛选,选择完成后进入详情页报名,系统生成活动需求信息上传至区块链存储以及保存到本地数据库,服务完成后志愿者获得相应时间币,同时时间币交易信息上传至区块链和本地数据库存储,如图6所示。



#### 图 6 接收需求界面

(4)兑换商城。用户点击首页下方"我的" 按钮进入用户个人界面,选择兑换商城,点击 商品进入商品详情页兑换,如图7所示。



图 7 兑换商城界面

(5)家庭通讯录。用户点击首页下方"我 的"按钮进入用户个人界面,选择家庭通讯录, 点击邀请按钮跳转至微信邀请界面,点击同意 关联即可绑定账号,同时将两者时间币余额进 行相加后作为新的账户余额数据,修改用户信 息,同时形成时间币交易信息并上传至区块链 和本地数据库存储,如图8所示。



图 8 家庭通讯录界面

## 3.3 管理端

管理端是利用 Spring Boot 和 Thymeleaf 实施的在线系统,包括账户、存取、信息、查询、基本业务等各种管理功能,由于篇幅原因,其实施过程在此不展开论述。

#### 4 结语

互助式服务平台把区块链技术与"时间银 行"概念结合起来,以政府为主导,以"时间 银行"为媒介,借助"时间币"进行价值衡量, 同时应用区块链去中心化管理数据并实现时间 币的结算流通,最终促进服务供需双方完成服 务交换。本平台的设计与实现,可以有效解决 社会老龄化过程中关于"时间银行"概念的新 型服务模式探索过程中的问题,为促进我国老 年人链接多元的社会资源,为我国养老公益志 愿服务常态化、规范化、机制化发展作出贡献。

#### 参考文献:

- [1] 樊新华,郑红.基于时间银行养老系统的设计与实现[J].智能计算机与应用,2019,9(2):73-77.
- [2] 肖凯,王蒙,唐新余,等.基于区块链技术的公益时间银行系统[J]. 计算机应用,2019,39(7):2156-2161.
- [3] 谢志妮.基于uni-app的微信小程序关键技术运用[J].电子技术与软件工程,2021(12):32-33.
- [4] 刘子凡,郭昱君.基于 SpringBoot+Mybatis 的个人 博客系统设计与实现[J].现代信息科技,2021,5
   (8):104-107,111.
- [5] 贾民政.基于区块链的投票系统的设计与实现[J]. 北京工业职业技术学院学报,2018,17(2):25-29.
- [6] 单智峰.论《招生录取管理系统》开发设计与实现[J].通讯世界,2017(4):203-206.
- [7] 李伟宏.面向医疗信息管理系统的云存储技术研究 与实现[D].广州:华南理工大学,2012.
- [8] 杨永明,徐海霞. RFID的图书馆自助借阅系统的身份识别实现[J]. 物联网技术,2022,12(3):74-76.
- [9] 宋东翔,林海威,王怡然,等.基于区块链的微信小程序简历设计与实现[J].电脑编程技巧与维护, 2022(4):80-82.
- [10] 王思泓,冯玲云,耿宜帅,等.基于志愿者服务系统
   的时间银行设计与实现[J].现代信息科技,2020,4
   (6):107-109.

(下转第120页)

文章编号:1007-1423(2023)02-0116-05

DOI: 10.3969/j.issn.1007-1423.2023.02.018

# 私有云平台数据云上云下备份体系设计

孙建刚\*, 高 颖, 杨庆甫, 常雨竹, 董耀众, 李伟良

(国家电网有限公司信息通信分公司,北京 100761)

摘要:云平台作为大型企业数字化转型的基础设施,平台的安全、稳定是保障业务连续性的基础。目前数据备份仅存储在平台,而且备份策略不规范。因此,结合电力行业私有云平台提出云平台云上、云下备份体系。参考电力行业相关要求,对数据进行分级分类,对相应级别数据的数据存放、备份方式、备份频率等进行设计,并提出恢复及安全需要考虑的内容。

关键词:私有云;备份体系;电力行业;数据分级

## 0 引言

随着云计算技术在电力行业的广泛普及, 电力行业主体单位依托主流的互联网云技术, 构建了新一代信息基础设施,用于承载服务于 各类业务的数字化转型应用。因此,保障云平 台自身的安全、稳定是提高信息系统连续运行 能力和业务持续运营能力的基础,也是防范数 据资产泄露、丢失、破坏或不正当使用,进而 充分挖掘资产使用价值的充分保障。

近年来开展了云平台的容灾建设,通过保 障云平台的高可用,从而实现云上业务的连续 性。而容灾建设仅是主备环境之间采用数据同 步方式,无法应对数据丢失与误操作等场景。 同时,也无法在发生机房故障、地域性自然灾 害等场景时实现对核心数据的保护。因此需要 开展平台级的数据备份,但是目前各行业均没 有对云平台的备份设置的标准,无法进行有效 参考开展实施工作。有鉴于此,本文对云平台 备份的关键内容进行了一个体系化的梳理与设 计。希望能够为电力行业、政府、金融以及其 他采用互联网云技术搭建的私有云平台,提供 可供参考借鉴的理论方法。

平台备份体系设计主要包括数据分级分类、 数据备份策略设计、数据恢复及安全策略设计、 数据备份验证四部分。

## 1 数据分级分类

#### 1.1 云上业务

平台组件备份的最终目标是保障云上业务 的连续性,组件的备份恢复情况,需要结合业 务备份恢复要求。参照行业的整体要求,将云 上业务划分为一类、二类、三类系统。在此基 础上,进一步结合《信息安全技术网络安全等 级保护定级指南》,按照信息系统的重要程度进 一步划分为等保三级、等保二级、等保一级系 统,如表1所示<sup>[1]</sup>。

参照行业的整体要求,其中一类系统RTO<8 小时,二类系统RTO<12小时,三类系统RTO<24 小时,所有系统RPO为24小时。在此基础上, 考虑到等保三级系统的重要性,将等保三级系 统的RPO<24小时要求提升为RPO<12小时,如 表2所示。

**收稿日期:** 2022-07-16 修稿日期: 2022-08-23

作者简介:\*通信作者:孙建刚(1992—),男,北京人,硕士研究生,研究方向为云平台、数据中台,E-mail: 749628618@qq.com;高颖(1994—),女,山西大同人,硕士研究生,研究方向为信息系统运维;杨庆甫(1995—),男,河南焦 作人,硕士研究生,研究方向为云平台、数据分析;常雨竹(1994—),女,北京人,硕士研究生,研究方向为技术中台;董耀 众(1989—),男,河北石家庄人,硕士研究生,研究方向为智能运维、大数据分析;李伟良(1979—),男,北京人,硕士研究 生,研究方向为云平台、数据中台、电力系统

#### 表 1 业务系统分类分级定义

分类	定义
一类系统	受政府严格监管、纳入国家关键信息基础设施 的信息系统
二类系统	受到政府一般监管、服务于公司特定用户、对 公司生产经营活动有一定影响的信息系统,或 者服务于公司全体员工、直接影响公司业务运 作的信息系统
三类系统	除一、二类以外的其他信息系统
等保三级	数据丢失后,对国家安全造成一般损害,且对 社会秩序、公共利益造成严重损害
等保二级	数据丢失后,对社会秩序、公共利益造成一般 损害,且对公民、法人和其他组织的合法权益 造成严重损害与特别严重损害的系统
等保一级	对公民、法人和其他组织的合法权益造成一般 损害的系统

表 2 业务系统备份要求 单位:小时

安坛八米	至弦八仞	备份指标要求	
矛机刀矢	杀犯刀级	RTO	RPO
一类系统	三级系统	<8	<12
	二级系统	<8	<24
	一级系统	<8	<24
二类系统	三级系统	<12	<12
	二级系统	<12	<24
	一级系统	<12	<24
	三级系统	<18	<12
三类系统	二级系统	<18	<24
	一级系统(未定级)	<18	<24

#### 1.2 云平台组件

为满足多业务、多场景、海量数据、数千 级的应用建设需求,云平台进行模块化的设计, 基于技术路线与框架结构,构建了不同用途的 平台组件。根据组件数据丢失对平台和业务的 影响范围、影响程度划分一级组件、二级组件、 三级组件,如表3所示<sup>[2-3]</sup>。

按照组件功能,将组件运行过程中产生的 数据划分为成元数据、配置数据、日志数据。

元数据:存储了运行该组件的基本信息, 通常用来描述该组件包含的其他数据的结构、 存储位置、访问权限、用户结构等。丢失会影 响平台或业务可用性。

配置数据:存储了该组件各个进程的环境、 调用文件等,用来支撑各个进程的正常启动与 运行。丢失会影响平台或业务可用性。 日志数据:存储了该组件运行过程中产生的所有操作事件、状态、告警、故障原因等。 丢失会影响平台或业务的优化<sup>[4]</sup>。

考虑到云平台故障以及元数据、配置数据 丢失的影响,参考一类系统的RTO和RPO指标 要求设计,将一级、二级组件中的RPO<24小时 要求提升为RPO<12小时,如表4所示<sup>[4-5]</sup>。

表 3 组件数据丢失影响内容

八米	合き	数据丢失影响		
万尖	定义	平台	业务	
一级组件	组件重要数据丢失导 致业务中断、无法运 行、访问缓慢。包括 断网、无法访问实 例、实例数据丢失、 安全防护失效等	平台无 法创建 实例等	影响业务系统 功能,系统完 全不可用式或部 分功能出现问题,或系统性 能出现问题	
二级组件	组件重要数据丢失导 致云平台相关日志无 法访问、运维与巡检 能力缺失。业务故障 无法定位	平台无 法创建 实例等	影响业务系统 功能故障处 理、性能优化	
三级组件	组件重要数据丢失导 致管理平台无法访 问、管理机制缺失, 但不影响现有实例服 务。包括无法新建、 更新策略与配置、无 法进行资源监控与统 计等	平台无 法创建 实例等	无影响	

#### 表 4 业务系统备份要求 单位:小时

加快人动	新也以未	备份指	标要求
组件分级	致惦 <u></u> 尔尖 -	RTO	RPO
一级组件	元数据	<8	<12
	配置数据	<8	<12
	日志数据	<8	<24
二级组件	元数据	<8	<12
	配置数据	<8	<12
	日志数据	<8	<24
三级组件	元数据	<8	<24
	配置数据	<8	<24
	日志数据	<8	<24

## 2 数据备份策略设计

备份策略是备份工作开展的基础,在进行

备份策略设计时,需要明确备份数据存放的具体地点、备份方式、备份频率、备份窗口,以及具体的保留周期<sup>[5-6]</sup>。

## 2.1 存储方式

备份数据存放方式有本地云上、本地云下、 异地云上三种,如表5所示。

表 5 备份数据存放方式

名称	定义	适用场景
本地云上	将备份数据保留 在云平台	仅用于防范云上业务或云组件 自身的逻辑故障,无法防范云 平台整体性故障的风险
本地云下	将备份数据保留 在云平台之外的 本地备份介质中	可用于防范云上业务或云组件 自身的逻辑故障、以及云平台 的整体性故障,无法防范数据 中心级别的故障
异地云上	将备份数据保留 在异地数据中心 备份介质中	用于防范云上业务或云组件自 身的逻辑故障、云平台的整体 性故障、以及数据中心级别的 故障

#### 2.2 备份方式

根据备份对象与数据访问特性,将备份方 式分为完全备份、增量备份、差异备份,如表6 所示。

表 6 备份方式

备份方式	特性	适用场景
完全备份	完全备份是增量备份与差 异备份的基础,完全备份 的恢复效率最高。但大数 据量情况下,备份窗口时 间较长	备份对象整体数据 量较小,备份过程 不会对备份对象产 生负载影响,完全 备份窗口时间较短 (如1小时内)的场景 或者备份对象的数 据变化频率不高的 场景
增量备份	增量备份的恢复需要结合 最近一次的完全备份与期 间所有的增量备份,恢复 效率次于完全备份与差异 备份	备份对象整体数据 量较大,且增量数 据较大的场景
差异备份	差异备份的恢复仅需要结 合最近一次的完全备份。 恢复效率次于完全备份但 优于增量备份	备份对象整体数据 量较大,但增量数 据较少的场景

## 2.3 备份频率

基于系统分级分类中的PRO 指标要求,设

置合理的备份频率。例如生产系统的PRO指标 要求为24小时,应当至少每24小时执行一次备 份。对于变化频率不高的数据(如软件配置类数 据),可在发生变更前后各执行一次备份<sup>[7-8]</sup>。

## 2.4 备份窗口

备份窗口即发起备份作业的具体时间或执行 一次备份作业所需的时间范围。应结合生产系统 的业务活动特性进行备份窗口的设置,避免在业 务活动频繁的时间段发起备份,如表7所示。

表 7 备份窗口评估与建议

评估因素	约束条件	备份窗口建议
网络带宽	分析备份对象所在环境的网 络带宽使用情况,避免在网 络带宽占用较大时执行备份 作业	结合备份对象的业 务活动时间特性进
性能负载	分析备份对象所在环境的性能负载情况,避免在性能负载情况,避免在性能负载转高时执行备份作业。常见的负载指标包括CPU、内存、磁盘IO	行设计 通常可设置为晚上 22:00—6:00

## 2.5 保留周期

应结合备份对象的业务活动特性与访问需 求设置备份数据的保留周期。保留周期的设计 需要考虑到备份对象的最长保留期限,以及备 份存储介质的空间占用情况。对于备份数据访 问频率较低但不能进行删除的,可根据需要转 储至离线存储介质中长期保留<sup>[9-10]</sup>。

## 2.6 备份策略

云组件备份策略如表8所示。

#### 表 8 云组件备份策略

组件 分级	数据分类	存放方式	备份 方式	保留周 期/天	备份 窗口
一级 组件	元数据	本地云下备份+ 异地云下备份	完全+ 差异	30	晚上 22:00— 06:00
	配置数据	本地云下备份+ 异地云下备份	完全+ 差异	30	
	日志数据	本地云下备份	增量	7	
二级 组件	元数据	本地云下备份+ 异地云下备份	完全+ 差异	30	
	配置数据	本地云下备份+ 异地云下备份	完全+ 差异	30	
	日志数据	本地云下备份	增量	7	
三级 组件	元数据	本地云下备份	增量	7	
	配置数据	本地云下备份	增量	7	
	日志数据	本地云下备份	增量	7	

## 3 数据恢复及安全策略设计

## 3.1 恢复策略

数据备份的目的是进行有效的恢复,为了 保障国网云数据恢复过程的安全与有效,本章 节主要针对数据恢复时的注意事项与恢复方式 提出基本要求。

在进行数据恢复时,需要明确以下内容:

(1)明确各类数据的恢复对象与适用场景。

(2)确认数据的恢复时间点满足业务访问与 运行需求。

(3)发布检修/停机窗口时,数据恢复所需 的时间应包括数据传输的时间,以及数据恢复 后进行配置的时间。

(4)采用最小化恢复原则,尽量避免恢复与 业务运行无关的数据,能够选择部分数据恢复 时,不要进行完全恢复。

(5)数据恢复时需要考虑恢复对象之间的关 联性与优先级,并设置分步恢复策略。

(6)数据完成恢复与配置后,需要进行灰度 发布确认数据一致性与有效性。

## 3.2 安全策略

在进行数据的备份与恢复期间,需要基于 安全策略确保生产系统(云上业务与云平台组 件)、备份系统、备份数据传输、备份数据存储 等维度的安全,如表9所示。

安全维度	安全策略要求			
生产环境安全 (テトル冬与平	备份作业的执行不能对生产环境产生明显的性能负载,包括CPU、内存、磁盘IO			
台)	备份作业的执行不能争抢业务活动的网络 带宽			
タ川环培立人	备份系统的元数据(备份作业的索引信息) 应实现定期备份			
留切坏現女王	备份系统应具备冗余特性,支持对执行中的备份作业进行故障转移			
	备份数据的传输应经过转码或加密处理			
粉捉住绘字会	备份网络应具备冗余特性			
<b>奴</b> 加 [2] 相 女 主	跨数据中心的备份网络应采用专线或VPN 通道			
	备份数据的存储应经过转码或加密处理			
数据存储安全	备份介质应具备冗余特性,支持备份数据 的冗余存储			

表 9 备份安全策略设计

## 4 结语

在行业数字化转型的背景下,电力行业稳 定快速发展的需求对数据备份的管理提出了更 高、更全面的要求,数据备份的范围不单单要 涵盖信息系统,对于底层的基础设施也有同样 的要求。本文重点针对私有云备份这一课题, 提出了体系设计理论研究内容,填补了行业内 该领域的空白。希望这些理论能够为其他行业 的私有云的数据完整性建设提供有价值的参考。

#### 参考文献:

- [1] 陈嘉升,黄瑶.云计算技术在计算机网络安全存储
   中的运用研究[J].数字通信世界,2022(11):84-86.
- [2] 卢凯,包伟华.企业信息数据安全保护的研究和对 策[J].信息化建设,2022(6):59-61.
- [3] 庞凯中.信息化时代下计算机网络安全问题的探讨[J].网络安全技术与应用,2022(6):155-157.
- [4] 李英豪,杨春,李文华,等.基于"两地三中心"模式 的大数据平台异地灾备研究[J].长江信息通信, 2022,35(3):171-173.
- [5] 薛俊海,李晋泰,张承,等.大数据技术在计算机信息安全中的应用研究[J].网络安全技术与应用, 2022(2):70-71.
- [6] 刘强.档案数据时代:国内研究现状及展望[J].办 公自动化,2022,27(2):40-43.
- [7] 王铠宁.行政事业单位网络安全建设研究[J].辽宁
   师专学报(自然科学版),2021,23(4):28-30,74.
- [8] 梁猛.基于大数据的计算机网络与信息安全策略[1].电子技术,2021,50(12):136-137.
- [9] 王梅.工作电脑文件备份可靠性和安全性的解决方案[J].电子技术与软件工程,2021(24):259-260.
- [10] 赵永华. 应用混合云备份的探讨[J]. 网络安全和信息化,2020(5):80.

## The private cloud platform data cloud backup system design

Sun Jiangang\*, Gao Ying, Yang Qingfu, Chang Yuzhu, Dong Yaozhong, Li Weiliang

(Information and Communication Branch of State Grid Co., Ltd., Beijing 100761)

Abstract: As the digital transformation infrastructure of large enterprises, the security and stability of the cloud platform are the basis for ensuring business continuity. At present, data backup is only stored on the platform, and the backup strategy is not standardized. Therefore, this paper proposes the cloud platform on cloud and under cloud backup system based on the private cloud platform of the power industry. According to the relevant requirements of the power industry, the data classification, the data storage, backup mode and backup frequency of the corresponding data are designed, and the contents that need to be considered for recovery and security are put forward.

Keywords: private cloud; backup system; electric power industry; data classification

(上接第115页)

# Design and development of mutual assistance service platform based on blockchain and "time bank"

## Xiang Jiaxin, Wang Hongjie, Liang Guiping, Lai Peixin

(School of Data Science, Guangzhou Huashang College, Guangzhou 511300)

Abstract: In order to relieve the pressure of social pension, the "time bank" model is used as a new pension model to supplement the traditional pension model. In the existing "time bank" system, the time dollars are stored on a centralized service node. This method not only has the risk of data loss and tampering, but also has the disadvantage of lack of transparency in the issuance and circulation of time currency. In this regard, a new type of mutual aid service platform is designed based on the blockchain technology and the concept of "time bank". The platform takes the government as the leading role, "time bank" as the medium, and uses "time dollars" to measure the value. It effectively uses the characteristics of blockchain decentralization, tamper proof, traceability, etc. to solve the problems in the traditional "time bank" system, thus boosting the development of the national elderly care.

Keywords: pension; time bank; time dollars; blockchain; smart contract