

M 现代计算机

XIANDAI JISUANJI
第29卷第11期（总第779期）
半月刊（1984年创刊）
2023年6月10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 梁嘉璐
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
电子邮箱 tougao@moderncomputer.cn

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



邮局订刊二维码



现代计算机
官方网站二维码



M 现代计算机

第29卷 第11期（总第779期）

2023年6月

2023年6月 第29卷
第11期（总第779期）



ISSN 1007-1423
CN 44-1415/TP

MODERN COMPUTER

现代计算机



中山大学出版社 主办

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。

目次

研究与开发

基于CenterNet的草原牛羊计数研究	孙 强, 郝 敏	(1)
结合加权核范数与3D全变分的目标检测	班 颖, 田 韵, 邵泽军	(9)
基于深度强化学习的配件库存决策研究	李雨情, 潘超凡	(16)
基于Transformer和CNN的真实场景下植物病害识别方法	刘 畅, 莫海芳, 马 春	(22)
基于粒子群算法的江西景点自驾游路线规划	宋艳艳, 李广宏, 朱 玲, 夏传林, 郑 巍	(28)
基于图注意力网络的突发事件触发词抽取	陈彦杰, 廖 涛	(33)
软件可测性和测试用例效率的研究综述	王 东	(38)
基于WOA-RBF的温室温度预测研究	李其操, 董自健	(43)
基于EEMD-KELM的淮安市二河溶氧量预测研究	陈 肖, 张 强, 姚年春	(47)
基于极限学习机算法的农业绿色智慧评价系统	王 琦	(52)
列生成算法的改进策略研究	罗凤娥, 张 鑫, 赵 强, 杨思瀚	(56)

实践与经验

一种基于功能点的项目资源日历计算方法	杨 帆, 刘雅莹, 唐海涛	(60)
基于区块链的电子病历访问控制方法研究	刘荣管, 王兆栋, 邓兰华, 陈长源, 陈佩江	(64)
BP神经网络在测控设备自动调平系统中的应用	薛缠明, 赵翔宇	(68)

开发案例

研究生业务费预算与支出管理系统设计与实现	张建宇, 徐怡彤, 张盈盈, 吕震宇	(71)
基于SpringBoot的物联网智能家居方案设计	陈 欣, 阮锦新, 邬依林, 张 谦, 陈勇昌, 熊 燕	(76)
基于区块链的公益捐赠系统研究	李保东, 高洪美	(84)
一种低功耗蓝牙MEMS振动传感器及其应用	方逸洲, 赵中瑞, 张青野, 张鹏程, 艾麦提·布拉丁, 曾佑轩, 王 珏	(90)
基于Android的防走失APP的设计与实现	周 漫, 赵嘉颖, 温毅娴	(95)
基于VizQL的汽车工业数据可视化系统的设计与实现	杨国丹	(99)
在线实验报告管理系统的设计与实现	欧阳习彪, 徐宝林, 蒋 谨	(104)
电子校历自动生成系统的设计与实现	刘 哲, 宋佳豪, 潘梦瑶	(109)
面向汽车旧件再制造组合赋权模糊综合评价模型	王华政	(113)
一种基于YOLO的口罩佩戴检测终端的设计	张 良, 陈周欢, 林耿名, 叶绮雯, 何镇轩	(117)

Modern Computer

(Vol. 29, No. 11; Jun. 10, 2023)

CONTENTS

Research and Development

CenterNet-based grassland cattle and sheep counting study	(1)
Object detection with weighted nuclear norm and 3D total variation	(9)
Research of inventory decision for auto parts based on deep reinforcement learning	(16)
Plant disease identification method in real scenarios based on Transformer and CNN	(22)
Self-driving route planning of Jiangxi attractions based on particle swarm optimization algorithm	(28)
Emergent triggers extraction based on graph attention network	(33)
The review of software testability and test cases effectiveness	(38)
Research on greenhouse temperature prediction based on WOA-RBF	(43)
Prediction of dissolved oxygen in Huaian Erhe River based on EEMD-KELM	(47)
Agricultural green intelligent evaluation system based on extreme learning machine algorithm	(52)
Research on improvement strategies for column generation algorithm	(56)

Practice and Experience

A method for calculating project resource calendar based on function points	(60)
Study on the access control method of electronic medical records based on blockchain	(64)
The application of BP neural network in the automatic leveling system of measurement and control equipment	(68)

Development Solution

Design and implementation of budget and expense system for graduate research-related subsidy management	(71)
Design of IoT smart home solution based on SpringBoot	(76)
Research on public welfare donation system based on blockchain	(84)
A low-power Bluetooth MEMS vibration sensor and its application	(90)
Design and implementation of Android-based anti-wandering APP	(95)
Design and implementation of data visualization system for automobile industry based on VizQL	(99)
Design and implementation of online experimental report management system	(104)
Design and implementation of electronic calendar automatic generation system	(109)
Weighting evaluation model for remanufacturing combination of used automobile parts	(113)
Design of a mask wearing detection terminal based on YOLO	(117)

研究与开发

文章编号: 1007-1423(2023)11-0001-08

DOI: 10.3969/j.issn.1007-1423.2023.11.001

基于 CenterNet 的草原牛羊计数研究

孙 强, 郝 敏*

(内蒙古农业大学机电工程学院, 呼和浩特 010000)

摘要: 对草原放牧牛羊数量快速准确的统计是进行科学放牧管理和草原草畜平衡分析的重要前提。随着市场对畜牧产品需求的增加, 内蒙古地区作为国内畜牧产品主要产地, 其草原放牧牛羊的数量逐年上升, 而目前以人工进行放牧牛羊数量统计工作存在难度大、效率低、统计准确度不稳定等问题。为实现牛羊数量的快速统计, 使用无人机进行航拍采集的草原放牧牛羊图像并建立数据集, 使用多尺度特征提取网络 Res2Net 提升 anchor-free 的 CenterNet 模型对航拍情景下放牧牛羊图像粘连小目标的检测效果, 加入 ECANet 注意力机制进一步提升模型性能, 并调节相关参数进行算法优化。研究表明, 该模型在无人机航拍获取的数据集中的类平均精度达到 88.22%, 体现密集目标检测效果的牛羊对数平均漏检率分别下降了 10% 和 11%。这一结果说明引入无接触式的计算机视觉技术在草原牛羊计数方面具有一定的价值。

关键词: CenterNet; 目标检测; 计数; Res2Net; 注意力机制

0 引言

内蒙古地区拥有大规模的畜牧养殖业, 是我国牛奶、牛肉、羊肉、羊绒的主要产地之一。随着经济快速发展, 国内市场对高蛋白的牛羊肉等需求逐年增加, 呈现出消费增幅快于生产增幅的态势^[1]。在内蒙古地区的牛羊饲养中, 草原放牧畜产品更受市场青睐, 但畜牧业发展要兼顾自然环境的保护, 放牧生产端需要定期统计放牧牛羊数量反馈给牧民和有关部门进行载畜量分析、生产调整和科学管理, 因此便捷准确地对同一放牧草原中不同放牧牲畜识别计数以折合为羊单位进行实际载畜量统计, 进行草畜平衡分析对保护生态平衡和促进畜牧业健康发展、促进草地可持续发展具有十分重要的意义和作用。

目前牲畜养殖中电子耳标使用广泛^[2], 在福利化养殖中主要针对奶牛这种长期养殖牲畜采用耳标用于记录牛的健康状况, 但这类接触

式的计数手段成本高, 不适用于短期养殖牲畜或草原自由放牧较多牲畜的场合。随着机器视觉的发展, 李琦等^[3]采用 YOLOv3 与 DeepSORT 算法相结合, 提出基于双线计数法的草原羊的自动计数, 但该方法需要在羊圈中预设检测设施, 不适用于大型草原。

随着智能化畜牧养殖业的发展, 结合航拍图像使用机器视觉技术代替人眼, 通过目标检测算法无接触进行目标场景下的牲畜识别和数量统计具有检测简单快速的优越性。伍天舟^[4]利用无人机航拍图像制作羊头数据集结合 SSD 算法进行放牧羊群的自动计数。赵建敏等^[5]提出基于无人机图像的羊群密集计数方法, 初步解决了羊群密度判断。杜永兴等^[6]采用标注点的方式标注羊只, 高斯核回归转换成标签, 提升了高密度场景下计数准确度。

但针对航拍图片中牛羊小目标粘连现象, 目前研究中以羊头为目标比以羊背作目标检测

收稿日期: 2023-02-16 修稿日期: 2023-03-14

基金项目: 内蒙古自然科学基金(2017MS0606)

作者简介: 孙强(1997—), 男, 安徽安庆人, 硕士研究生, 研究方向为数字化农牧业关键技术及设备; *通信作者: 郝敏(1982—), 女, 内蒙古包头人, 副教授, 工学博士, 研究方向为图像处理、模式识别、数字化农牧业关键技术及设备, E-mail: hhhthaomin@163.com

效果好,但航拍图像中存在羊头被遮导致漏检;设计不同大小检测框无法满足牛羊等形态各异的检测需求。本文利用深度学习算法研究适应性更强的特征提取网络,以羊背作为检测目标,引入 anchor-free 类的 CenterNet 关键点检测算法,针对小目标检测难同时对牛羊进行分类识别,引入多尺度的特征提取网络 Res2Net 提取更多特征^[7],再添加注意力机制优化模型,以快速准确地对牛、羊两种放牧牲畜进行识别与计数。

1 材料与方法

1.1 图像采集与实验平台配置

在内蒙古自治区乌兰察布草原,考虑到不同高度下待检测目标的分辨率,使用 DJI Mavic Air2 无人机在 20 m 和 25 m 两个高度下对放牧场景下的多个羊群和牛群在雪地、草原、人工养殖场等多个不同场景、早中晚等不同光照条件下分别进行图像采集,研究所获取的无人机可见光原图像的分辨率最高为 5472×3648 像素,航拍图像如图 1 所示。

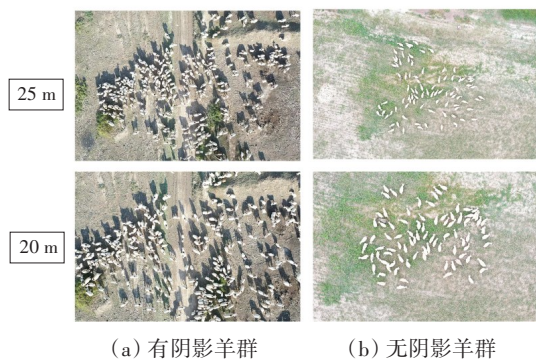


图 1 无人机航拍图像

本研究的深度学习实验平台的环境配置如表 1 所示。

表 1 深度学习实验平台搭建

项目	配置
操作系统	Ubuntu 18.04
GPU	NVIDIA GeForce RTX 3090
编程语言	Python 3.8
深度学习框架	PyTorch 1.7.1
计算架构	Cuda 11.1

1.2 构建数据集

为了提升模型的泛化性能,从拍摄目标与拍摄条件两方面着手来提升数据的丰富度。牛羊数量较大,且觅食喜欢群聚,识别准确率与图像所含目标数量相关,其识别难度体现在牲畜群聚时识别对象粘连程度,因此图像裁剪和数据集构建时考虑单张图像内包含的各类识别对象的类型、数量。另外手动剔除了一些由于图片过曝过暗、目标遮挡过多、拍摄角度倾斜的低质量图像。

1.2.1 图像预处理

考虑到深度学习方法需要足够数量训练集图像进行模型训练,但原图像数量有限且含有较多目标,对原图像的不同类型目标进行单个目标的整只裁剪和包含目标大约 20~400 个不同数量目标区域裁剪以扩充数据集,达到每张样本图像中都含有多个待识别的目标的基本要求。

整只裁切后的图片长宽比并不规则,算法模型需要输入固定长宽比的图像,一般的 resize 处理强行改变长宽比导致牛羊目标体型变形影响后续模型训练,如图 2 所示。

本研究引入等比例放缩的 letterbox 图像预处理输出特定尺寸的图片。由于裁剪图片的长宽比不同导致缩放填充后两端的灰度大小不同,灰度填充较多导致信息冗余影响推理速度,对 letterbox 的缩放策略进行改进,即:先计算原图像宽高与输入尺寸的缩放比例 R_w 和 R_h ,若是放大选取 $R=\min(R_w, R_h)$,若是缩小选取 $R=\max(R_w, R_h)$,再将原图像按 R 进行相应缩放;此时原图像宽或高里有一边与规定输入尺寸相同,没有达到输入尺寸的另一边计算与规定输入尺寸的差值,然后进行上下(或左右)的填充,实现了对原图像自适应的灰度填充,如图 3 所示。

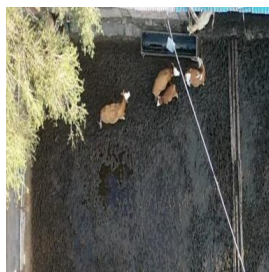


图 2 resize 效果图

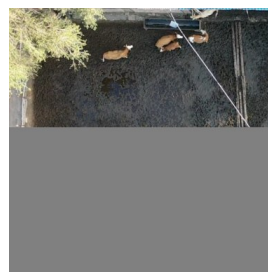


图 3 letterbox 效果图

1.2.2 自建数据集

制作的数据集包括 1000 张牛羊航拍图片，其中有 200 张来自互联网，800 张图片来自乌兰察布草原。由程序随机划分 100 张做测试集，剩余 900 张按照 9 : 1 的比例由程序随机划分训练集和验证集。表 2 是自建数据集所含目标数量分布情况。使用图像注释工具 LabelImg 对裁剪及缩放后的图像进行牛羊分类标注，通过最小外接矩形将样本中的牛、羊背部分别标出，记录该矩形的左上、右下的顶点坐标并分别设置类别标签为 cattle 和 sheep。

表 2 数据集的目标数量分布

目标类型	目标数量/只	目标图片数量/张
羊	<100	50
	100~200	400
	>200	150
牛	<50	50
	50~150	100
	>150	100
牛、羊	<50	50
	100~150	100

注：图片目标类型按照只含有牛、只含有羊、同时含有牛羊分类

2 载畜量计算模型

本次研究采集的航拍牛、羊目标像素占原图像总像素的比例小，属于小目标检测范畴。待检测牛、羊目标自身包含的像素信息过少，因而包含的判定性特征过少；由于小目标的 ground truth box 和 anchor 都较小，加上小目标对定位的要求较高，anchor 和 ground truth box 稍微偏移就会造成 IoU (intersection over union) 过低，导致该目标被模型判断为负样本^[8]。本实验图像中牛羊所占比例不同，针对性设计 anchor 无

法满足自建数据集同一类型目标所需 anchor 不同的要求，本研究使用不需要生成 anchor 的 anchor-free 模型。

2.1 CenterNet 模型框架

CenterNet 主要分为 Backbone、Decoder、Head 三个部分，将目标检测转换成中心点预测。经过 Backbone 提取初步特征，在 Decoder 中通过反卷积模块三次上采样得到更高的分辨率特征输入到 Head，通过预测目标点的偏移量 (Offset) 与宽高 (Width&Height) 来获取预测的矩形框，模型对于牛、羊都会产生一个单独的 Heatmap 图，对于生成的 Heatmap 图，如果某个坐标处含有牛、羊的中心点，则在该坐标处生成一个关键点。模型只需要从 Heatmap 中提取目标，不使用 NMS (non-maximum suppression)，降低了计算量，最后将 Head 的 3 个分支预测结果进行整合，得到目标的检测框和所属的类别^[9]。

2.2 模型的改进与提升——预防过拟合

本研究对于模型的改进与提升主要体现在防止小数据集训练过度拟合和提升模型对于牛羊计数的精度上。自建牛羊数据集图片较少，训练模型易产生过拟合。为获得准确的预测，使用下列三种方法预防过拟合。

2.2.1 运用迁移学习

迁移学习是先在可用的大型数据集上训练通用模型，然后在小型数据集上对其进行微调。本研究是对牛、羊的目标检测，模型在 VOC 07+12 数据集上预先训练，再对模型进行微调。比起从零开始使用小数据集训练的模型，使用大型数据集预先训练的模型更可能产生准确的预测^[10]。

2.2.2 模型选择与组合

组合模型通过个体模型预测值加权平均计算出最终预测值，提升模型泛化能力。为防止

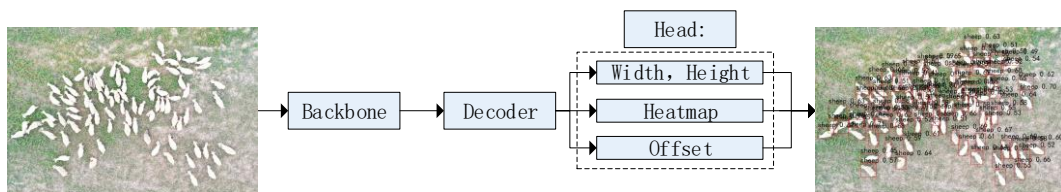


图 4 CenterNet 网络结构

过拟合, 限制权重和参数量排除掉可能具有非线性或特征交互的模型, 本研究使用 CenterNet 论文中平衡性最佳的 ResNet 类作为特征提取网络, 降低出现训练过拟合的可能。

2.2.3 平衡样本

小数据集中异常值对预测结果影响较大, 处理数据时需要识别并删除异常值^[11], 而特殊数据不足导致数据种类不平衡时可使用合成样本。另一方面, 在模型的目标中心点与分类中用 Focal Loss 控制正负样本、容易分类和难分类样本的权重^[12]。

交叉熵损失在二分类问题的函数为

$$CE(P, y) = \begin{cases} -\ln(P), & y = 1 \\ -\ln(1 - P), & y = -1 \end{cases} \quad (1)$$

上式中 y 表示目标的类别, 值为 1 和 -1, 类别的概率分别用 P 和 $(1-P)$ 表示。

$$P_t = \begin{cases} P, & y = 1 \\ 1 - P, & y = -1 \end{cases} \quad (2)$$

此时, $CE(P, y) = CE(P_t) = -\ln(P_t)$ 。

通过引入一个在 $[0, 1]$ 范围内的系数 α_t 实现正负样本对 loss 的控制:

$$CE(P_t) = -\alpha_t \ln(P_t), \text{ 其中 } \alpha_t = \begin{cases} \alpha, & \text{if } y = 1 \\ 1 - \alpha, & \text{otherwise} \end{cases} \quad (3)$$

通过 $(1 - P_t)$ 可以计算出某个类属于容易分类还是难分类, 引入调制系数 $(1 - P_t)^\gamma$ 来控制模型趋向于难分类样本:

$$FL(P_t) = -(1 - P_t)^\gamma \ln(P_t) \quad (4)$$

综合得

$$FL(P_t) = \begin{cases} -\alpha_t(1 - P_t)^\gamma \ln(P_t), & \text{if } y = 1 \\ -(1 - \alpha_t)P_t^\gamma \ln(1 - P_t), & \text{otherwise} \end{cases} \quad (5)$$

通过将 Focal loss 控制正负样本的加权, 降低易分类样本参与模型训练的可能性。

2.3 模型的改进与提升——提升精度

小目标可利用像素较少难以提取特征, 随着网络层数的增加, 目标的特征信息与位置信息也逐渐丢失, 难以检测。因此小目标检测引入多尺度学习提取深层语义信息与浅层表征信息。

2.3.1 替换多尺度 Backbone

Res2Net 是 ResNet 升级而来的一种多尺度特征提取网络。

图 5(b) 中的 Bottleneck block 网络第一个 1×1 卷积核负责降低通道维度, 最后一个 1×1 卷积核负责提升通道维度, 增加了网络深度, 相比 Basicblock 的计算量反而有所下降^[13]。

如图 5(c), Res2Net 在单个残差块内构造具有等级制的类似残差连接, 取代了 Bottleneck block 的 3×3 卷积核, 称为 Bottle2neck。Res2Net 在更细粒度级别表示多尺度特征, 并增加了每个网络层的感受野。

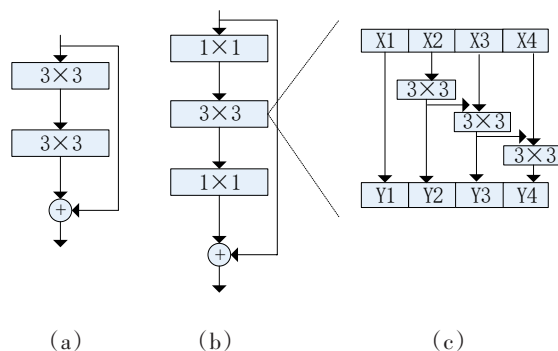


图 5 Basicblock、Bottleneck、Bottle2neck(4s) 结构

在 Res2Net 中引入了 Scale 层, Scale 表示 block 中分支的个数, 图 5(c) 中 X 分成 4 份, 表示 $s = 4$ 。而 w 表示每一个分支的通道数目, Res2Net 用一组较小的卷积核组替代 n 个通道的 3×3 卷积核, 这些卷积核组以分层类似残差的连接, 用来增加尺度表示输出特征^[14]。由于 s 和 w 的引入, Res2Net 具有多个版本。

表 3 ImageNet 数据集的目标检测效果

Res2Net 版本	Params/M	MACCs	Top1 error/%
Res2Net-50-48w-2s	25.29	4.2	22.68
Res2Net-50-26w-4s	25.70	4.2	22.01
Res2Net-50-14w-8s	25.06	4.2	21.86
Res2Net-50-26w-6s	37.05	6.3	21.42
Res2Net-50-26w-8s	48.40	8.3	20.80
Res2Net-101-26w-4s	45.21	8.1	20.81
Res2Net-v1b-50	25.72	4.5	19.73
Res2Net-v1b-101	45.23	8.3	18.77

注: MACCs (multiply-accumulate operations) 体现运行速度, MACCs 越小模型速度越快

选择参数量较少的版本替换原有的 Backbone 在自建的牛羊数据集上运行。

表 4 自建数据集的目标检测效果

Backbone	Residual Block	Parameters/M	mAP/%	$AP_{cattle-sheep}/\%$	FPS
ResNet-18	Basicblock	14.04	82.44	5.18	148.60
ResNet-34	Basicblock	24.15	83.73	5.25	114.51
ResNet-50	Bottleneck	32.67	85.33	5.95	88.50
ResNet-101	Bottleneck	51.66	84.08	5.14	52.76
ResNet-152	Bottleneck	67.30	77.30	6.69	37.48
Res2Net-14w-8s	Bottle2neck	32.17	85.67	6.63	45.80
Res2Net-26w-4s	Bottle2neck	32.81	85.95	5.45	42.70
Res2Net-48w-2s	Bottle2neck	32.10	85.36	4.49	83.25
Res2Net-v1b-50	Bottle2neck	32.83	87.49	6.83	46.00

多类目标的检测模型需要直观地展现各指标的联系，选择各项指标平衡性最佳的模型。由于参数量大的模型训练小数据集更易过拟合，如图 6 所示，选择精度和参数量两个指标平衡性最好的 Res2Net-v1b-50。

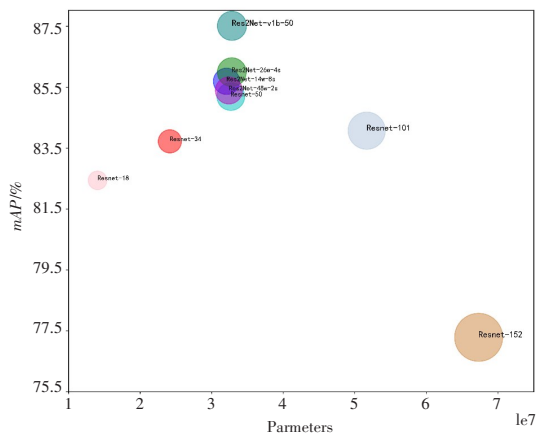


图 6 Paraeters-mAP示意图

牛、羊两类目标具有一定差异，训练效果存在差异，绘制直方堆叠图展示模型性能，红色星星柱和蓝色斜线柱的高度分别代表 AP_{sheep} 和 AP_{cattle} 的值，整体的高度反映 mAP 高低，直方堆叠图从 mAP 和两类 AP 两个角度比较模型的性能。

图 7 中选择 Res2Net-v1b-50 的模型 AP 最高且参数量适中， $AP_{cattle-sheep}$ 虽然较高，但同类型的 Res2Net-26w-4s 的 $AP_{cattle-sheep}$ 较低，Res2Net-v1b 仅修改了 Res2Net-26w-4s 代码的顶层，用三层卷积组合代替了原来的一层卷积层，后续有改善 $AP_{cattle-sheep}$ 的可能。选用 Res2Net-v1b-50 的模型

综合表现更好。

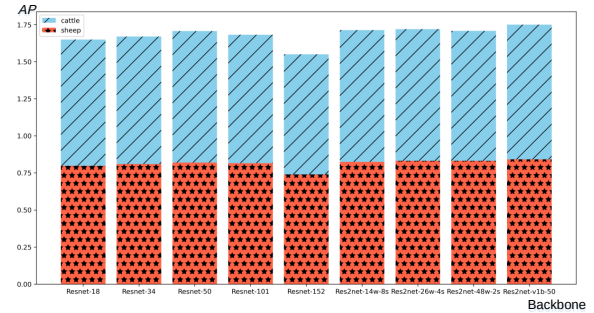


图 7 $AP_{cattle-sheep}$ 直方堆叠图

2.3.2 引入注意力机制

在深度卷积神经网络中嵌入注意模块可以提升部分性能^[15]。引入常用的 3 种注意力机制，传统的通道注意力 SENet^[16]，拥有空间及通道注意力 CBAM^[17]，以及优化 SENet 而来的轻量化 ECANet^[18]，选择效果最佳的注意力机制。注意力机制是一个即插即用的模块，理论上可以置于任何一个特征层后，由于是改进后的组合模型，放在主干网络会导致预训练权重无法使用，选择加在 Backbone 中，针对 Res2Net 中添加了 Scale 层，为不影响 Res2Net 在更细粒度级别表示多尺度特征，选择 Res2Net-v1b-50 在 Backbone 的 Attention Module-1 和 Attention Module-2 两个位置分别加入这三类注意力机制。训练结果见表 5，在 Attention Module-1 处添加 ECANet 提升效果最佳，最终载畜量计数模型选择用此改进模型。

表 5 添加注意力机制后自建数据集的检测效果

注意力机制	添加位置	mAP/%	$AP_{cattle-sheep}/\%$	FPS
SENet	Attention Module-1	86.58	5.18	45.96
	Attention Module-2	86.24	6.11	46.70
CBAM	Attention Module-1	87.04	7.04	47.18
	Attention Module-2	87.86	7.27	46.45
ECANet	Attention Module-1	88.22	5.74	46.07
	Attention Module-2	87.52	9.31	46.55

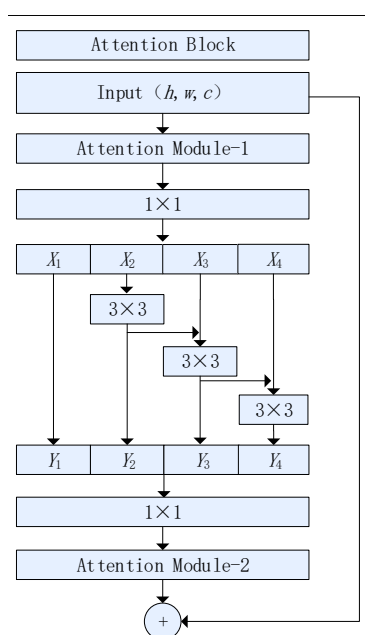


图8 注意力机制位置添加示意图

2.3.3 网络训练及参数设置

模型其他参数设置：选择adam优化器来更新学习率从而对网络进行训练，达到快速收敛的效果；使用余弦退火指数衰减策略，使得模型收敛达到最优解，避免陷入局部最小值；训练时将初始学习率设置为 5×10^{-6} ，网络训练时epoch值设置为500，Batch_size值设置为32。

2.4 评价指标

2.4.1 检测速度评价

模型的参数量与计算速度有一定的关系，而评价模型检测速度主要是比较FPS(frames per second)。

通过代码记录检测每帧图片所用的时间 T (单位：秒)，得：

$$FPS = \frac{1}{T} \quad (6)$$

2.4.2 检测精度评价

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

其中， TP ：被正确识别为正样本的数量； FP ：被错误识别为正样本的数量； TN ：被正确识别为负样本的数量； FN ：被错误识别为负样本的

数量。

根据训练数据绘制Precision-Recall曲线，得到：

$$AP = \int_0^1 P(R) dr \quad (9)$$

本实验类平均精度为

$$mAP = \frac{AP_{sheep} + AP_{cattle}}{2} \quad (10)$$

为了反映对牛羊类别的检测差异，计算牛羊 AP 差值 $AP_{cattle-sheep}$ ，得：

$$AP_{cattle-sheep} = AP_{cattle} - AP_{sheep} \quad (11)$$

2.4.3 漏检率评价

参考行人检测的评价标准，对数平均漏检率(log-average miss rate, LAMR)是反映多目标图像检测能力的指标^[19]，每幅图误检个数(false positive per image, FPPI)：

$$FPPI = \frac{FP}{FN + TN} \quad (12)$$

漏检率(miss rate, MR)反映正确预测负样本的能力。

$$MR = \frac{FN}{TP + FN} \quad (13)$$

绘制类似于Precision-Recall曲线的MR-FPPI曲线，在横坐标FPPI的 10^{-2} 到 10^0 之间均匀取9个点，获取这9个点的MR值。

$$LAMR = e^{\left[\frac{\sum \ln(MR)}{9} \right]} \quad (14)$$

图像中FPPI上限与牛羊密度有关，建立在MR-FPPI曲线上的LAMR对模型有一定参考价值，LAMR越小模型性能越好。

3 结果与分析

3.1 模型检测显示

为了检验改进后模型的检测能力，用自建牛羊数据集以同样的参数设置，对以ResNet-50为初步特征提取网络的CenterNet模型进行训练。牛和羊的检测框设为绿色和红色(如图9所示)，在只含牛的图片检测中，使用ResNet-50提取特征信息不足，存在将背景识别为牛的情况；在同时含有牛羊目标的图像中，最终模型提升了多只羊粘连的检测以及牛羊的分类识别能力；在羊群密集时，最终模型对羊目标检测改善明显。

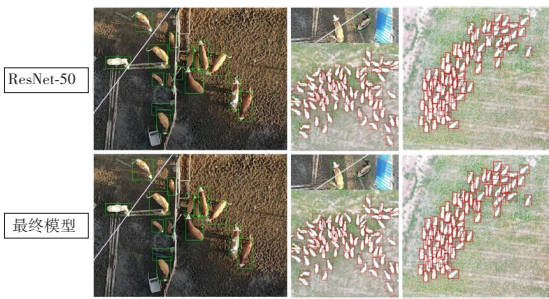


图 9 目标检测框效果

CenterNet 检测框的生成依靠 Heatmap 图，见图 10，在 Heatmap 中，若某个坐标处有物体目标的中心点，即在该坐标处产生一个高斯圆。由于检测框是在 Heatmap 的基础上生成，通过 Heatmap 图可以看到黑色牛，尤其是偏黑色背景中 ResNet-50 上的检测效果较差，而最终模型提取更多特征信息，对于拥有较小信息量的目标和受背景干扰目标的检测效果更佳。

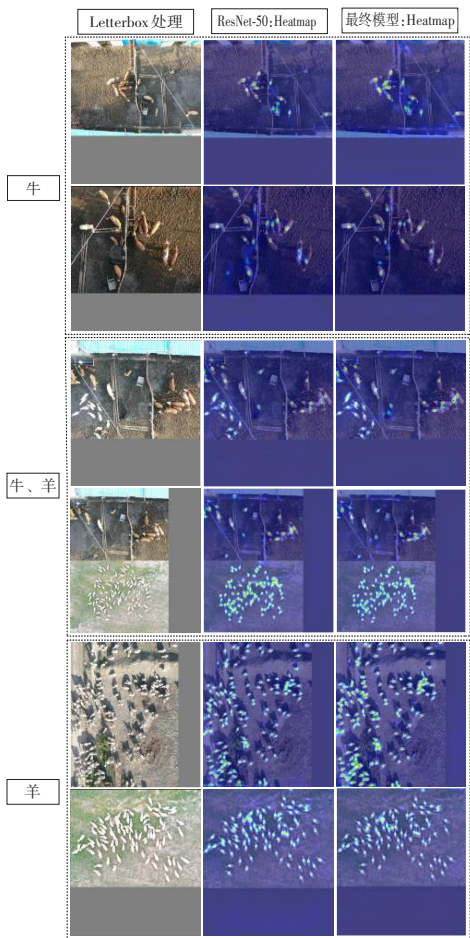


图 10 Heatmap 目标检测效果

3.2 实验结果

参数量接近时，牺牲部分检测速度，最终模型在牛羊计数任务上取得了更好的结果，见表 6。

表 6 模型参数比较

Backbone	ResNet-50 (比较模型)	Res2Net-v1b-ECA (最终模型)
<i>Params/M</i>	32.67	32.83
<i>FPS</i>	88.50	46.10
<i>mAP/%</i>	85.33	88.22
<i>AP_{cattle-sheep}/%</i>	5.95	5.74
<i>LAMR_{cattle}</i>	0.21	0.11
<i>LAMR_{sheep}</i>	0.69	0.58

4 结语

改进后的最终模型对比使用 ResNet-50 的模型 *mAP* 提升了 2.89 个百分点，对比 CenterNet 原文中使用的 ResNet-18 和 ResNet-101 分别提升了 5.78 和 4.15 个百分点，最终模型中牛和羊的 LAMR 对比 ResNet-50 都下降了约 10 个百分点，提升了密集目标的检测能力。最终模型提升了个体分界不明显和粘连目标的检测能力。

参考文献：

[1] 张秀青,谢兰兰. 国际牛羊肉市场发展趋势及我国面临的风险点分析[J]. 价格理论与实践,2022(2): 79-83,200.

[2] 耿丽微,钱东平,赵春辉. 基于射频技术的奶牛身份识别系统[J]. 农业工程学报,2009,25(5):137-141.

[3] 李琦,尚绛岚,李宝山. 基于头部图像特征的草原羊自动计数方法[J]. 中国测试,2020,46(11):20-24.

[4] 伍天舟. 基于改进 SSD 算法的草原羊群个体识别与计数[D]. 呼和浩特市:内蒙古农业大学,2021.

[5] 赵建敏,李雪冬,李宝山. 基于无人机图像的羊群密集计数算法研究[J]. 激光与光电子学进展,2021, 58(22):220-229.

[6] 杜永兴,苗晓伟,秦岭,等. 基于 VNet 卷积神经网络的羊群计数[J]. 激光技术,2021,45(5):675-680.

[7] 高新波,莫梦竟成,汪海涛,等. 小目标检测研究进展[J]. 数据采集与处理,2021,36(3):391-417.

[8] CHENG G, YUAN X, YAO X, et al. Towards large-scale small object detection: survey and benchmarks[EB/OL]. arXiv:2207.14096,2022.

[9] ZHOU X, WANG D, KRÄHENBÜHL P. Objects

- as points[EB/OL]. arXiv:1904.07850,2019.
- [10] MAIMAITI M, LIU Y, LUAN H, et al. Enriching the transfer learning with pre-trained lexicon embedding for low-resource neural machine translation[J]. Tsinghua Science and Technology, 2022, 27 (1) : 150-163.
 - [11] OKSUZ K, CAM B C, KALKAN S, et al. Imbalance problems in object detection: a review [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2021, 43(10):3388-3415.
 - [12] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 42 (2):318-327.
 - [13] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA. New York:IEEE, 2016:770-778.
 - [14] GAO S, CHENG M M, ZHAO K, et al. Res2Net: a new multi-scale backbone architecture [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(2):652-662
 - [15] ZHANG H, ZU K, LU J, et al. EPSANet: an efficient pyramid squeeze attention block on convolutional neural network [EB/OL]. arXiv: 2105.14447, 2021.
 - [16] HU J, Li. SAMUEL A, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 42(8):2011-2023.
 - [17] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018:3-19.
 - [18] WANG Q, WU B, ZHU P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C]//Proceeding of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Piscataway, NJ:IEEE Press, 2020:11531-11539.
 - [19] DOLLAR P, WOJEK C, SCHIELE B, et al. Pedestrian detection: an evaluation of the state of the art [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(4):743-761.

CenterNet-based grassland cattle and sheep counting study

Sun Qiang, Hao Min*

(College of Mechanical and Electrical Engineering, Inner Mongolia Agricultural University, Hohhot 010000, China)

Abstract: Rapid and accurate statistics on the number of cattle and sheep grazing on grassland is an important prerequisite for scientific grazing management and grassland grass-livestock balance analysis. As the market demand for livestock products increases, the number of grassland grazing cattle and sheep in Inner Mongolia, the main production area of livestock products in China, increases year by year, while the current manual work of counting the number of grazing cattle and sheep has problems such as difficulty, low efficiency and unstable statistical accuracy. In order to realize the fast counting of cattle and sheep, we use the aerial photography of the cattle and sheep images collected by UAV and build the dataset, use the multi-scale feature extraction network Res2Net to improve the detection effect of the anchor-free CenterNet model on the small targets adhered to the grazing cattle and sheep images under the aerial photography scenario, add the ECANet attention mechanism to further improve the model. The performance of the model is further improved by adding ECANet attention mechanism, and the relevant parameters are adjusted for algorithm optimization. The research results show that the class average accuracy of the model in the dataset acquired by UAV aerial photography reaches 88.22%, and the logarithmic average missed detection rate of cattle and sheep, which reflects the effect of dense target detection, decreases by 10% and 11%, respectively. This result indicates the value of introducing contactless computer vision technology in grassland cattle and sheep counting.

Keywords: CenterNet; object detection; counting; Res2Net; attention mechanism

文章编号: 1007-1423(2023)11-0009-07

DOI: 10.3969/j.issn.1007-1423.2023.11.002

结合加权核范数与 3D 全变分的目标检测

班 颖*, 田 韵, 邵泽军

(燕京理工学院建筑学院, 廊坊 065201)

摘要: 针对现实复杂场景使目标检测精确度下降的问题, 提出了一种结合加权核范数与 3D 全变分的目标检测模型。基于扩展的鲁棒主成分分析模型, 首先将视频分解为低秩静态背景、稀疏平滑前景和稀疏动态背景, 利用加权核范数对背景进行低秩约束, 考虑了不同奇异值对秩函数的影响; 为加强前景的时空连续性, 利用 3D-TV 来约束运动目标, 有效抑制了复杂背景的干扰作用。实验表明, 所提算法检测运动目标的准确率较高, 能有效抑制复杂背景的干扰作用。

关键词: 目标检测; 鲁棒主成分分析; 加权核范数; 3D 全变分; 交替方向乘法

0 引言

近年来, 扩展的鲁棒主成分分析模型(robust principal component analysis, RPCA)^[1]在运动目标检测领域得到了较好的发展, 其模型简单, 求解高效。该模型将视频序列组成的矩阵分解为低秩静态背景、稀疏平滑前景和稀疏动态背景, 并分别利用核范数和 L_1 范数进行低秩和稀疏约束。

上述扩展的 RPCA 存在如下不足: ①核范数被定义为背景矩阵所有奇异值之和, 没有考虑不同奇异值对秩函数的影响; ②RPCA 模型将前景看作背景中存在的异常像素点, 在复杂背景中前景检测精确度下降, 而 L_1 范数不能有效应对复杂背景对前景提取造成的干扰。

为解决问题①, Gu 等^[2]提出了加权核范数, 考虑了不同奇异值对秩函数的影响, 加强了背景的低秩性。为提高背景建模在复杂场景中的稳定性, 加权 Schatten-p 范数最小化模型被提出^[3]。文献[4]基于 t-product 提出了一种新的张量核范数, 为准确恢复低秩和稀疏成分提供了保证。为解决问题②, 并提高目标检测的准确

率, 高阶鲁棒主成分分析模型(Higher-order RPCA, HoRPCA)被提出^[5]。在文献[6]中, 3D 全变分(3D Total Variation, 3D-TV)被提出用来约束稀疏前景, 有效抑制了由动态背景造成的噪声干扰。文献[7]利用超像素生成步骤和树结构稀疏性动态估计前景, 有效去除了背景。

基于以上认识, 为加强对实际背景的近似和应对复杂背景的变化, 提出了一种结合加权核范数与 3D 全变分的目标检测模型。该模型将视频序列组成的矩阵分解为低秩静态背景、稀疏平滑前景和稀疏动态背景, 利用加权核范数对背景进行低秩约束, 考虑了不同奇异值对秩函数的影响; 为加强前景的时空连续性, 利用 3D-TV 对运动目标进行稀疏约束, 有效抑制了动态背景的噪声干扰。

1 扩展的 RPCA 模型

假设给定的视频序列 $I \in \mathbf{R}^{m \times n \times t}$, m 和 n 分别表示视频的宽和高, t 表示视频的帧数, 将每帧图像向量化形成矩阵 O 。矩阵 O 可分解为

$$O = B + M \quad (1)$$

收稿日期: 2023-01-11 修稿日期: 2023-02-23

作者简介: *通信作者: 班颖(1995—), 女, 河北廊坊人, 硕士, 研究方向为目标检测, E-mail: yingmi_9586@126.com; 田韵(1997—), 女, 河北保定人, 硕士, 研究方向为复杂系统优化控制理论; 邵泽军(1981—), 男, 山东菏泽人, 硕士, 研究方向为最优化方法

其中: $B \in \mathbf{R}^{mn \times t}$ 是静态背景, $M \in \mathbf{R}^{mn \times t}$ 为残差。

为应对现实生活场景的复杂多变, 将残差 M 分解为稀疏前景 F 和动态背景 E :

$$M = E + F \quad (2)$$

其中: $F \in \mathbf{R}^{mn \times t}$ 和 $E \in \mathbf{R}^{mn \times t}$ 均具有较强的稀疏性。

由于静态背景具有较强的相关性, 因此对背景 B 采用低秩约束。对前景 F 和动态背景 E 分别采用稀疏约束, 具体模型如下^[1]:

$$\begin{aligned} \min_{B, M, E, F} & \|B\|_* + \gamma_1 \|M\|_1 + \gamma_2 \|E\|_1 + \gamma_3 \varphi(F) \\ \text{s.t. } & O = B + M, M = E + F \end{aligned} \quad (3)$$

其中: $\|B\|_* = \sum_i |\sigma_i(B)|$, $\sigma_i(B)$ 是 B 的第 i 个奇异值, $\|\cdot\|_0$ 表示 L_0 范数, γ_1 , γ_2 和 γ_3 是权重系数。 $\varphi(\cdot)$ 用于约束前景目标的稀疏性和时空连续性。

2 结合加权核范数与 3D 全变分的目标检测

2.1 背景模型

在 RPCA 模型中, 核范数定义为背景矩阵的所有奇异值之和, 没有考虑不同奇异值对秩函数的影响, 因此对秩函数的近似程度不够, Gu 等^[2]提出了加权核范数, 定义如下:

$$\|B\|_* = \sum_i |\omega_i \sigma_i(B)| \quad (4)$$

由于较大的奇异值对秩函数的影响大, 因此施加较大的权重系数, 对较小的奇异值施加小的权重系数。上述加权核范数考虑了不同奇异值对秩函数的影响, 使其更加接近实际背景的秩。基于以上考虑, 本文采用上述的加权核范数对背景进行低秩约束, 进一步加强了背景的低秩性。

2.2 前景模型

运动目标, 一般是视频中的显著性运动对象, 且在整个图像上的分布占比较小, 因此前景目标具有稀疏性。另外, 运动目标往往是占据一定比例的连续区域, 因此具有时空连续性和光滑性, 而视频背景中的雪花和摇晃的树叶等呈现不连续的噪声特征^[2]。在数学上, 3D 全变分^[6]具有平滑信号的功能, 可以有效抑制视

频中不连续的噪声干扰。对于前景张量 $F(i, j, k) \in \mathbf{R}^{m \times n \times t}$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n; k = 1, 2, \dots, t$), 利用 3D 全变分来约束前景的稀疏性和时空连续性, 如下^[6]:

$$\begin{aligned} \text{TV}_{i,j,k}(f) = & |F(i+1, j, k) - F(i, j, k)| + \\ & |F(i, j+1, k) - F(i, j, k)| + \\ & |F(i, j, k+1) - F(i, j, k)| \end{aligned} \quad (5)$$

3D 全变分通过将上述 $\text{TV}_{i,j,k}$ 相加得到:

$$\|F\|_{3D-TV} = \sum_{i,j,k} \text{TV}_{i,j,k} \quad (6)$$

为方便计算, 引入沿水平方向、垂直方向以及时间方向的向量差分算子:

$$\begin{cases} D_h f = \text{vec}(F(i+1, j, k) - F(i, j, k)) \\ D_v f = \text{vec}(F(i, j+1, k) - F(i, j, k)) \\ D_t f = \text{vec}(F(i, j, k+1) - F(i, j, k)) \end{cases}$$

令 $Df = [(D_h f)^T, (D_v f)^T, (D_t f)^T]^T$ 表示三个向量差分算子的联级, 则上述 3D 全变分模型可改写为

$$\|F\|_{3D-TV} = \|Df\|_1 = \|D_h f\|_1 + \|D_v f\|_1 + \|D_t f\|_1 \quad (7)$$

从 3D 全变分的定义看出, 沿水平和垂直方向的差分算子说明 3D 全变分考虑了运动目标在二维空间上的连续性; 沿时间方向的差分算子说明其考虑了运动目标在时间上的连续性。另外, 3D 全变分能有效抑制动态背景造成的不连续噪声干扰, 且考虑了运动目标在时空上的连续性, 加强了模型在复杂背景中的稳定性, 提高了目标检测的准确性。基于以上考虑, 本文采用 3D 全变分约束运动目标的稀疏性和时空连续性。

2.3 新模型的建立

综上所述, 本文提出了一种新的结合加权核范数和 3D-TV 的目标检测模型, 具体模型如下:

$$\begin{aligned} \min_{B, M, E, F} & \|B\|_{\omega^*} + \gamma_1 \|M\|_1 + \gamma_2 \|E\|_1 + \gamma_3 \|F\|_{3D-TV} \\ \text{s.t. } & O = B + M, M = E + F \end{aligned} \quad (8)$$

利用 3D-TV 的变形公式(7), 上述模型可转换成如下:

$$\begin{aligned} \min_{B, M, E, F} & \|B\|_{\omega^*} + \gamma_1 \|M\|_1 + \gamma_2 \|E\|_1 + \gamma_3 \|Df\|_1 \\ \text{s.t. } & O = B + M, M = E + F \end{aligned} \quad (9)$$

其中： $\gamma_1, \gamma_2, \gamma_3$ 为权重系数。

3 模型的求解

引入辅助变量 g 求解上述模型，式(9)变形为

$$\min_{B, M, E, F} \|B\|_{\omega^*} + \gamma_1 \|M\|_1 + \gamma_2 \|E\|_1 + \gamma_3 \|g\|_1$$

$$\text{s.t. } O = B + M, M = E + F, g = Df \quad (10)$$

利用交替方向乘子法(alternating direction multiplier method, ADMM)^[8]求解式(10)，其增广拉格朗日函数为

$$L_u(B, M, E, F, g, Y_1, Y_2, Y_3, u) =$$

$$\|B\|_{\omega^*} + \gamma_1 \|M\|_1 + \gamma_2 \|E\|_1 + \gamma_3 \|g\|_1 + \langle O - B - M, Y_1 \rangle +$$

$$(\mu/2) \|O - B - M\|_F^2 + \langle M - E - F, Y_2 \rangle + (\mu/2) \|M - E - F\|_F^2 +$$

$$\langle g - Df, Y_3 \rangle + (\mu/2) \|g - Df\|_F^2$$

其中： Y_1, Y_2, Y_3 是拉格朗日乘子， μ 为参数。

进行变量分离，上述增广拉格朗日函数转化成几个子问题。

(1) 固定变量 $M^k, E^k, F^k, g^k, Y_1^k, Y_2^k, Y_3^k$ 和 μ^k ，更新 B^{k+1} ：

$$B^{k+1} = \arg \min_B \|B\|_{\omega^*} + \frac{\mu^k}{2} \|B - Q\|_F^2 \quad (11)$$

问题的解为^[2]

$$B^{k+1} = US_{\omega} \left(\sum \right) V^T \quad (12)$$

其中： $Q = O - M^k + Y_1^k/\mu^k$ ， $Q = U \sum V^T$ 为奇异值分解， ω 为权重向量， $S_{\omega} \left(\sum \right)_{ii} = \max \left(\sum_{ii} - \omega_{ii}, 0 \right)$ 是具有权重向量的广义软阈值算子。

(2) 固定变量 $B^{k+1}, E^k, F^k, g^k, Y_1^k, Y_2^k, Y_3^k$ 和 μ^k ，更新 M^{k+1} ：

$$M^{k+1} = \arg \min_M \gamma_1 \|M\|_1 + \mu^k \|M - P\|_F^2 \quad (13)$$

该子问题可以通过软阈值算子求得^[9]：

$$M^{k+1} = S_{\mu^k/2\gamma_1}(P) \quad (14)$$

其中： $P = (O + E^k - F^k - B^{k+1})/2 + (Y_1^k - Y_2^k)/2\mu^k$ ， $S_a(b) = \text{sgn}(b) \max(|b| - a, 0)$ 是软阈值算子。

(3) 固定变量 $B^{k+1}, M^{k+1}, F^k, g^k, Y_1^k, Y_2^k, Y_3^k$ 和 μ^k ，更新 E^{k+1} ：

$$E^{k+1} = \arg \min_E \|E\|_1 + \frac{\mu^k}{2} \|E - N\|_F^2 \quad (15)$$

通过软阈值算子求得^[9]：

$$E^{k+1} = S_{\mu^k/2\gamma_2}(N) \quad (16)$$

其中： $N = M^{k+1} - F^k + Y_2^k/\mu^k$ 。

(4) 固定变量 $B^{k+1}, M^{k+1}, E^{k+1}, g^k, Y_1^k, Y_2^k, Y_3^k$ 和 μ^k ，更新 F^{k+1} ：

$$\mu^k (I + D^T D) f = G \quad (17)$$

其中： $G = \text{vec}[Y_2^k + \mu^k M^{k+1} - \mu^k E^{k+1} + D^T(Y_3^k + \mu^k g^k)]$ 。

通过3D快速傅里叶变换(3D Fast Fourier Transform, 3D FFT)^[6]求解有：

$$f^{k+1} = \text{ifftn} \left(\frac{\text{fftn}(G)}{\mu^k I + \mu^k \left(|\text{fftn}(D_h)|^2 + |\text{fftn}(D_v)|^2 + |\text{fftn}(D_t)|^2 \right)} \right)$$

$$F^{k+1} = \text{reshape}(f^{k+1}) \quad (18)$$

其中： fftn 和 ifftn 分别为快速3D傅里叶变换和逆变换。

(5) 固定变量 $B^{k+1}, M^{k+1}, E^{k+1}, F^{k+1}, Y_1^k, Y_2^k, Y_3^k$ 和 μ^k ，更新 g^{k+1} ：

$$g^{k+1} = \arg \min_g \gamma_3 \|g\|_1 + \frac{\mu^k}{2} \left\| g - (Df^{k+1} + \frac{Y_3^k}{\mu^k}) \right\|_F^2 \quad (19)$$

通过软阈值算子求得^[9]：

$$g^{k+1} = S_{\mu^k/\gamma_3} \left(Df^{k+1} + \frac{Y_3^k}{\mu^k} \right) \quad (20)$$

(6) 更新拉格朗日乘子 Y_1^k, Y_2^k, Y_3^k 和参数 μ 。

4 数值实验结果和分析

4.1 实验环境与参数选取

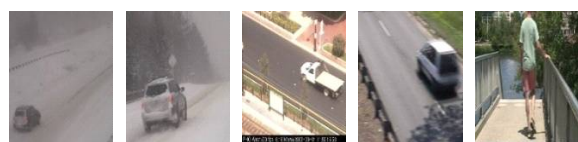
为了验证所提算法的效果，将本文算法与RPCA^[10]，IALM-RPCA^[11]，HoRPCA^[5]，KBR-RPCA^[12]，TRPCA-TNN^[4]，以及KBR-L₁₁₂^[13]分别从主观和客观两个方面进行比较。所有实验的运行环境为Matlab 2014a，Inter Core i5-6500处理器，8 GB的内存，Win10 64位操作系统。实验的测试数据均来自于CD.net数据库^[14]。

为了在客观上准确评估本文算法的性能，采用Recall(查全率)、Precision(查准率)和综合评价指标F值(F-measure)作为提取运动前景的评价指标^[15]。

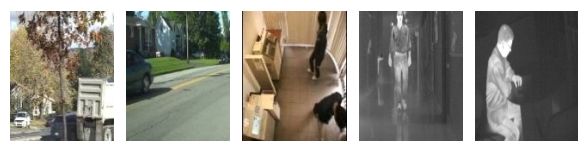
4.2 主观分析

在仿真实验中，本文主要选取了5类监控视

频进行结果展示和分析, 分别为: Bad Weather (a-b)、Camera Jitter(c-d)、Dynamic background (e-f)、Shadow(g-h)和 Thermal(i-j), 如图 1 所示。



(a)Blizzard (b)Snowfall (c)Boulevard (d)Traffic (e)Overpass



(f)Fall (g)Bungalows (h)Copymach (i)Corridor (j)Library

图 1 前景目标时空连续性图

图 2 前 6 行分别给出了复杂天气、相机抖动和动态背景的视觉对比图。从图中看出, 对于复杂天气 Blizzard 和 Snowfall, IALM-RPCA 和 HoRPCA 提取的目标受动态背景的干扰较大, 提取的运动目标出现了较多的背景部分, RPCA 和 KBR-RPCA 提取的前景较为模糊空洞, TRPCA-TNN 和 KBR-L₁₁₂对目标的提取效果较好, 但有部分的前景误判, 本文算法提取的运动目标较为完整清晰, 且对前景的误判较小, 有效抑制了复杂天气对前景提取的干扰。对于相机抖动 Boulevard 和 Traffic, IALM-RPCA、HoRPCA 和 KBR-L₁₁₂, 虽然提取的前景目标较为完整, 但同时由于相机抖动前景视觉图中提取了较多的背景线条, RPCA、KBR-RPCA 和 TRPCA-TNN 提取前景的精确度较低, 前景目标中的空洞现象较大,

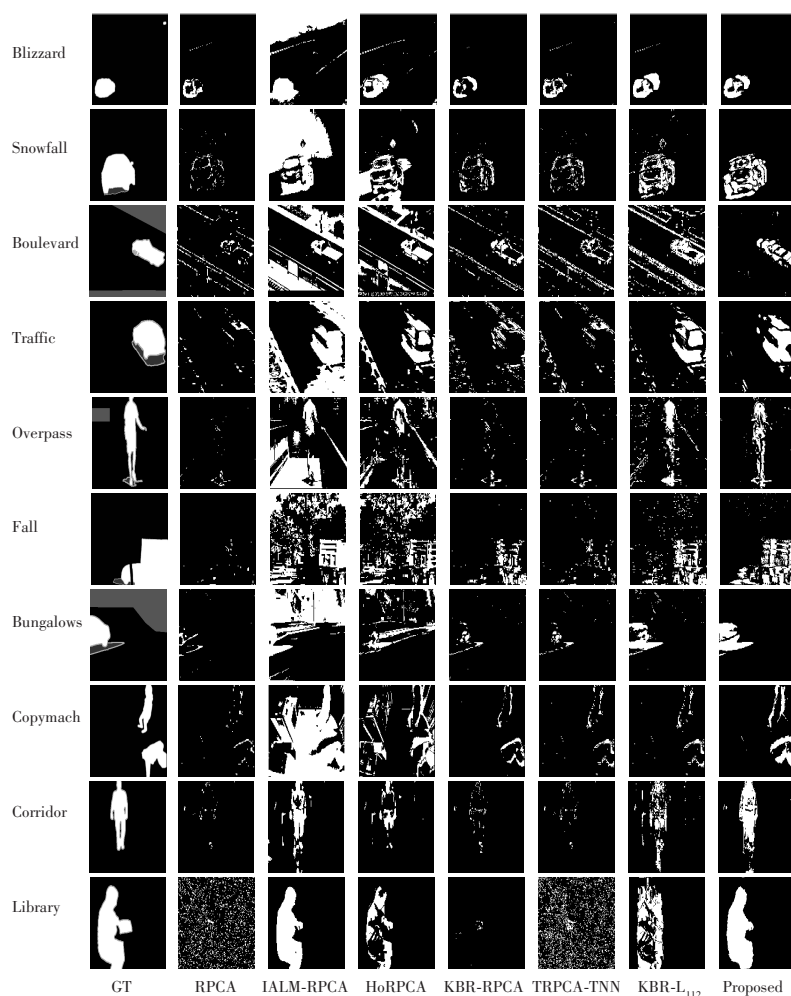


图 2 不同视频的视觉对比

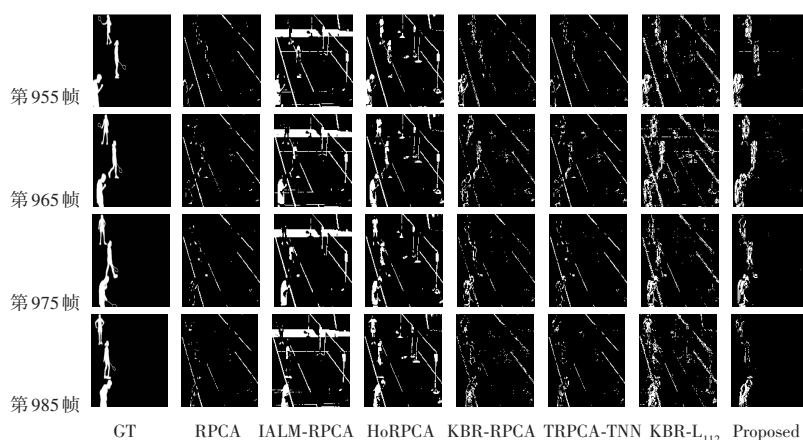


图3 不同算法下不同帧的视觉对比(Badminton)

且把部分背景误判为前景目标，本文算法提取的运动目标较为完整，对背景的干扰作用去除较好，且对前景的误判较小。对于动态背景 Overpass 和 Fall，RPCA、KBR-RPCA 和 TRPCA-TNN 提取的前景目标只出现部分轮廓，运动目标不完整，IALM-RPCA 和 HoRPCA 的视觉效果图中出现了大量的背景物体，KBR- L_{112} 和本文算法提取的前景较为完整，但由于湖水荡漾和树叶摆动，KBR- L_{112} 的前景视觉图中出现较多的噪点，而本文算法对前景的误判较小，有效抑制了动态背景的干扰作用。

图2后4行分别给出了阴影和热成像的视觉对比图。对于阴影图像 Bungalows 和 Copymach，IALM-RPCA 和 HoRPCA 的视觉效果图中出现了较多的背景部分，前景背景的区分度较低，RPCA、KBR-RPCA、TRPCA-TNN 和 KBR- L_{112} 提取的前景较为空洞，前景目标的提取精确度较低。本文算法提取目标的精确度较高，去除背景的效果较好。对于热成像图像 Corridor 和 Library，RPCA、KBR-RPCA 和 TRPCA-TNN 提取前景的效果较差，HoRPCA 和 KBR- L_{112} 的视觉效果图中出现较大的空洞现象，IALM-RPCA 和本文算法提取前景的精确度较好，但两者相比本文算法去除背景的效果更好。

图3给出了视频 Badminton 在4个不同帧下的视觉对比图。从中看出，IALM-RPCA 和 HoRPCA 提取的前景较为完整，但同时把背景误判为前景的错误率较高，RPCA、KBR-RPCA、TRPCA-TNN 和 KBR- L_{112} 提取的运动目标空洞较

多，同时含有较多的背景线条，本文算法提取的前景目标较为完整，且去除背景的准确率较高。

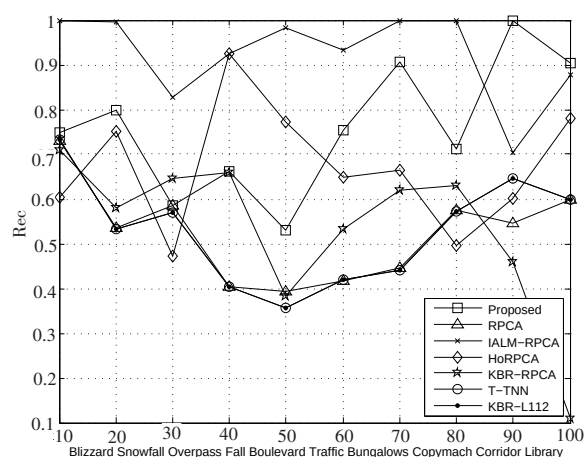
综上所述，与其余6种算法对比，本文算法提取的前景较为完整，且有效去除背景，对前景的误判较低，整体获得最佳的检测效果。

4.3 客观分析

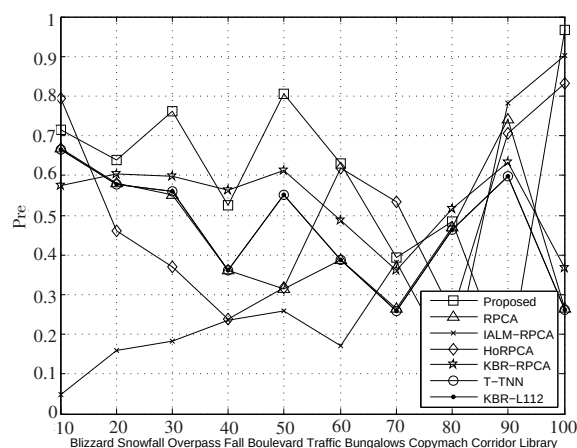
将本文算法的 *Recall* 和 *Precision* 值分别在5类不同的监控视频上与其他6种算法进行对比，其对比如图4所示。从图4可知，除 IALM-RPCA 外，本文算法的 *Recall* 和 *Precision* 值基本上高于其他算法，虽然本文算法的 *Recall* 和 *Precision* 低于 IALM-RPCA，但综合评价指标 *F* 值远高于 IALM-RPCA，说明本文算法有较好的前景查全率和查准率。

由于 *Recall* 只能反映丢失运动目标内部信息的相关性，*Precision* 只能反映丢失目标外部信息的相关性，且两者指标值有时会出现矛盾的情况，因此采用它们的调和平均值 *F* 值来综合判断提取效果更为准确。本文算法与其他6种算法的 *F* 值对比见表1，其中加粗为最优值，下划线为次优值。从表1可以看出，本文算法的 *F* 值均处于最优或次优的情况，说明本文算法有较好的前景提取效果，对前景的误判较小。

表2给出了视频 Badminton 在955、965、975和985帧不同算法的 *F* 值。从表2看出，所提算法在相同视频不同帧下的 *F* 值基本上处于最优的情况，说明所提算法能稳定地分离视频的前景与背景，且分离准确率较高。



(a) Recall 值



(b) Precision 值

图 4 Recall和 Precision数值对比

表 1 不同算法下的 F 值

F 值	RPCA	IALM-RPCA	HoRPCA	KBR-RPCA	TRPCA-TNN	KBR-L ₁₁₂	本文
Blizzard	0.560	0.093	0.667	0.652	0.559	<u>0.672</u>	0.723
Snowfall	0.495	0.274	0.537	0.587	0.492	<u>0.607</u>	0.612
Boulevard	0.313	0.406	0.452	<u>0.436</u>	0.312	0.419	0.655
Traffic	0.314	0.294	0.605	0.346	0.313	<u>0.592</u>	0.657
Overpass	0.388	0.303	0.392	0.516	0.384	<u>0.557</u>	0.663
Fall	0.380	0.380	0.388	0.699	0.381	0.506	<u>0.662</u>
Bungalows	0.505	0.552	0.626	0.530	0.503	<u>0.666</u>	0.724
Copymach	0.432	0.222	0.311	<u>0.519</u>	0.426	0.522	0.522
Corridor	0.292	<u>0.742</u>	0.634	0.403	0.291	0.555	0.815
Library	0.367	0.926	0.805	0.167	0.364	0.710	<u>0.915</u>

表 2 不同算法下不同视频帧的 F 值

帧数/ F 值	RPCA	IALM-RPCA	HoRPCA	KBR-RPCA	TRPCA-TNN	KBR-L ₁₁₂	本文
955	0.306	0.262	0.484	0.413	0.356	0.412	0.529
965	0.383	0.311	<u>0.502</u>	0.429	0.380	0.411	0.549
975	0.377	0.285	<u>0.477</u>	0.404	0.378	0.440	0.545
985	0.402	0.307	0.484	0.380	0.399	0.415	<u>0.460</u>

5 结语

本文提出了一种结合加权核范数与 3D 全变分的目标检测模型。该模型以改进的 RPCA 为基础,将视频分解为低秩静态背景、稀疏平滑前景和稀疏动态背景。为加强背景的低秩性,利用加权核范数对背景进行低秩约束,考虑了不同奇异值对秩函数的影响,提高了背景的去噪率;并利用 3D-TV 对前景进行稀疏约束,加强了前景的时空连续性,提高了目标检测的准确率。从实验视觉对比图看出,所提算法提取的运动目标较为清晰完整,去除背景的准确率较高,能有效抑制复杂背景对目标检测的干扰作用。所提算法的综合评价指标 F 值均处于最优或次优,说明本文算法的目标检测准确率较高。

参考文献:

- [1] CAO X C, YANG L, GUO X J. Total variation regularized RPCA for irregularly moving object detection under dynamic background[J]. IEEE Transactions on Systems, Man, and Cybernetics, 2016, 46 (4):1014-1027.
- [2] GU S H, ZHANG L, ZUO W M, et al. Weighted nuclear norm minimization with application to image denoising[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington D. C., USA: IEEE Press, 2014:2862-2869.
- [3] 陈利霞,刘俊丽,王学文. 结合加权 Schatten- p 范数与 3D 全变分的前景检测[J]. 计算机应用, 2019, 39 (4):1170-1175.
- [4] LU C Y, FENG J S, CHEN Y D, et al. Tensor robust principal component analysis with a new tensor nuclear norm [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42 (4) : 925-938.
- [5] GOLDFARB D, QIN Z. Robust low-rank tensor recovery: models and algorithms[J]. SIAM Journal on

- Matrix Analysis and Applications, 2014, 35 (1) : 225-253.
- [6] CAO W, WANG Y, SUN J, et al. Total variation regularized tensor RPCA for background subtraction from compressive measurements [J]. IEEE Transactions on Image Processing, 2016, 25(9): 4075-4090.
- [7] EBADI S E, IZQUIERDO E. Foreground segmentation with tree-structured sparse RPCA [C]//Proceedings of European Conference on Computer Vision, Amsterdam, Netherlands: IEEE Press, 2016: 314-329.
- [8] TOM A J, YIN W T, ZENG J S. Global convergence of ADMM in nonconvex non-smooth optimization [J]. Journal of Scientific Computing, 2019, 78 (1): 29-63.
- [9] WANG Y, GEORGE S-N. Tensor total variation regularized moving object detection for surveillance videos [C]//Proceedings of the 2018 International Conference on Signal Processing and Communications, Bangalore, India: IEEE Press, 2018: 327-331.
- [10] BOUWMANS T, ZAHZAH E H. Robust PCA via principal component pursuit: a review for a comparative [J]. Computer Vision and Image Understanding, 2014, 122: 22-34.
- [11] LIN Z C, CHEN M M, MA Y. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices [J]. Journal of Structural Biology, 2010, 181(2): 116-127.
- [12] XIE Q, ZHAO Q, MENG D Y, et al. Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40 (8): 1888-1902.
- [13] CHEN L X, LIU J L, WANG X W. Background subtraction with Kronecker-basis-representation based tensor sparsity and L112 norm [J]. Multidimensional Systems and Signal Processing, 2021, 32: 77-90.
- [14] WANG Y, JODOIN P-M, PORIKLI F, et al. CD-net 2014: an expanded change detection benchmark dataset [C]//Proceedings of the 2014 The IEEE Conference on Computer Vision and Pattern Recognition, Washington D. C., USA: IEEE Press, 2014: 387-394.
- [15] 左军辉, 贾振红, 杨杰, 等. 基于改进背景减法的视频图像运动目标检测 [J]. 计算机工程与设计, 2020, 41(5): 1367-1372.

Object detection with weighted nuclear norm and 3D total variation

Ban Ying*, Tian Yun, Shao Zejun

(School of Architecture, Yanching Institute of Technology, Langfang 065201, China)

Abstract: In view of the fact that the accuracy of object detection decreases in complex scenes, a new object detection model based on weighted nuclear norm and 3D Total Variation (3D-TV) is proposed. Based on extended Robust Principal Component Analysis (RPCA), the data are decomposed into low-rank static background, sparse smooth foreground and sparse dynamic background. The weighted nuclear norm is applied to constrain the background, which considers the influence of different singular values on the rank function. In order to enhance the spatio-temporal continuity of the foreground, 3D-TV is used to constrain the moving foreground, which effectively suppresses the disturbance of complex background. The experimental results demonstrate that the proposed algorithm has higher accuracy in detecting objects and can effectively suppress the disturbance of complex background.

Keywords: object detection; RPCA; weighted nuclear norm; 3D-TV; ADMM

文章编号: 1007-1423(2023)11-0016-06

DOI: 10.3969/j.issn.1007-1423.2023.11.003

基于深度强化学习的配件库存决策研究

李雨情^{1*}, 潘超凡²

(1. 西南交通大学制造业产业链协同与信息化支撑技术四川省重点实验室, 成都 611756;

2. 西南石油大学计算机科学学院, 成都 610500)

摘要: 针对供应链中制造厂配件中心库的库存决策问题, 充分考虑库存限制因素及复杂多变的配件需求特性, 将库存决策这类序贯决策问题建模为马尔可夫决策过程, 并通过深度强化学习算法对该库存决策模型进行求解, 以探索不确定需求下不同需求特性配件的最优订货量, 从而达到企业利润最大化的目的。实验结果表明, 基于深度强化学习的库存决策能够在充分贴合现实情况的前提下有效提高企业利润, 具有实际应用价值。

关键词: 深度学习; 强化学习; 库存决策

0 引言

随着中国国民经济的快速发展, 我国的汽车工业也得到了迅猛发展。据公安部统计, 2022 年全国机动车保有量达 4.17 亿辆, 其中汽车 3.19 亿辆^[1]。伴随着汽车保有量的增长, 汽车的保养和维修需求也迎来高峰期。众所周知, 汽车配件是汽车售后市场获取利润的直接源泉。由于汽车配件种类繁多, 需求呈多样性, 根据需求发生间断程度的大小, 一般将汽车配件分为两类: 连续性需求配件和非连续性需求配件^[2]。面对这样不确定的配件需求, 制定合理的库存决策是企业急需解决的重大问题, 也是扩大市场份额的重要支撑。本文根据配件的不同需求特性, 通过深度强化学习算法研究汽车配件库存决策问题。在现有研究的基础上, 本文的研究更贴合现实, 也为企业库存决策提供了一定的依据。

1 相关理论及方法

1.1 强化学习

强化学习(reinforcement learning, RL), 是

基于智能体与环境进行交互, 并从环境中获得信息奖赏并映射到动作的一种学习方式, 如图 1 所示, 其主要思想是根据系统当前环境确定下一步决策应采取的行动, 行动的结果是系统状态发生转移, 而且根据行动的后果产生一定的奖励评价决策的效果并进行反馈, 从而逐渐改进决策方法^[3]。RL 更接近于自然界生物学习的本质, 因而在很多领域取得了成功的应用。

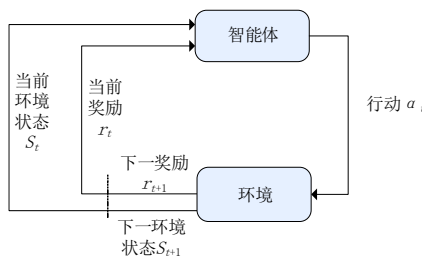


图 1 强化学习基本结构

随着深度学习的兴起, 它与基于价值函数的 RL 算法结合产生了 DQN^[4]等深度强化学习算法^[5]。而将深度学习与基于策略函数的 RL 结合, 相继产生了 DDPG^[6], PPO^[7], TD3^[8], SAC^[9]等算法。

收稿日期: 2023-02-19 修稿日期: 2023-03-04

作者简介: *通信作者: 李雨情(2000—), 女, 湖南涟源人, 硕士, 研究方向为供应链管理、强化学习, E-mail: liyuqing.lucy@foxmail.com; 潘超凡(1998—), 男, 四川成都人, 硕士, 研究方向为强化学习、自监督学习

1.2 基于强化学习的库存决策研究

针对需求不确定的情况，目前大部分库存决策研究方法通常分为两步：首先建立需求预测模型得到预测的需求，其次设计算法求解待求问题的数学模型。这样的研究方法需要依赖需求预测模型保证极佳的预测精度，难以考虑动态及随机因素^[10]。

而基于强化学习算法的库存决策研究不依赖高精度的需求预测模型，且可充分考虑库存决策过程中的多阶段序贯决策和不确定性。库存决策是一个典型的多阶段序贯决策问题，Stranieri等^[11]将需求表示为余弦函数来模拟季节性行为，通过实验比较VPG^[12]、A3C^[13]、PPO三种算法在两阶段供应链库存决策问题下的性能，最终得出PPO更能适应不同环境的结论，且PPO的平均利润也高于其他算法。毕文杰等^[14]针对生鲜产品的易逝性提出了基于深度强化算法的生鲜产品联合库存控制与动态定价方法，通过DQN、DDPG、Q-learning等算法进行对比实验，结果表明基于DQN算法的收益表现最佳。

综上所述，考虑到产品需求复杂性的研究较少，而这是当前企业面临的现实问题。本文针对这些问题，基于深度强化学习来建模求解库存决策问题，以最大化期望总收益，并探讨了不同需求下各前沿强化学习算法的表现优劣。

2 模型构建

2.1 问题定义

汽车制造厂内的配件中心库是整个配件产业链的统筹计划中心，其存储了上万种可能用于自身品牌汽车售后维修保养的配件，企业投入了大量人力和财力用于其日常运营。本文对由制造厂的配件中心库和供应商组成的两阶段供应链库存决策问题进行建模，从供应商订购两种不同需求特性的配件，满足下级代理商的需求，如图2所示。



图2 供应链结构

假设供应商可以提供无限数量的配件，从供应商到仓库的交货时间为零。假定制造厂有法律义务履行所有订单，供应过程中允许缺货，但是一旦缺货，所欠的货必须补上。其中各数学符号含义如表1所示。

表1 数学符号含义说明

数学符号	说明
$a_{i,t}$	i 类配件的订购数量
$n_{i,t}$	i 类配件 t 时间步的库存水平
IC_i	i 类配件的最大库存容量
c_i	i 类配件的单位订购成本
c_i^S	i 类配件的单位存储成本
c_i^P	i 类配件的单位罚款成本
p_i	i 类配件的售价
$d_{i,t}$	i 类配件 t 时间步的需求

为了更贴合现实环境，需求采用每个时间段的均值(λ)的泊松分布进行建模。连续性需求和非连续性需求的详细设置将在第4节介绍。实验中考虑1080天的需求值，每15天为一个时间步，故有72个时间步。连续性需求和非连续性需求配件的需求值的某个实例如图3所示。

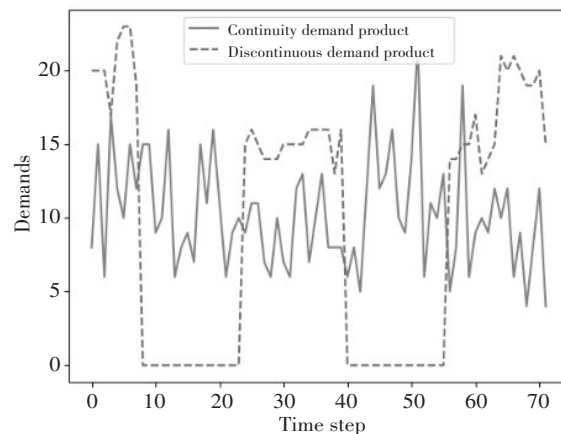


图3 不同需求特性配件需求值的某实例

2.2 马尔可夫决策过程

强化学习将库存控制这类序贯决策问题建模为马尔可夫决策过程并进行求解。本文通过马尔可夫决策过程将待求解问题定义为一个四元组 $M = (S, A, P, R)$ ，本文定义马尔可夫决策过程的各要素如下：

状态空间：状态向量包括仓库的每种配件类型的当前库存水平，加上历史 x 个时间步的需求值，并定义如下：

$$s_t = (n_{0,t}, n_{1,t}, \dots, n_{i,t}, d_{t-x}, \dots, d_{t-1}) \quad (1)$$

其中 $d_{t-1} = (d_{0,t-1}, d_{1,t-1}, \dots, d_{i,t-1})$ 。当前 t 时间步的实际需求 d_t 要到下一个 $t+1$ 时间步才能知道。类似于由 Kemmer 等^[15]所做的工作，在后面加上需求值是为了使智能体可以从需求历史中学习，从而将一种需求预测直接集成到策略中。

动作空间：通过神经网络直接生成动作值实现了一个连续的动作空间，包括在供应商采购每种配件类型 i 的数量：

$$a_t = (a_{0,t}, a_{1,t}, \dots, a_{i,t}) \quad (2)$$

其中每个值的下界都是零，上限是仓库对每种配件类型的最大容量，也即 $(0 \leq a_{i,t} \leq IC_i)$ 。

状态转移：下一状态的更新规则定义为

$$s_{t+1} = \left(\min \left[\left(n_{0,t} + a_{0,t} - d_{0,t} \right), IC_0 \right], \dots, \right. \\ \left. \min \left[\left(n_{i,t} + a_{i,t} - d_{i,t} \right), IC_i \right], d_{t+1-x}, \dots, d_t \right) \quad (3)$$

在下一个时间步开始时，制造厂的库存等于初始库存加上订购的数量，减去当前的需求。最后，更新包含在状态向量中的需求值。

奖励函数：每个时间步 t 的奖励函数定义如下：

$$r_t = \sum_{i=0}^i p_i \cdot d_{i,t} - \sum_{i=0}^i c_i \cdot a_{i,t} - \sum_{i=0}^i c_i^S \cdot \max(n_{i,t}, 0) + \\ \sum_{i=0}^i c_i^P \cdot \min(n_{i,t}, 0) \quad (4)$$

等式右边的第一项是收入，第二项是订购成本，第三项是存储成本，实现 $\max$ 函数是为了避免负库存(积压订单)被计算在内。最后一项表示惩罚成本，它是用加号引入的，因为在不满足订单的情况下，库存水平已经为负。因此，智能体的目标是利润最大化。

3 算法设计

3.1 状态抽象

强化学习中的状态抽象是指将环境状态映射到一个更低维的表示形式的过程。具体来说，

在库存决策环境中进行状态抽象指的是对库存状态中与订单决策不相关的状态值进行映射，以达到状态空间压缩的效果。

本文针对库存决策的环境提出一种基于状态抽象的状态归一化模块，该模块设计了一个针对库存决策的状态抽象函数，使智能体能以更快的速度做出决策。基于订单数量仅与一定范围内的库存水平相关的思想，本文设计了如下的状态抽象函数：

$$f(s) = \begin{cases} -1, & s < s_l \\ 2 \frac{s - s_l}{s_h - s_l} - 1, & s_l \leq s \leq s_h \\ 1, & s > s_h \end{cases} \quad (5)$$

其中 s 表示环境的状态向量，状态向量分为库存水平和历史需求两部分，因此状态向量的下限也分为两部分，其中，需求下限为 0，当积压的订单数大于制造厂每个 t 时间步订购的数量 a_t 时，智能体都需要最大化订单数量，因此状态向量的下限 $s_l = (-a_t^{\max}, 0)$ 。需求的上限 d_h 取决于需求的具体分布，当库存水平高于两倍需求的上限时，决策系统都需要将订单数量设为 0，以减小库存成本，因此状态向量的上限 $s_h = (2d_h, d_h)$ 。通过该状态抽象函数可以将原始状态空间 S 映射到新的状态空间 $S' = [-1, 1]$ 中。

3.2 算法选取及设计

强化学习方法应用于库存管理领域已经有一定研究，Q-learning 算法与 DQN 算法应用较为广泛。DQN 算法适用于离散动作空间，为了适用于连续动作空间，产生了 DDPG 算法。和 DDPG 相比，SAC 算法使用的是随机策略，目标是最大化带熵的累积奖励，使其具有更强的探索能力，且更具有鲁棒性，面对干扰的时候能更容易做出调整。

结合第 2 节所设的库存决策环境的 SAC 算法流程如下：

输入：环境

使用参数 θ_1 、 θ_2 、 ϕ 初始化策略网络 π_ϕ 、第一个评价网络 Q_{θ_1} 和第二个评价网络 Q_{θ_2} 。初始化经验池 B ，设置其大小为 N 。

对于每一幕执行：

- (1) 初始化环境和状态抽象后的状态 S_0 ;
- (2) 从 $t = 1$ 到 $t = T$, 执行以下操作:
 - ① 观察状态 S 从策略网络中选择动作 A 。
 - ② 执行所选动作 A , 观察下一个状态 S' 和奖励 r 。
 - ③ 在经验池中存储四元组 (S, A, R, S') 。
 - ④ 如果经验池中存储经验的数量大于批次大小, 执行以下操作:
 - (a) 从经验池中随机采样经验批次 $(S_i, A_i, R_i, S'_i), \forall i = 1, \dots, N$ 。
 - (b) 使用两个评价网络中最小的 Q 值来计算目标值, 利用梯度下降算法最小化评估值和目标值之间的误差, 从而更新两个评价网络 Q_θ 。
 - (c) 通过梯度上升算法更新策略网络 π_θ 。自动调整熵的系数 α , 让策略的熵保持一种适合训练的动态平衡。
 - (d) 引入动量 τ , 用软更新稳定训练, 更新目标网络 $\theta'_i \leftarrow \tau \theta_i + (1 - \tau) \theta'_i$ 。

SAC 算法是异策略的随机策略梯度算法, 在异策略的确定性策略梯度算法中, 本文采用了相对于 DDPG 算法优化的 TD3 算法进行对比实验。此外, 同策略算法中也采用了较先进的 PPO 算法进行对比实验。

4 仿真实验

引入了深度神经网络的算法模型的超参数一般需要手动设置, 其中, 学习率和 gamma 值是深度强化学习模型训练中非常重要的参数, 影响到智能体训练的稳定性和收敛性。因此, 本文通过控制变量法, 分别测试了学习率 $\alpha = 0.005$ 、 $\alpha = 0.001$ 、 $\alpha = 0.0005$ 的情况, 最终选定 $\alpha = 0.0005$ 。固定其他参数, 分别测试了 gamma 值 $\gamma = 0.9$ 、 $\gamma = 0.95$ 、 $\gamma = 0.99$ 的情况, 最终选定 $\gamma = 0.95$ 。由于篇幅有限, 具体效果图不做展示。

本文设计并进行了一组丰富的实验, 分不同需求特性设定了三个场景, 比较 PPO、TD3、SAC 三种深度强化学习算法在不同场景下的性能, 以此分析算法在仿真库存决策环境中的表现, 并判断算法能否应用于现实环境。实验中对 i 类配件的通用参数常量设置如表 2 所示。

表 2 实验环境常量

数学符号及含义	连续性需求配件	非连续性需求配件
售价, p_i	20	10
单位订购成本, c_i	8	4
单位存储成本, c_i^s	2	1
单位惩罚成本, c_i^p	10	5
仓库最大容量, IC_i	60	50

4.1 连续性需求场景

本场景下的连续性需求配件的需求服从于均值为 10 的泊松分布, 结合上述参数设置进行实验, 对比分析 TD3、SAC、PPO 三种深度强化学习算法在连续性需求下的性能。实验结果如图 4 所示, 在连续性需求场景下, SAC 和 TD3 算法的最终奖励都比 PPO 算法高, 且 SAC 算法和 TD3 算法都比 PPO 算法收敛稳定。SAC 算法比 TD3 算法先收敛。尽管 TD3 算法比 SAC 算法先达到较高的奖励, 但 TD3 算法前期比较震荡。

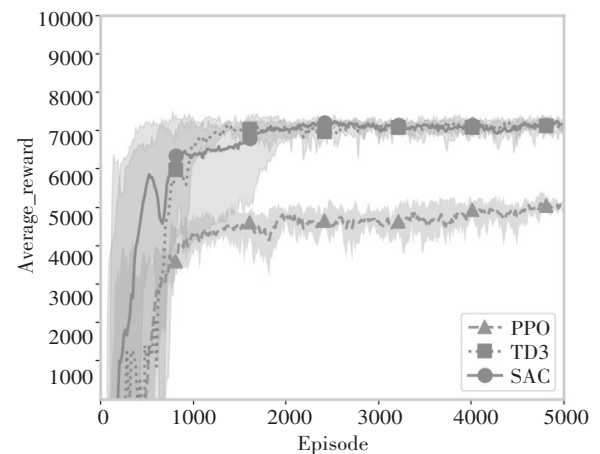


图 4 TD3、SAC、PPO 在连续性需求下的实验对比结果

4.2 非连续性需求场景

本场景下的非连续性需求配件的需求依然服从于泊松分布, 均值每 120 天从 $[10, 15, 20, 0]$ 中等概率随机选取。本研究结合上述参数设置进行实验, 对比分析了 TD3、SAC、PPO 三种深度强化学习算法在非连续性需求下的性能。实验结果如图 5 所示, 在非连续性需求场景下, SAC 和 TD3 算法都比 PPO 算法奖励高, 且 SAC 算法和 TD3 算法都比 PPO 算法收敛稳定。SAC

算法比TD3算法先收敛也先达到较高的奖励。

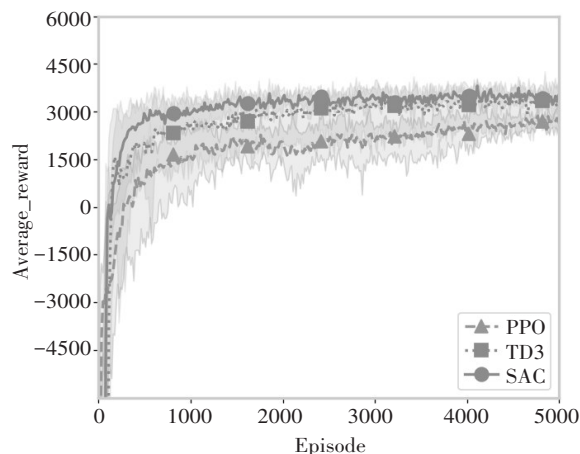


图5 TD3、SAC、PPO在非连续性需求下的实验对比结果

4.3 混合连续性及非连续性需求场景

本场景下配件的需求结合前两节的需求设置,奖励为这两类配件的利润之和。本研究结合上述参数设置进行实验,对比分析TD3、SAC、PPO三种深度强化学习算法在混合连续性需求和非连续性需求下的性能。实验结果如图6所示,在混合连续性需求和非连续性需求场景下,SAC算法达到的奖励最高,最先收敛,稳定性也最强;TD3算法其次,PPO算法最后。

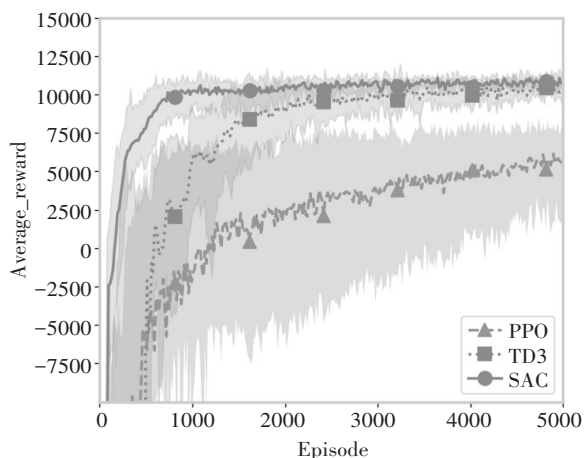


图6 TD3、SAC、PPO在混合连续性及非连续性需求下的实验对比结果

综合上述三幅图可得,SAC算法的收益表现和稳定性无论在连续性需求、非连续性需求,还是混合连续性及非连续性需求上表现都最佳,其次是TD3算法,PPO算法的收益最低。

现实中的需求变化极其复杂,各种因素导致需求呈现随机性强、非稳态的波动变化。SAC算法能够解决动态需求问题,也能为企业提供近似最优的订货策略。由此可见,基于SAC算法的库存决策模型具有非常广泛的应用价值。

5 结语

目前汽车后市场规模相应扩大,配件需求也逐年递增,对于制造厂而言如何合理控制库存是非常重要的决策问题。本文研究了汽车配件的库存决策问题,根据配件的需求特性,分为连续性需求配件和非连续性需求配件并制定了三个场景。通过将库存决策问题建模为马尔科夫决策过程,本文在更现实的实验下评估了多个深度强化学习算法的有效性和稳健性。本文还设计了一个针对库存决策的状态抽象函数,从而有效减少无用状态信息,简化库存决策环境,提高决策效率。

本文改进的深度强化学习算法,可以根据当前可用库存来动态调整订货量,使预期利润最大化。然而,由于条件限制,本研究只关注了两阶段供应链的库存决策问题,未考虑运输及订购提前期等因素。未来可以在考虑运输和订购提前期等情况下的库存决策方面进行深入的研究。

参考文献:

- [1] 公安部. 全国机动车保有量达4.17亿辆驾驶人超过5亿人[EB/OL]. [2023-01-15]. http://www.gov.cn/shuju/2023-01/11/content_5736278.htm.
- [2] 申利俊. 汽车售后配件安全库存研究[D]. 秦皇岛:燕山大学,2020.
- [3] 高阳,陈世福,陆鑫. 强化学习研究综述[J]. 自动化学报,2004,30(1):86-100.
- [4] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Humanlevel control through deep reinforcement learning[J]. Nature,2015,518:529-533.
- [5] 杨思明,单征,丁煜,等. 深度强化学习研究综述[J]. 计算机工程,2021,47(12):19-29.
- [6] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[EB/OL]. [2023-01-15]. arXiv:1509.02971,2015.
- [7] SCHULMAN J, WOLSKI F, DHARIWAL P, et al.

- Proximal policy optimization algorithms [EB/OL]. [2023-01-15]. arXiv:1707.06347, 2017.
- [8] FUJIMOTO S, VAN HOOFF H, MEGER D. Addressing function approximation error in actor-critic methods[EB/OL]. [2023-01-15]. arXiv:1802.09477, 2018.
- [9] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor[EB/OL]. [2023-01-15]. arXiv:1801.01290, 2018.
- [10] 徐翔斌,李志鹏. 强化学习在运筹学的应用:研究进展与展望[J]. 运筹与管理, 2020, 29(5): 227-239.
- [11] STRANIERI F, STELLA F. A deep reinforcement learning approach to supply chain inventory management[EB/OL]. [2023-01-15]. arXiv:2204.09603v2, 2022.
- [12] WILLIAMS R J. Simple statistical gradient-following algorithms for connectionist reinforcement learning [J]. Reinforcement Learning, 1992, 8(3/4): 229-256.
- [13] MNIH V, BADIA A P, MIRZA M, et al. Asynchronous methods for deep reinforcement learning [C] //Proceedings of the International Conference on Machine Learning, New York, USA: ACM Press, 2016:1928-1937.
- [14] 毕文杰,周玉冰. 基于深度强化学习的生鲜产品联合库存控制与动态定价研究[J]. 计算机应用研究, 2022, 39(9): 2660-2664.
- [15] KEMMER L, VON KLEIST H, DE ROCHEBOUËT D, et al. Reinforcement learning for supply chain optimization [C] //Proceedings of the 14th European Workshop on Reinforcement Learning, Lille, France, 2018.

Research of inventory decision for auto parts based on deep reinforcement learning

Li Yuqing^{1*}, Pan Chaofan²

(1. Sichuan Key Laboratory of Manufacturing Industry Chain Collaboration and Information Support Technology, Southwest Jiaotong University, Chengdu 611756, China;

2. School of Computer Science, Southwest Petroleum University, Chengdu 610500, China)

Abstract: The aim of this study is to address the inventory decision-making problem in the factory warehouse for auto parts within the supply chain. The study takes into consideration the inventory constraints and the complex and changeable characteristics of auto parts' demand. To model the sequential decision-making problem, such as inventory decision-making, the study utilizes a Markov decision-making process. Furthermore, the study applies deep reinforcement learning to solve the inventory decision-making model. The goal is to explore the optimal order quantity of auto parts with varying demand characteristics under uncertain demand conditions, so as to maximize the enterprise's profit. The experimental results demonstrate that the inventory decision based on deep reinforcement learning can effectively improve the enterprise's profit while accurately fitting the actual situation and offer high practical value.

Keywords: deep learning; reinforcement learning; inventory decision

文章编号: 1007-1423(2023)11-0022-06

DOI: 10.3969/j.issn.1007-1423.2023.11.004

基于 Transformer 和 CNN 的真实场景下植物病害识别方法

刘 畅^{1,2}, 莫海芳^{1,3*}, 马 春^{1,2}

(1. 中南民族大学计算机科学学院, 武汉 430074; 2. 农业区块链与智能管理湖北省工程研究中心, 武汉 430074;
3. 湖北省制造企业智能管理工程技术研究中心, 武汉 430074)

摘要: 针对真实场景下的植物病害图像, 设计了一种融合 Transformer 和 CNN 的植物病害识别模型 CLT。该模型采用多阶段层次设计, 以 Conv Stem 作为初始特征提取模块提取浅层局部特征, 结合 Transformer 模块以学习全局特征; 利用卷积 Token 嵌入, 改变模型每个阶段 Token 的序列长度和特征维数, 实现对多层次的局部空间上下文进行建模; 将卷积模块融入到 Transformer 模块, 并将多头注意力机制中的线性投影替换为卷积投影, 提升模型对局部空间的特征提取。实验结果表明, CLT 模型能够有效表达植物病害的各种特征, 在真实场景下平均准确率达到 77.91%, 分类效果优于其他模型, 为真实场景下植物病害识别提供参考。

关键词: 植物病害识别; Transformer; CNN; 真实场景

0 引言

植物病害影响农产品的质量和产量, 是造成农业经济损失的主要原因之一。通过人工判断植物病害的方式复杂、耗时且受主观因素影响, 因此需要一种自动、快速、准确的植物病害识别方法来保证植物的品质和产量。

近年来, 基于卷积神经网络(convolutional neural networks, CNN)的模型在图像识别领域取得了巨大的成功, 这些模型能够作为骨干模型, 自动从图像中提取特征。目前常见的卷积神经网络模型有 AlexNet^[1], VGG^[2], InceptionNet^[3] 和 ResNet^[4]等。很多研究人员利用卷积神经网络模型替代人工识别的方法, 在植物病害识别领域进行研究。Lyv 等^[5]针对 AlexNet 模型在复杂的自然环境中难以获取玉米病害特征的问题, 使用膨胀卷积保留特征的细节信息, 使用多尺度卷积增加特征的多样性, 取得了 98.62% 的准

确率。鲍文霞等^[6]将 ResNet 模块和 Inception 模块结合, 增大模型的特征映射区域, 识别准确率达到 98.73%。何欣等^[7]为了进一步准确识别葡萄叶片的病害程度, 在 ResNet 中引入多尺度卷积, 识别准确率达到 90.83%。

以上研究中, 模型都取得了很高的准确率, 但都是针对单一物种的数种病害, 数据集中的病害种类少且没有不同植物同种病害的情况, 应用场景受限, 难以适应复杂的农业生产环境。为此, 有学者针对非单一物种的植物病害识别进行研究, 发现模型在真实场景下的识别精度不高。Ferentinos^[8]使用 Plant Village 数据和自制的真实环境数据, 分别对模型进行训练, 实验表明 AlexNet 与 VGG 模型使用实验环境的病害图像进行训练并识别真实场景下的病害图像时准确度仅为 32.23% 和 33.27%。王东方等^[9]提出基于深度残差网络和迁移学习的农作物病害识别

收稿日期: 2023-02-17 修稿日期: 2023-03-09

基金项目: 国家民委中青年人才培养计划(MZR20007); 新疆维吾尔自治区区域协同创新专项(科技援疆计划)(2022E02035)

作者简介: 刘畅(1996—), 男, 湖北武汉人, 硕士生, 研究方向为图像处理和深度学习; *通信作者: 莫海芳(1974—), 女, 广西宜州人, 博士, 副教授, 硕导, 研究方向为智能计算、机器视觉等, E-mail: 278415737@qq.com; 马春(1997—), 女, 云南曲靖人, 硕士生, 研究方向为图像处理和深度学习;

模型, 在真实农业场景数据集 PlantDoc 上平均识别率达到 47.37%。李进^[10]提出一种基于多尺度的残差网络, 在 PlantDoc 数据集上最高识别准确率达到 64.86%。以上的研究均采用 CNN 架构模型, 这些模型会更多地关注局部特征信息, 缺乏对远程依赖关系的提取, 模型泛化能力弱, 导致对复杂背景的植物病害图像识别不够精准。因此, 增强模型的泛化能力, 有助于提高真实场景下植物病害图像的准确识别率。

TransFormer^[11]在视觉任务上也取得了巨大成果。该模型是一种非局部模型^[12], 不使用卷积核, 对于长距离特征的提取能力较强, 并且具有较强的泛化能力^[13-15]。Dosovitskiy 等^[16]提出的 ViT 使用图像块(patch)作为图像分类的输入, 首次证明了 Transformer 模型可以达到与 CNN 同样的图片分类精度。于是, 基于 Transformer 提取全局特征的优势和 CNN 提取局部特征的优势, 有学者提出将两种模型结构结合起来。谷歌研究人员提出的 BotNet^[17]利用 Transformer 中的多头注意力模块去替换 Res-Net 中的 3×3 卷积模块; Peng 等^[18]提出的 Conformer 混合 Transformer 和 CNN 两种结构, 融合了不同分辨率下的全局特征和局部特征; Wang 等^[19]提出的 PVT 将 Transformer 模型设计为渐进的金字塔结构来实现较大的分辨率输出, 同时减少特征图的计算。

为了提升模型的泛化能力, 提高真实场景下的植物病害识别准确率, 本文设计了一种基于 Transformer 和 CNN 的深度学习模型 CLT。CLT 利用卷积改进 Token 嵌入, 并将模型设计为多阶段渐进式结构, 降低 Token 的长度, 同时增加了 Token 的维度, 从而保证了模型容量。同时, 在 Transformer 模块中, 融入卷积模块以增强模型的局部特征提取能力。

1 模型设计

1.1 模型整体结构

CLT 模型融合了视觉 Transformer 模块和 CNN 模块, 并利用卷积操作增强 Transformer 模块的局部感知能力, 结合 Transformer 架构在全局特征提取上的优势, 提升对真实场景下植物病害图像的特征提取能力。CLT 模型在 Transformer 模块中融入一个带残差的深度可分离卷积模块和一个卷积投影模块, 并借鉴 CNN 中多阶

段层次的设计思想, 将模型设计为三个阶段, 整体结构如图 1 所示。

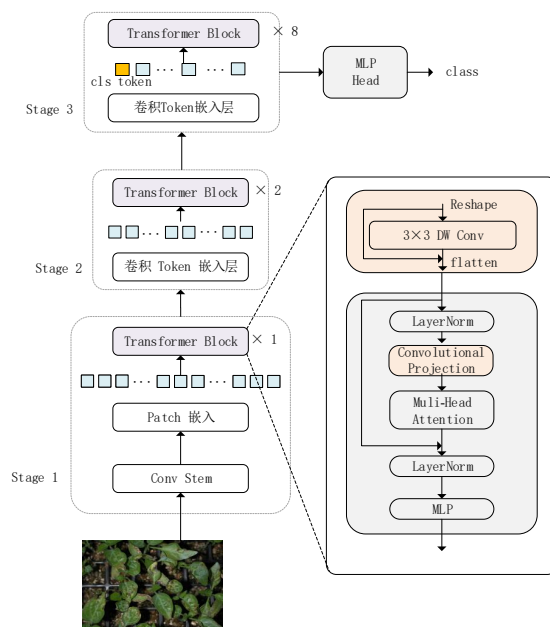


图 1 CLT 模型整体架构

第一阶段, 首先图像输入到 Conv Stem 模块, 如图 2 所示。Conv Stem 由 4 个卷积模块构成。前三个模块使用卷积核大小为 3×3 的卷积提取图像的浅层特征, 使模型能够捕捉到更多局部细节, 其中第一个模块的步长设置为 2, 起到对输入图像维度缩减的作用。最后一个模块使用卷积核大小为 1×1 卷积, 用于加强通道之间的信息交互。然后输出的特征图经过分块操作, 输入到 Transformer 模块中。

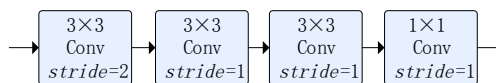


图 2 Conv Stem 模块

第二阶段, 将第一阶段输出的 Token 序列重塑为 Token map, 经过卷积 Token 嵌入层, 对 Token map 进行卷积操作, 减少 Token 的数量, 同时增加 Token 的维度, 从而实现空间下采样。然后输出的 Token 序列直接输入到 Transformer 模块中。

第三阶段, 和第二阶段的设计基本一样, 不同的是在卷积 Token 嵌入层后, 为新的 Token

添加一个 class Token 用于进行分类。最后，输出的分类标记通过 MLP 来预测结果。

1.2 融入卷积的 Token 嵌入层

在传统的视觉 Transformer 中，直接将输入的图像切分成大小相同的块，然后将 2D 的分块图像展平为 Token 向量作为模型的输入，图 3(a) 展示了 ViT 中通过 patch 嵌入获取 Tokens 的过程。这种简单的切分方式会导致图像的边缘信息丢失。CLT 通过对上一个阶段得到的 Token 序列进行卷积运算，得到下一个阶段的 Token 序列，充分利用图像的边缘和线条等局部结构信息。整个过程如图 3(b) 所示，图中的 Token map 都是由 Token 序列重塑所得，第 $i-1$ 阶段得到的 Token map 记为 $x_{i-1} \in \mathbf{R}^{H_{i-1} \times W_{i-1} \times C_{i-1}}$ ，这里通过一个卷积函数 $f(\cdot)$ 将 x_{i-1} 映射到一个新的通道大小为 C_i 的 Token map，其中 $f(\cdot)$ 卷积核大小为 $k \times k$ ，padding 为 p 。那么新得到的 Token map $f(x_{i-1}) \in \mathbf{R}^{H_i \times W_i \times C_i}$ ，其宽度和高度为

$$\begin{cases} H_i = \left\lfloor \frac{H_{i-1} + 2p - k}{stride} + 1 \right\rfloor \\ W_i = \left\lfloor \frac{W_{i-1} + 2p - k}{stride} + 1 \right\rfloor \end{cases} \quad (1)$$

再将这个 Token map 展平成 $H_i W_i \times C_i$ ，就得到了第 i 阶段的 Token 序列。

卷积 Token 嵌入层允许通过改变卷积运算的参数来调整每个阶段的 Token 特征维数和数量。通过这种方式，在每个阶段逐步减少 Token 序列长度，同时增加 Token 特征维数。这使得 Token 能够在越来越大的空间上表示越来越复杂的视觉模式，类似于 CNN 的特征层。

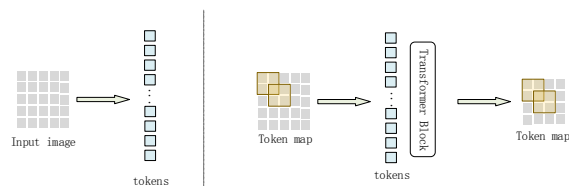


图 3 传统的 Token 嵌入和融入卷积的 Token 嵌入

1.3 融入卷积的 Transformer 模块

为了增强模型的泛化能力，减少背景干扰，提升模型真实场景下病害特征的提取能力，CLT

在 Transformer 模块中融入两个卷积模块，一个带残差的深度可分离卷积模块和一个卷积投影模块，详细说明如下。

1.3.1 带残差的深度可分离卷积模块

在原始的视觉 Transformer 模块中通常采用绝对位置编码存储图片中像素的空间位置关系，这导致模型对图像中目标的旋转、平移、缩放不敏感。而 CNN 对局部特征提取能力强，目标出现的位置不影响模型的输出，具有平移不变性。因此，本文在 Transformer 模块中引入 3×3 的深度分离卷积 (DW conv)，用于增强网络的平移不变性，并利用残差连接稳定网络训练，如图 1 中 Transformer 模块所示。考虑到模型中堆叠了多个 Transformer 模块，为了不增加过多的运算参数，这里使用深度分离卷积。

1.3.2 卷积投影

Transformer 模块中的多头注意力模块 (MHA) 中 $Q/K/V$ 矩阵的 Token 输入采用的是线性投影。本文在 MHA 前面添加一层卷积操作，将线性投影替换为卷积投影，实现对局部空间上下文的额外建模，如图 4 所示。首先将 Token 重塑为 Token map。然后使用卷积核大小为 3 的深度可分离卷积来实现 $Q/K/V$ 矩阵的映射，计算方式如下：

$$x_i^{Q/K/V} = \text{Flatten}(\text{Conv2d}(\text{Reshape2D}(x_i, k))) \quad (2)$$

其中： $x_i^{Q/K/V}$ 表示第 i 阶段 $Q/K/V$ 矩阵的 Token 输入； x_i 表示卷积投影之前 Token。Conv2d 这里同样采用深度可分离卷积，其卷积核大小为 k 。

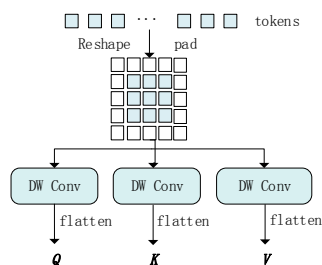


图 4 卷积投影层结构

2 实验与分析

2.1 数据集

PlantDoc^[20]数据集是印度理工学院的研究人

员于2019年秋季发布的用于图像分类和对象检测的数据集。该数据集花费了300多个人工小时收集和标注，包含13个植物物种和27个类别（17个疾病；10个健康）的2598张图像。本文将网络爬取的真实场景下的病害图片和数据集PlantDoc融合构建扩充的PlantDoc数据集，扩充的PlantDoc数据集与PlantDoc具有相同的植物物种和类别。共计3171张病害图片。通过此数据集评价各个模型对真实环境下的植物病害识别性能。图5给出了该数据样本的展示图。



图5 扩充后的PlantDoc植物病害图片示例

2.2 实验环境与参数设置

2.2.1 实验环境

实验环境如表1所示，GPU配置为Nvidia Geforce RTX 3090 24 GB，并使用CUDA计算框架进行加速，实验平台为Centos 64位，处理器为Intel(R) Xeon(R) Silver 4110，并以PyTorch为深度学习框架，Python 3.8为编程语言。

表1 实验环境

配置	型号
GPU	Nvidia Geforce RTX 3090 24 GB
CPU	Intel(R) Xeon(R) Silver 4110
编程语言	Python 3.8
深度学习框架	PyTorch 1.10.0
GPU加速库	CUDA 10.2

2.2.2 参数设置

参数设置对模型的训练有着重要的影响。本实验对学习率、每个批次病害图片的数量

(batch size)、损失函数和优化算法进行统一设置。学习率的大小影响着模型的收敛速度，本实验将学习率设置为0.0001。Batch Size会影响模型的性能和速度，为了提升GPU的使用效率，平衡内存和算力，使用不同的Batch Size进行测试，最终选定为128。实验选择交叉熵损失函数，优化算法使用自适应矩估计(adaptive moment estimation, Adam)。相较于广泛使用的随机梯度下降算法(stochastic gradient descent, SGD)，Adam的计算效率更高。

2.3 实验结果与分析

2.3.1 不同模型对比实验

为验证本文提出的模型CLT在真实场景下的植物病害识别效果，选取InceptionV3、ResNet50、ViT和Swin Transformer作为对比模型，在扩充后的PlantDoc数据集上训练并进行实验对比。其中，InceptionV3和ResNet50为CNN架构的模型，ViT和Swin Transformer为Transformer架构的模型。

为了评价各个模型的性能，结合植物病害识别分类的特点，本实验选择平均准确率(Average accuracy rate)和加权F1(Weighted-F1)作为模型性能的评价指标。各模型均在PlantVillage上进行预训练，然后迁移到PlantDoc上迭代300次。

本文提出的CLT在迭代50次以后，已经取得了优于其他模型的平均准确率，且在后续训练中模型的准确率仍在缓慢提升，如图6所示。

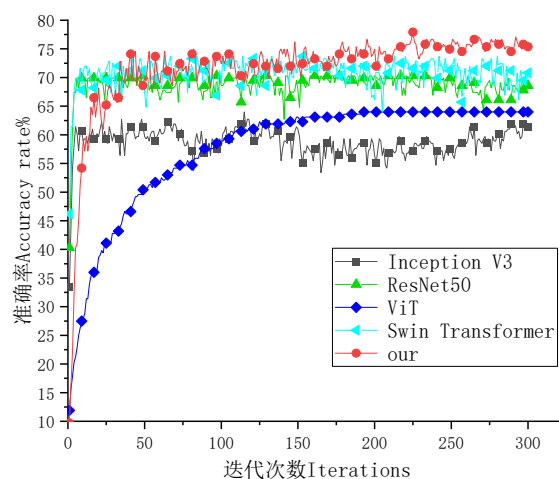


图6 不同模型平均识别率对比结果

虽然 CLT 的收敛速度略慢于其他模型,但是取得了最好的平均准确率,这说明融合 CNN 和 Transformer 结构提升了模型的泛化性能,能更好地识别真实场景下的植物病害。

表 2 展示了不同植物病害识别模型在测试集上的平均准确率和加权 F1 值。各模型的平均准确率在 60%~80% 之间,说明深度学习模型在真实场景下的植物病害分类性能还有很大提升空间。其中,Transformer 架构的模型比 CNN 架构的模型表现的更好,说明全局特征提取有助于提升模型在复杂背景下的病害特征提取。本文所提出的 CLT 模型结合 Transformer 和 CNN 对不同距离特征的提取优势,突出植物病害信息,抑制无用背景信息,在测试集上平均准确率达到 77.91%,分类效果优于其他模型。同时加权 F1 值也高于其他模型,证明 CLT 对不同类别的植物区分能力更强,更适合多物种的病害识别场景。

表 2 不同模型识别性能对比

模型	平均准确率/%	加权 F1/%
InceptionV3	62.06	61.17
ResNet50	70.30	68.12
ViT-base	64.00	60.58
Swin Transformer	73.70	70.62
CLT	77.91	73.15

2.3.2 消融实验

对 CLT 的 Transformer 模块中添加的 DW conv 和卷积投影部分进行消融实验,实验均在扩充后的 PlantDoc 数据集上进行。表 3 中 CLT_No_DW 表示在 Transformer 模块中不使用带残差的深度可分离卷积模块。CLT_No_ConvProj 表示在 Transformer 模块中,多头注意力机制中的 $Q/K/V$ 矩阵使用线性投影而非卷积投影获取。在不使用 DW Conv 时,模型的平均准确率为 75.53%,对比 CLT 下降了 2.38 个百分点。而不使用卷积投影的 CLT 模型的平均准确率为 74.95%,下降了 2.96 个百分点。证明了融入卷积的 Transformer 模块能更好地提取病害图像中的特征,使用卷积投影的多头注意力机制能更好地对局部空间上下文进行建模,使得模型在真实场景下的植物病害识别能力得到进一步的提升。

表 3 局部感知单元 LPU 消融实验

方法	平均准确率/%	加权 F1/%
CLT	77.91	73.15
CLT_No_DW	75.53	71.36
CLT_No_ConvProj	74.95	70.67

3 结语

本文基于 CNN 模型与 Transformer 模型,提出了一种植物病害分类模型 CLT,用于不指定具体植物种类的病害检测,避免了传统方法繁琐的人工特征设计,提升模型在真实场景下病害的检测分类性能。CLT 以 Conv stem 作为初始特征提取模块提取浅层局部特征,结合 Transformer 模块以学习全局特征。通过卷积的方式改进 Token 嵌入,调整每个阶段 Token 的序列长度和特征维数,结合模型的多阶段结构设计,增强模型空间局部特征提取能力。同时,在 Transformer 模块中插入卷积层并在多头注意力中使用卷积投影,增加模型的平移不变性和局部感知能力。通过在扩充后的 PlantDoc 数据集上实验,CLT 获得了 77.91% 的平均准确率,高于单独的 CNN 网络和 Transformer 网络,并通过消融实验证明了融合模型的有效性,更适合真实场景下的植物病害分类任务。

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: 1409.1556, 2014.
- [3] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 2015: 1-9.
- [4] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016: 770-778.
- [5] LV M, ZHOU G, HE M, et al. Maize leaf disease identification based on feature enhancement and DMS-robust alexnet [J]. IEEE Access, 2020, 8:

- 57952-57966.
- [6] 鲍文霞, 黄雪峰, 胡根生, 等. 基于改进卷积神经网络模型的玉米叶部病害识别(英文)[J]. 农业工程学报, 2021, 37(6): 160-167.
 - [7] 何欣, 李书琴, 刘斌. 基于多尺度残差神经网络的葡萄叶片病害识别[J]. 计算机工程, 2021, 47(5): 285-291.
 - [8] FERENTINOS K P. Deep learning models for plant disease detection and diagnosis [J] Computers and Electronics in Agriculture, 2018, 145: 311-318.
 - [9] 王东方, 汪军. 基于迁移学习和残差网络的农作物病害分类[J]. 农业工程学报, 2021, 37(4): 199-207.
 - [10] 李进. 基于卷积神经网络的多种作物病害识别研究[D]. 镇江: 江苏大学, 2022.
 - [11] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] //Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA, 2017: 6000-6010.
 - [12] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local neural networks [C] //Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018: 7794-7803.
 - [13] ZHANG C, ZHANG M, ZHANG S, et al. Delving deep into the generalization of vision transformers under distribution shifts [C] //Proceedings of the Conference on Computer Vision and Pattern Recognition, New Orleans, Louisiana, USA, 2022: 7277-7286.
 - [14] FANG Y X, LIAO B C, WANG X G, et al. You only look at one sequence: rethinking transformer in vision through object detection [EB/OL]. arXiv: 2106.00666, 2021.
 - [15] NASEER M M, RANASINGHE K, KHAN S H, et al. Intriguing properties of vision transformers [C] //Proceedings of the Advances in Neural Information Processing Systems, 2021, 34: 23296-23308.
 - [16] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale [EB/OL]. arXiv: 2010.11929, 2020.
 - [17] SRINIVAS A, LIN T Y, PARMAR N, et al. Bottleneck transformers for visual recognition [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 2021: 16519-16529.
 - [18] PENG Z, HUANG W, GU S, et al. Conformer: local features coupling global representations for visual recognition [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 2021: 367-376.
 - [19] WANG W, XIE E, LI X, et al. Pyramid vision transformer: a versatile backbone for dense prediction without convolutions [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 2021: 568-578.
 - [20] SINGH D, JAIN N, JAIN P, et al. PlantDoc: a dataset for visual plant disease detection [C] //Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, Hyderabad, India, 2020: 249-253.

Plant disease identification method in real scenarios based on Transformer and CNN

Liu Chang^{1,2}, Mo Haifang^{1,3*}, Ma Chun^{1,2}

(1. School of Computer Science, South-Central MinZu University, Wuhan 430074, China;

2. Hubei Provincial Engineering Research Center of Agricultural Blockchain and Intelligent Management, Wuhan 430074, China;

3. Hubei Provincial Engineering Research Center for Intelligent Management of Manufacturing Enterprise, Wuhan 430074, China)

Abstract: A plant disease recognition model CLT incorporating Transformer and CNN is designed for plant disease images in real scenarios. the model adopts a multi-stage hierarchical design, using Conv Stem as the initial feature extraction module to extract shallow local features, combined with Transformer module to learn global features; using convolutional Token embedding to change the model The Convolutional Token embedding is used to change the sequence length and feature dimension of Token at each stage to model the multi-level local space context; the Convolutional module is incorporated into the Transformer module and the linear projection in the multi-headed attention mechanism is replaced with the Convolutional projection to enhance the model's feature extraction of the local space. The experimental results show that the CLT model can effectively express various features of plant diseases, with an average accuracy of 77.91% in real scenes, and the classification effect is better than other models, providing a reference for plant disease recognition in real scenes.

Keywords: plant disease identification; Transformer; CNN; real scenarios

文章编号: 1007-1423(2023)11-0028-05

DOI: 10.3969/j.issn.1007-1423.2023.11.005

基于粒子群算法的江西景点自驾游路线规划

宋艳艳^{1,2}, 李广宏¹, 朱 玲^{3*}, 夏传林³, 郑 巍³

(1. 桂林理工大学旅游与风景园林学院, 桂林 541000; 2. 南昌航空大学数学与信息科学学院, 南昌 330063;
3. 南昌航空大学软件学院, 南昌 330063)

摘要: 随着服务产业的快速增长, 每逢假期各个旅游景点都成为了度假胜地。越来越多的人喜欢自驾前往旅游景点, 如果想自驾去多个景点, 如何规划旅游路线让旅游路径花费的时间最短成为了一个难题。粒子群优化算法具有收敛速度快、参数少、算法简单易实现的优点, 通过不断迭代更新找到最优解。利用粒子群优化算法对江西的 13 个 5A 级景区进行了路径规划。通过百度拾取坐标系获得景点经纬度坐标, 进一步转换为距离进行计算, 从而得出途径 13 个旅游景点的最短路径, 并把导航的真实距离作为实验结果。通过实例证明, 粒子群优化算法在解决路径优化这类问题上的有效性。

关键词: 粒子群算法; 路径规划; 江西景点

0 引言

随着科技的发展, 通过携程、美团、飞猪等 APP 可以在网上查询各地的旅游景点, 十分方便人们的出行, 在一定程度上也促进了旅游业的发展。随着汽车的普及, 自驾游也越来越受到年轻人的喜爱, 人们会根据出行的时间合理地规划旅游行程, 选择合适的景点个数, 但是实际出行中, 还会伴随其他问题的产生, 例如旅游路径的规划^[1-2]、旅游费用等, 这个时候需要合理的方法解决这个问题。许多学者对这个问题进行研究, 最早源于 TSP 旅行商问题^[3], 随后对该问题的算法进行深入研究, 解决这类问题常用的方法有蚁群^[4-7]、粒子群^[8-11]、遗传算法^[12-15]等。本文结合粒子群算法, 对江西 13 个 5A 旅游景点进行路径规划, 粒子群算法可以为我们提供最佳的旅游顺序^[16]。

1 算法介绍

粒子群优化算法(particle swarm optimization, PSO)又称粒子群算法^[17], 是通过模拟鸟群觅食行为而发展起来的一种基于群体协作的随机搜索算法^[18-20], 主要应用在连续优化的问题中, 对于离散问题, PSO 应用在 TSP 旅行商问题是一个新的研究方向。

一群鸟在寻找食物, 在这个区域中只有一只虫子, 所有的鸟都不知道食物在哪。但是它们知道自己的当前位置距离食物有多远, 同时它们知道离食物最近的鸟的位置^[21-24]。如图 1 所示, 表示的是鸟群在觅食过程中的几种状态。小黑点表示一只鸟, 箭头指示的方向表示鸟群飞行的方向。图 1(a)表示鸟群的初始觅食状态, 鸟群都是分散开来的, 朝着区域内最可能找到食物的方向运动; 图(b)表示鸟群中间觅食的状态, 可以看到鸟群已经呈现出聚集状态, 分布

收稿日期: 2023-02-17 修稿日期: 2023-03-04

基金项目: 国家自然科学基金项目(61867004)

作者简介: 宋艳艳(1985—), 女, 甘肃庆阳人, 硕士, 研究方向为旅游规划、优化算法; 李广宏(1966—), 男, 河南睢县人, 硕士, 副教授, 研究方向为旅游资源开发与规划; *通信作者: 朱玲(1997—), 女, 江西赣州人, 硕士, 研究方向为优化算法, E-mail: 1429881125@qq.com; 夏传林(1995—), 男, 江西九江人, 硕士, 研究方向为优化算法; 郑巍(1982—), 男, 江西萍乡人, 博士, 教授, 研究方向为优化算法

在食物周围，找到局部最优解；图(c)表示鸟群的最最终觅食状态，鸟群分布在食物的周围，此时找到了最优解^[25]。

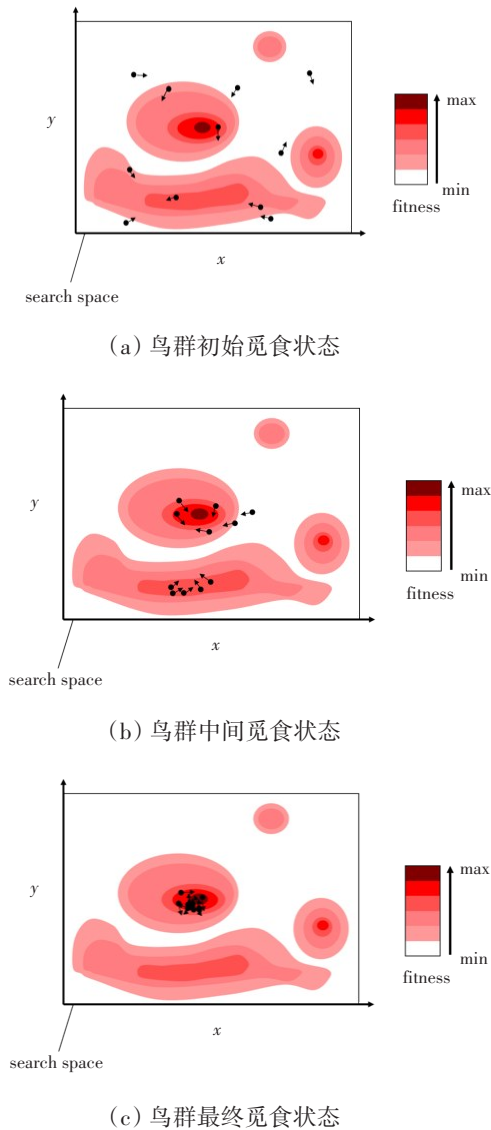


图1 鸟群觅食状态

图2描述了粒子群优化算法的运算流程，如下^[26]：

步骤1：初始化粒子群，包括群体规模 N ，每个粒子的位置 x_i 和速度 v_i 。

步骤2：计算每个粒子的适应度值 $fit(i)$ 。

步骤3：对每个粒子，用它的适应度值 $fit(i)$ 和个体极值 $p_{best}(i)$ 比较。如果 $fit(i) < p_{best}(i)$ ，则用 $fit(i)$ 替换掉 $p_{best}(i)$ 。第 i 个粒子迄今为止搜索到的最优位置称为个体极值，记为

$$p_{best} = (p_{i1}, p_{i2}, \dots, p_{iD}), i = 1, 2, \dots, N \quad (1)$$

(4) 对每个粒子，用它的适应度值 $fit(i)$ 和全局极值 g_{best} 比较。如果 $fit(i) < g_{best}$ ，则用 $fit(i)$ 替换掉 g_{best} 。整个粒子群迄今为止搜索到的最优位置为全局极值，记为

$$g_{best} = (g_1, g_2, \dots, g_D) \quad (2)$$

(5) 迭代更新粒子的速度 v_i 和位置 x_i 。根据公式(1)和(2)得到两个最优值 p_{best} 和 g_{best} ，可根据公式(3)和(4)来更新粒子的速度和位置。

$$v_{ij}(t+1) = v_{ij}(t) + c_1 r_1(t) [p_{ij}(t) - x_{ij}(t)] + c_2 r_2(t) [p_{gj}(t) - x_{ij}(t)] \quad (3)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad (4)$$

其中： c_1 和 c_2 称为学习因子，分别调节 p_{best} 和 g_{best} 方向飞行的最大步长； r_1 和 r_2 为 $[0,1]$ 范围内的均匀随机数。

(6) 进行边界条件处理。

(7) 判断算法终止条件是否满足：若是，则结束算法并输出优化结果；否则返回步骤(2)。

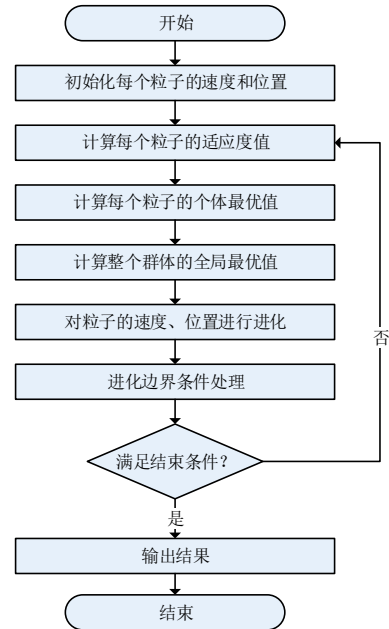


图2 粒子群优化算法的运算流程^[27]

2 实例验证

2.1 实例介绍

江西省是一个具有浓厚历史韵味和人文特色的著名旅游城市，本文选取江西省13个5A景区作为分析对象，其中包含自然景观、历史遗

址、现代建筑等相对具有代表性的旅游景点。其中包含滕王阁、庐山、庐山西海、龙虎山、三清山、龟峰、井冈山、江湾、明月山、共和国摇篮、武功山、大觉山、古窑民俗博览区 13 个景点。这些景点有的在南昌市内,有的在九江市,分布在江西省的各个地区。如果想去多个景点旅游,要是没有做好路线规划,会造成去过相同的路线,非常浪费时间。因此,本文重点对 13 个景区进行一个顺序排序,对景点进行路径规划,使得总路线最短。各个景点的位置坐标如表 1 所示,景点坐标是通过百度的拾取坐标系统获得的经纬度坐标,在后续实验计算中,需要将经纬度坐标转换为距离坐标。

表 1 江西省 13 个 5A 景区坐标

景点序号	景点名称	景点坐标
1	滕王阁	(28.69, 115.89)
2	庐山	(29.56, 116.00)
3	庐山西海	(29.20, 115.48)
4	龙虎山	(28.12, 116.99)
5	三清山	(28.91, 118.08)
6	龟峰	(28.32, 117.43)
7	井冈山	(26.64, 114.14)
8	江湾	(29.37, 118.06)
9	明月山	(27.61, 114.30)
10	共和国摇篮	(25.89, 116.02)
11	武功山	(27.50, 114.14)
12	大觉山	(27.70, 117.15)
13	古窑民俗博览区	(29.29, 117.19)

2.2 实验结果分析

本文中景点坐标是经纬度坐标,而地球是

一个三维的球体,不能够直接通过经纬度计算距离,所以需要将景点的经纬度坐标转换为距离进行计算。

把表 1 中景点坐标数据输入 Matlab 中,通过粒子群优化算法对江西 13 个 5A 景区进行路径规划。根据若干次 Matlab 仿真实验结果对蚁群算法的其他参数进行设定:城市数量为 13,种群大小为 50,学习因子 c_1 为 2,学习因子 c_2 为 2, $p_{\text{best}} - x_i$ 的保留概率 r_1 为 0.7, $g_{\text{best}} - x_i$ 的保留概率 r_2 为 0.8,算法最大迭代次数为 1000。利用同样的参数和程序对江西省旅游景点进行多次 Matlab 仿真计算,最优结果如图 3 所示。

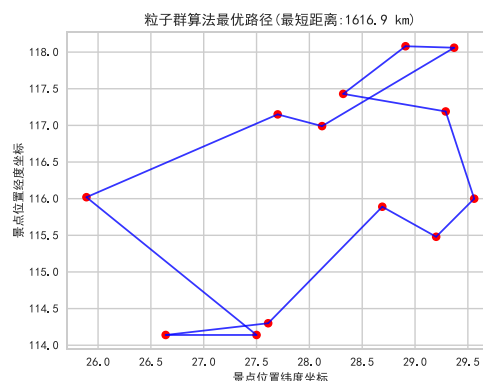


图 3 粒子群优化算法最优路径

根据仿真实验结果,得到最短路径顺序为: 11→10→12→4→8→5→6→13→2→3→1→9→7→11。

表 2 除了统计了景点间的直线距离(根据经纬度计算),还展示了景点间的真实距离(通过

表 2 最优路径各区间距离

景点路径	路径坐标	直线距离/km	导航真实距离/km
11→10	(27.50, 114.14) → (25.89, 116.02)	258.5	429
10→12	(25.89, 116.02) → (27.70, 117.15)	229.9	288
12→4	(27.70, 117.15) → (28.12, 116.99)	49.1	108
4→8	(28.12, 116.99) → (29.37, 118.06)	173.5	276
8→5	(29.37, 118.06) → (28.91, 118.08)	51.0	102
5→6	(28.91, 118.08) → (28.32, 117.43)	91.2	147
6→13	(28.32, 117.43) → (29.29, 117.19)	110.0	173
13→2	(29.29, 117.19) → (29.56, 116.00)	119.3	159
2→3	(29.56, 116.00) → (29.20, 115.48)	64.3	104
3→1	(29.20, 115.48) → (28.69, 115.89)	69.2	97
1→9	(28.69, 115.89) → (27.61, 114.30)	196.8	249
9→7	(27.61, 114.30) → (26.64, 114.14)	108.6	248
7→11	(26.64, 114.14) → (27.50, 114.14)	95.3	202

高德导航, 选择驾驶模式)。从表2可以看出导航的真实距离和通过经纬度计算得到的直线距离有较大差别, 而导航的真实距离才是日常生活中使用的, 所以本文把导航的真实距离作为实验结果。

由表2中的导航真实距离得到最短距离为2582 km, 相邻景点间的距离依次为429 km、288 km、108 km、276 km、102 km、147 km、173 km、159 km、104 km、97 km、249 km、248 km和202 km。

3 结语

TSP旅行商问题是一个生活中很常见的问题, 本文通过粒子群优化算法模拟对江西省13个5A景区进行路径规划, 通过对参数的多次调整, 最终找到一条最短的旅游路径, 为乘客节省旅途时间的成本, 完成了本文需要实现的目标。本文得到的最短距离是根据导航的真实距离, 更加贴近真实的生活, 但是没有考虑到旅途中路线等真实费用问题, 在日后的研究中应当把这些实际因素考虑进来, 这样才能更好地解决生活中的实际问题, 让科学研究走进生活。

参考文献:

- [1] DAMOS M A, ZHU J, LI W, et al. A novel urban tourism path planning approach based on a multiobjective genetic algorithm [J]. ISPRS International Journal of Geo-Information, 2021, 10(8): 530.
- [2] ZHOU X. Path planning of coastal tourism based on the improved firefly algorithm [J]. Journal of Coastal Research, 2020, 106(S1): 263-266.
- [3] 许伊萍. 浅谈旅行商问题与粒子群优化算法[J]. 电子制作, 2015(9): 268.
- [4] SKINDEROWICZ R. Improving Ant Colony Optimization efficiency for solving large TSP instances [J]. Applied Soft Computing, 2022, 120: 108653.
- [5] 李梦丹. 基于蚁群算法西安旅游路线的优化研究[J]. 价值工程, 2020, 39(20): 136-137.
- [6] 卢曦. 求解TSP问题的改进蚁群算法研究[J]. 无线互联科技, 2016(19): 121-122.
- [7] UĞUR A, AYDIN D. An interactive simulation and analysis software for solving TSP using Ant Colony Optimization algorithms [J]. Advances in Engineering Software, 2009, 40(5): 341-349.
- [8] SHI X H, LIANG Y C, LEE H P, et al. Particle swarm optimization-based algorithms for TSP and generalized TSP [J]. Information Processing Letters, 2007, 103(5): 169-176.
- [9] 蒋冬梅. 粒子群算法在TSP中的应用及其LabVIEW实现[J]. 信息化研究, 2018, 44(4): 24-29.
- [10] 王伟, 吴天红, 姜英姿, 等. 混合粒子群算法在TSP问题中的研究 [J]. 中国新通信, 2020, 22(9): 126-127.
- [11] GUO X, JI M, ZHAO Z, et al. Global path planning and multi-objective path control for unmanned surface vehicle based on modified particle swarm optimization (PSO) algorithm [J]. Ocean Engineering, 2020, 216: 107693.
- [12] 李金, 付春龙. 基于TSP问题的遗传算法和蚁群算法研究 [J]. 电脑编程技巧与维护, 2021(9): 17-18.
- [13] 陶丽华, 马振楠, 史朋涛, 等. 基于TSP问题的动态蚁群遗传算法 [J]. 机械设计与制造, 2019(12): 147-149.
- [14] 王港华. 基于遗传算法的小规模TSP问题研究分析 [J]. 物流工程与管理, 2022, 44(3): 111-114.
- [15] SU F, KONG L, WANG H, et al. Modeling and application for rolling scheduling problem based on TSP [J]. Applied Mathematics and Computation, 2021, 407: 126333.
- [16] 屈新怀, 单笛, 孟冠军. 基于靠近目标粒子群算法的AGV路径规划 [J]. 合肥工业大学学报(自然科学版), 2022, 45(1): 1-6.
- [17] 吴庆洪, 张颖, 马宗民. 粒子群优化算法及其应用综述 [J]. 微计算机信息, 2010, 26(30): 34-35.
- [18] 刘紫亮, 居翔, 张永芳, 等. 基于改进随机搜索算法的随机森林调参优化 [J]. 网络安全技术与应用, 2022(4): 49-51.
- [19] 孙艳丰, 戴春荣. 几种随机搜索算法的比较研究 [J]. 系统工程与电子技术, 1998(2): 44-48.
- [20] M. Y K, B. M R, N. N G. Mathematical model and random search algorithm for the optimal planning problem of replacing traditional public transport with electric [J]. Automation and Remote Control, 2020, 81(8).

- [21] 杨韬. 粒子群算法的参数分析[J]. 科技信息(学术研究), 2008(14):295-297.
- [22] 全芙蓉. 粒子群算法的理论分析与研究[J]. 硅谷, 2010(23):193.
- [23] 莫愿斌, 刘贺同, 陈德钊. 粒子群优化算法的发展趋势[J]. 计算机与应用化学, 2009, 26(4):430-434.
- [24] GUPTA A, SRIVASTAVA S. Comparative analysis of ant colony and particle swarm optimization algorithms for distance optimization[J]. Procedia Computer Science, 2020, 173:245-253.
- [25] 邹腊英. 基于TSP问题的旅游路线安排[J]. 兰州文理学院学报(自然科学版), 2015, 29(5):23-25.
- [26] 叶平. 基于粒子群算法的最优物流运输线路规划方法研究[J]. 微型电脑应用, 2021, 37(12):158-161.
- [27] BEN B B, BENJAMIN G, DICKSON K W, et al. A comprehensive review of particle swarm optimization[J]. International Journal of Engineering Research in Africa, 2016, 23:141-161.

Self-driving route planning of Jiangxi attractions based on particle swarm optimization algorithm

Song Yanyan^{1,2}, Li Guanghong¹, Zhu Ling^{3*}, Xia Chuanlin³, Zheng Wei³

(1. School of Tourism and Landscape Architecture, Guilin University of Technology, Guilin 541000, China;

2. School of Mathematics and Information Science, Nanchang Hangkong University, Nanchang 330063, China;

3. School of Software, Nanchang Hangkong University, Nanchang 330063, China)

Abstract: With the rapid growth of the service industry, every holiday season various tourist attractions have become vacation destinations. More and more people like to drive to tourist attractions, and if they want to drive to multiple attractions, how to plan the travel route so that the travel path takes the shortest time becomes a challenge. Particle swarm optimization algorithm has the advantages of fast convergence, few parameters, simple and easy to implement algorithm, and find the optimal solution through continuous iterative update. Path planning for 13 5A scenic spots in Jiangxi using particle swarm optimization algorithm. The longitude and latitude coordinates of the attractions are obtained through the Baidu pickup coordinate system and further converted into distances for calculation, so as to derive the shortest paths through 13 tourist attractions, and the real distance of navigation is taken as the experimental result. The effectiveness of the particle swarm optimization algorithm in solving such problems as path optimization is demonstrated by examples.

Keywords: particle swarm optimization algorithm; path planning; Jiangxi attractions

文章编号: 1007-1423(2023)11-0033-05

DOI: 10.3969/j.issn.1007-1423.2023.11.006

基于图注意力网络的突发事件触发词抽取

陈彦杰*, 廖 涛

(安徽理工大学计算机科学与工程学院, 淮南 232001)

摘要: 针对目前突发事件触发词抽取方法存在由于分词引起的误差传递而导致触发词提取不准确的问题, 提出基于图注意力网络的突发事件触发词抽取模型(ETEGAN)。ETEGAN 首先使用 Word2vec 和 BERT 预训练语言模型对文本序列进行向量化, 将获取到的向量表示与动态词向量相结合, 使用双向门控循环单元 BiGRU 提取上下文特征, 并利用图注意力网络 GAT 提取文本特征, 调整重要特征的权重, 突出重要词对事件触发词抽取的贡献。实验结果表明, 本文模型有效地提高了突发事件触发词识别准确率。

关键词: 触发词; 图注意力网络; 特征融合; 突发事件

0 引言

随着信息技术的发展以及传统产业的数字化转型, 互联网已经渗透到人类社会生活的各个方面, 将互联网作为获取信息的首要途径的人逐渐增多。互联网用户的增加, 导致数据呈现爆炸式的增长, 大量的数据以电子文本的形式在互联网中传播, 这些电子文本数据内容丰富、要素多样, 并且包含了一种蕴藏着极大价值的信息资源, 然而目前网页上的大量文本均呈非结构化状态并且包含大量冗余信息, 使得用户无法迅速从中找到有用的资源。因此面对互联网海量的数据, 如何针对性地从文本中抽取对用户有用的信息, 再通过结构化形式展现给用户, 成为一项重要的课题。

1 相关工作

事件抽取在信息抽取研究领域^[1]中至关重要, 它的主要任务是把含有事件信息的非结构化文本以结构化^[2]的形式呈现出来, 在自动文摘、自动问答、信息检索、知识图谱^[3]构建等

领域有着广泛的应用。事件触发词抽取是事件抽取的子任务, 其任务是要识别事件句中触发词的位置同时识别出其所属事件类型。触发词是事件句中最能体现事件发生的词语, 直接决定事件类型。事件触发词抽取的方法主要有三种: 基于模式匹配^[4]、基于机器学习^[5]和基于深度学习^[6]的方法。

在基于模式匹配的方法中, 目标语料使用适当的算法进行匹配, 并根据给定的模板提取事件要素。Liu 等^[7]手动创建模板进行提取, 并在初步处理后将文本信息嵌入模板。这种方法只遵循一个固定的模板, 所以给定的模板并没有涵盖真实场景中可能出现的所有情况。

基于机器学习的方法通过事件触发词的语义、句法等信息构建特征, 然后根据特征采用统计学模型进行分类以达到抽取的目的, 主要的机器学习模型包括最大熵模型、隐马尔可夫模型和支持向量机等。Chieu 等^[8]针对事件抽取问题引入了最大熵模型, 从人事管理中提取研讨会的公告和事件。

收稿日期: 2023-01-29 修稿日期: 2023-03-03

基金项目: 国家自然科学基金资助项目(62076006); 安徽省高等学校自然研究基金资助项目(KJ2016A202); 安徽省高校优秀青年人才支持计划项目(gxyq2017007); 安徽理工大学 2021 年研究生创新基金项目(2021CX2108)

作者简介: *通信作者: 陈彦杰(1996—), 女, 河南濮阳人, 硕士研究生, 研究方向为自然语言处理, E-mail: 15903751719@163.com; 廖涛(1977—), 男, 安徽淮南人, 博士, 副教授, 硕士生导师, 研究方向为自然语言处理

近年来,随着词嵌入技术的崛起,深度学习已成为一种新的研究趋势,并被应用于各个领域。基于深度学习的方法将事件抽取建模成为一个不依赖外部NLP工具的端到端模型,使用特征丰富的词向量作为输入,避免了复杂的手工作业。Nguyen^[9]解决了不平衡语料中的事件触发词抽取问题,通过卷积神经网络捕获句子中的重要特征信息。武国亮等^[10]提出了一种FB-Latiice-BiLSTM-CRF模型,以解决事件抽取任务中BiLSTM-CRF处理的低维问题,增强了模型学习词语和实体语义特征的能力,并取得了更好的识别性能。

本文提出一种基于图注意力网络的模型进行突发事件的触发词抽取,实验结果表明文中提出的模型在突发事件领域事件触发词抽取任务中是行之有效的。

2 ETEGAN 模型

2.1 模型整体架构

本文构建的突发事件触发词抽取模型(emergency triggers extraction based on graph attention network, ETEGAN)如图1,包括:词嵌入层、BiGRU层、GAT层、CRF层。

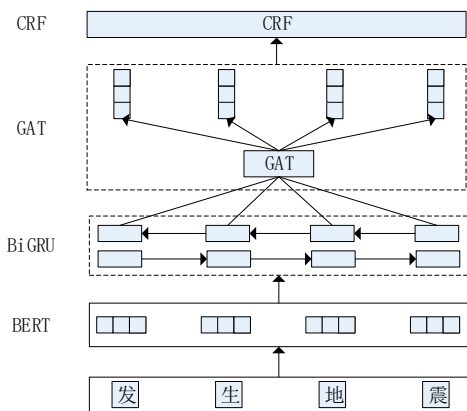


图1 模型整体架构

首先使用Word2vec进行文本向量化表示,再利用BERT模型提取动态词向量,将预训练词向量和动态词向量特征融合作为模型的输入;然后通过BiGRU对长距离语义特征进行抽取;接着采用GAT网络对特征进行权重调整,作为

最终的特征表示;最后在解码层通过CRF进行解码得到模型最终输出的结果。

2.2 词嵌入层

词嵌入层本质上是对输入字符的向量映射,即从离散的字符到分布式表示,可以很好地表示文本中的语义信息和语法关系。为了缓解由不正确的分词造成的问题,本文将语义特征进行融合作为模型的输入。

Word2vec模型先通过去标签等操作把语料库处理成文本序列,再通过分词的工具把输入数据处理成词序列数据 S ,最后通过Word2vec将文本向量化处理成预训练词向量 W 。

BERT预训练语言模型可以产生与上下文相关的词特征,并可以描述字的歧义性和句子的句法特征等。本文使用预训练模型BERT,根据同一个词的不同语境来计算每个单词的语境表征,对于由 n 个字符组成的文档作为输入 X ,经过BERT预训练模型得到相对应的特征向量 T 。

$$U = W \oplus T \quad (1)$$

把Word2vec和BERT分别训练的词向量 W 和动态词向量 T 相结合,得到联合词向量 U 。式中 $\oplus$ 代表将两个向量进行连接,把一个维度的向量结合起来成为新向量。

2.3 BiGRU层

在自然语言处理中,有大量的数据无法用传统的前向神经网络来建模,因此出现了循环神经网络RNN;它在理论上能够处理任意长度的序列信息,但在实践中,当序列过长时,就会出现梯度消失的问题,而且很难学到长期依赖的特征。

GRU是一种特殊的循环神经网络,混合了细胞状态和隐藏状态,其中遗忘门和输入门被合并在一个单一的更新门中。

GRU的结构更简单,参数更少,从而减少了训练时间。GRU由于其优越的序列建模能力,被广泛应用于语音识别、事件抽取和词性标注等方面。在本文中,BiGRU模型被用来对来自BERT层的向量进行初步建模;BiGRU计算两种不同的特征表示,即前向和后向两种特征,然后对这两种特征表示进行加权求和,产生涵盖

语境特征的向量。

2.4 GAT层

相比于卷积神经网络、循环神经网络等网络模型，因为图注意力网络GAT在图结构数据中表现出了优越的特征提取能力，所以研究者将它应用在了事件触发词抽取的任务中，图注意力网络的基本思想是通过计算邻节点的注意力来更新每个顶点隐藏层的输出，进而调节邻节点的权重。

如果 m 表示图卷积网络的层数，GAT层的输入与 BiGRU 的输出的隐藏层息息相关。GAT 层中第 i 层节点输入的隐藏状态 $\mathbf{h}_i^{(m)} = \{h_1, h_2, \dots, h_n\}$ ， n 为序列长度，同时是图的顶点数。

首先做一个线性变换，可训练的权重参数为 $\mathbf{W}^{(l)}$ 。

$$\mathbf{z}_i^{(m)} = \mathbf{W}^{(m)} \mathbf{h}_i^{(m)} \quad (2)$$

然后通过计算得到节点初始的注意力，通过计算 i 的邻居节点 j 得到初始的注意力权重 $e_{ij}^{(m)}$ 。将节点 i 和节点 j 的 $\mathbf{z}$ 向量相结合，再把 $\mathbf{z}$ 向量以及可学习的权重向量 $\vec{\mathbf{a}}^{(m)\top}$ 进行点积操作，其中 LeakyReLU 是非线性激活函数。

$$\mathbf{e}_{ij}^m = \text{LeakyReLU}(\vec{\mathbf{a}}^{(m)\top} (\mathbf{z}_i^m \parallel \mathbf{z}_j^m)) \quad (3)$$

接着将初始注意力进行归一化，计算出节点 i 对邻居节点 j 的注意系数，也就是注意力权重：

$$\alpha_{ij}^{(m)} = \frac{\exp(\mathbf{e}_{ij}^m)}{\sum_{k \in N(i)} \exp(\mathbf{e}_{ik}^m)} \quad (4)$$

因为每个节点的邻居节点都不一样，所以节点 i 对邻居节点 j 的注意系数与节点 j 对邻居节点 i 的不同，这样的结构和词与词之间的依赖关系有相似之处，这样使得关联性强的特征得到更多的关注。

然后对所有邻居节点的特征做基于注意力的加权求和，更新节点的特征向量：

$$\mathbf{h}_i^{(m+1)} = \sigma \left(\sum_{j \in N(i)} \alpha_{ij}^{(m)} \mathbf{z}_j^{(m)} \right) \quad (5)$$

接着使用多头注意力机制拼接输出结果：

$$\mathbf{h}_i' = \parallel_{K=1}^K \sigma \left(\sum_{j \in N(i)} \alpha_{ij}^{(k)} \mathbf{z}_j^{(k)} \right) \quad (6)$$

其中： $\parallel$ 代表拼接； α_{ij}^k 代表第 k 组注意力机制的

权重系数。

多头注意力进行多次运算，在多重子空间中学习节点之间的关联，战胜了循环神经网络并不能成功捕获长距离单词之间依赖特征的困难，还拥有并行的特征，不同子层之间互不影响，最后得到输出结果 $\mathbf{h}' = \{h'_1, h'_2, \dots, h'_n\}$ 。

2.5 CRF层

条件随机场(CRF)通常被用作自然语言处理任务中的解码器，根据连续标签之间的依赖关系对标签进行优化，缓解了输出标签存在的独立假设问题。

对于输入序列 $\mathbf{h}' = \{h'_1, h'_2, \dots, h'_n\}$ ，如果其对应的标注序列为 $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$ ，则标注序列 $\mathbf{y}$ 的概率为

$$P(\mathbf{y}|\mathbf{h}') = \frac{\exp \left(\sum_i (\mathbf{W}^{y'} \mathbf{h}'_i + \mathbf{T}_{(y'_{i-1}, y'_i)}) \right)}{\sum_{y'} \exp \left(\sum_i (\mathbf{W}^{y'} \mathbf{h}'_i + \mathbf{T}_{(y'_{i-1}, y'_i)}) \right)} \quad (7)$$

其中： y' 表示标签序列， $\mathbf{T}$ 为转移得分矩阵。通过计算输入序列 $\mathbf{h}'$ 对应的输出标签序列 $\mathbf{y}$ 得到预测的标签序列结果。

3 实验与分析

3.1 数据集及评测指标

实验选取中文突发事件语料库 CEC 作为语料集。上海大学语义智能实验室遵守《国家突发公共事件总体应急预案》的分类规则，统计了5类突发事件(地震、交通事故、火灾、食物中毒和恐怖袭击)的新闻报道，当作初始语料，然后进行文本分析、文本预处理、一致性检查以及事件标注等操作，最后在语料库中保存标注结果，语料库中对突发事件以及突发事件要素的标注非常全面。

本文将准确率 P 、召回率 R 、 F 值当作事件触发词抽取性能的评测指标。

3.2 实验结果及分析

3.2.1 与其他模型对比分析

本文训练和测试了下面的几种模型，以进一步比较模型的性能。表1显示了几种模型的触

发词抽取性能比较。

表1 与其他模型抽取效果的对比分析

模型	准确率 $P/\%$	召回率 $R/\%$	$F1/\%$
GAT+BiRNN	65.72	74.53	69.85
GAT+BiLSTM	58.73	78.26	67.10
BiGRU	69.52	64.28	66.80
GAT+GRU	63.57	75.19	68.89
ETEGAN(本文)	75.69	73.78	74.72

根据表1数据,在同一个数据集中训练和测试,与其他对比模型相比较,本文模型的 $F1$ 值和准确率表现较好。相较于GAT+BiRNN和GAT+BiLSTM,本文模型不仅避免了RNN存在的梯度消失问题,而且更容易训练,使模型的准确率分别提高9.97和16.96个百分点, $F1$ 值提高了4.87和7.62个百分点,表明了BiGRU的有效性;相较于BiGRU,本文模型添加了GAT,从数据可以看出,GAT通过对重要特征权重的调整使得模型可以着重关注相关性较高的语义特征,使模型的准确率提升了6.17个百分点, $F1$ 值提升了7.92个百分点,表明GAT提高了触发词抽取的效果;相较于GAT+GRU,本文模型多加了一层GRU用于获取双向的语义信息,得到了更为丰富的特征,使模型的准确率增长了12.12个百分点, $F1$ 值增长了5.83个百分点。

3.2.2 与其他工作结果对比分析

为了更全面地评估本文提出的模型的性能,这里将其结果与以往其他工作者的研究结果进行比较,对比情况如表2所示。

表2 与其他研究结果对比分析

模型	准确率 $P/\%$	召回率 $R/\%$	$F1/\%$
DMCNN ^[11]	61.60	58.80	60.20
HBTNGMA ^[12]	49.86	57.69	53.49
RBBLC ^[13]	73.51	68.87	71.11
HNN ^[14]	77.10	53.10	63.00
ETEGAN(本文)	75.69	73.78	74.72

对表2中的数据进行对比分析,相比于其他的神经网络方法,本文提出的基于图注意力网络的突发事件触发词抽取模型表现出了最好的

性能,表现了BiGRU和图注意力网络的优秀的性能,可以更深层次地学习到丰富的语义特征,同时使得高度相关特征得到了更多的关注;同时采用Word2vec和BERT预训练语言模型,得到了丰富的语义信息,有助于本文的突发事件触发词特征识别更准确。

4 结语

事件抽取在突发事件抽取以及知识图谱等研究中具有不可或缺的作用。针对触发词提取不准确的问题,本文提出了基于图注意力网络的突发事件触发词抽取模型ETEGAN。模型在实验数据集上取得了较好的效果,有助于突发事件触发词抽取任务。但是仍有进步的空间,在后续的工作中,将争取触发词抽取准确率进一步的提高。

参考文献:

- [1] YU W, YI M, HUANG X, et al. Make it directly: event extraction based on tree-LSTM and Bi-GRU [J]. IEEE Access, 2020, 8: 14344-14354.
- [2] 马春明, 李秀红, 李哲, 等. 事件抽取综述[J]. 计算机应用, 2022, 42(10): 2975-2989.
- [3] 冀相冰, 朱艳辉, 詹飞, 等. 基于门控多层次注意力机制的事件主体抽取[J]. 计算机应用与软件, 2021, 38(9): 173-179, 187.
- [4] 贾阵, 丁泽华, 陈艳平, 等. 面向司法数据的事件抽取方法研究[J]. 计算机工程与应用, 2023, 59(6): 277-282.
- [5] WU Y, ZHANG J. Chinese event extraction based on attention and semantic features: a bidirectional circular neural network [J]. Future Internet, 2018, 10: 95-95.
- [6] 高李政, 周刚, 罗军勇, 等. 元事件抽取研究综述[J]. 计算机科学, 2019, 46(8): 9-15.
- [7] LIU S, CHEN Y, HE S, et al. Leveraging framenet to improve automatic event detection [C] // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 2016: 2134-2143.
- [8] CHIEU H L, NG H T. A maximum entropy approach to information extraction from semi-structured and

- free text[C]//Proceedings of the Eighteenth National Conference on Artificial Intelligence, 2002:786-791.
- [9] NGUYEN T H, GRISHMAN R. Relation extraction: perspective from convolutional neural networks [C]//Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing, Denver, Colorado, 2015:39-48.
- [10] 武国亮, 徐继宁. 基于命名实体识别任务反馈增强的中文突发事件抽取方法[J]. 计算机应用, 2021, 41(7):1891-1896.
- [11] CHEN Y, XU L, LIU K, et al. Event extraction via dynamic multi-pooling convolutional neural networks [C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 2015:167-176.
- [12] CHEN Y, YANG H, LIU K, et al. Collective event detection via a hierarchical and bias tagging networks with gated multi-level attention mechanisms [C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 2018:1267-1276.
- [13] 杨登辉, 刘靖. 基于RBBLC模型的中文事件抽取方法[J]. 南京师范大学学报(工程技术版), 2022, 22(3):38-44, 82.
- [14] FENG X, QIN B, LIU T. A language-independent neural network for event detection[J]. Science China Information Sciences, 2018, 61(9):1-12.

Emergent triggers extraction based on graph attention network

Chen Yanjie*, Liao Tao

(School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232001, China)

Abstract: To address the problem of inaccurate extraction of trigger words due to error transmission caused by splitting of words in current emergency trigger word extraction methods, we propose emergency triggers extraction based on graph attention network (ETEGAN). ETEGAN first uses Word2vec and BERT pre-trained language models to vectorize text sequences, combines the obtained vector representations with dynamic word vectors, extracts contextual features using BiGRU, a two-way gated cyclic unit, and extracts text features using GAT, adjusts the weights of important features, and highlights the contribution of important words to event-triggered word extraction. The experimental results show that the model in this paper effectively improves the accuracy of event-triggered word recognition.

Keywords: triggers; graph attention network; characteristics of the fusion; emergency

文章编号: 1007-1423(2023)11-0038-05

DOI: 10.3969/j.issn.1007-1423.2023.11.007

软件可测性和测试用例效率的研究综述

王 东*

(广东白云学院大数据与计算机学院, 广州 510450)

摘要: 软件度量是软件开发的健壮性和可维护性的主要方法。软件度量可有效降低软件设计过程中的复杂性。为保证软件系统的可测试性, 在软件开发过程中采用可测试性指标, 确保软件的可维护性, 也是软件质量保证的一个有效手段。该研究内容包括的问题涉及静态代码、测试度量、可测试性和测试用例有效性四者之间的关系。分析结果表明, 大小和复杂性指标适用于预测面向对象类的可测试性。

关键词: 软件工程; 软件质量; 可测试性; 测试用例

0 引言

在软件工程领域中, 软件质量实际上是一个多维的概念。其中, 软件可测试性是软件质量保证体系中的重要内容, 它可直接反映出软件测试工作的复杂性和工作量。在软件设计和开发过程中, 增强可测试性亦可增加软件测试过程中发现缺陷的概率。ISO/IEC 9126-1 标准定义的可测试性是指: 对修改后的软件进行测试的能力。ISO/IEC 25010:2011 系统和软件工程标准定义的系统 and 软件质量需求和评估标准 (SQuaRE) 中的软件系统质量模型将可测试性作为可维护性的一部分。

可测试性从另一方面也可反映出是对软件进行测试工作的便利程度, 其本身就是一种质量属性。测试是软件开发过程中评估软件质量的活动, 需要专业的方法和工具。测试工作包括方方面面, 主要有: 功能、性能、安全性、易用性等测试。目前, 自动化测试虽能提高效率、降低成本, 但在相当长的一段时间内尚不能实现对人工手动测试的替代。

当一个软件具有较高可测试性时, 在相同的测试强度下, 软件测试人员更容易发现软件缺陷, 也可以说发现更多软件缺陷; 而当软件的

可测试性较低时, 软件系统中的缺陷将不易被发现, 测试工作量也将随之上升。由此可见, 软件的可测试性将影响软件测试成本, 进而影响软件质量。

影响可测试性的五个主要因素有: 可理解性、可观察性、可控性、可追溯性和对测试的支持能力。本文所述的对软件可测试性和测试用例有效性的研究问题为: 静态代码和测试度量对测试用例有效性和可测试性的影响。

1 软件可测性

可测试性是软件质量模型的重要内容之一, 对可测试性的度量工作因影响因素较多而具有较高难度。低可测试性的软件系统具有较低的可靠度, 这与测试工作的严密性无关, 而且低可测试性的软件系统往往在软件开发后期的维护和更新成本更高, 更难控制。软件系统的可测试性关系到软件的可靠性、测试效率、测试成本和工作量。对于软件开发而言, 测试的成本占整个软件开发成本的 50% 甚至更高, 因此改进可测试性对提高测试效率和降低整体开发成本有着重大影响。理想状态下, 测试人员在设计和执行测试用例所花费的时间应该占到全部

收稿日期: 2023-02-22 修稿日期: 2023-03-03

作者简介: *通信作者: 王东 (1979—), 男, 北京人, 硕士, 高级工程师, 研究方向为软件工程、人工智能、大数据和云计算技术, E-mail: dongwang@baiyunu.edu.cn

开发时间的25%~50%之间。当前热门道德软件开发方法,如:敏捷方法、极限编程和Scrum迭代式增量软件开发方法等,均体现出了测试工作的重要性。以测试驱动开发(TDD)为例,为确保软件系统存诸单元的可靠性,TDD在编写代码前已由开发人员或测试人员完成设计和开发测试的工作^[1]。单元测试是提高软件质量的主要方式,可开发过程的早期阶段发现和低成本修复bug。在实际生产环境中,测试代码量与软件开发代码量的比例约为1:1。

总之,提高软件可测试性对提升软件系统的质量有着积极作用。但是,对软件可测试性的度量和评估工作具有一定的复杂性。

1.1 可测试性的定义

IEEE定义的软件可测试性(software testability)是指一个软件工件(软件系统、模組、需求文件或设计文件等)在一个给定的测试环境下,可支援测试的程度。ISO标准将可测试性描述为“与验证软件产品所需工作有关的软件属性”。因此,ISO从测试软件产品所需工作的角度出发定义了可测试性,而IEEE从测试标准的角度定义了可测试性^[2]。

McCall质量模型是第一个定义软件质量的质量模型。McCall模型的主要内容包括:便捷性、仪器和模块的可测试性,并将其定义为验证需求的能力。Boehm改进了McCall模型,将可测试性定义为验证需求是否满足实现的软件的便捷性。ISO9126质量模型定义可测试性为验证软件产品所需的必要工作。

1.2 软件度量和可测量性

在软件工程领域,软件测量的主要目标是帮助软件工程师和项目管理者做出针对软件开发过程的准确的预测和决策,通过监控项目进展并适时介入质量控制机制判断软件项目及其各组件的功能和性能是否符合需求。软件度量是指量化和度量软件特性的指标参数,软件指标是软件质量的重要组成部分,在分析和提高软件质量方面发挥了重要作用。此外,软件指标可用于验证软件质量的要求是否被满足^[3]。

软件测试中的软件度量包括:产品、过程和项目度量。而过程和产品度量与软件质量关系密切。过程度量标准反映了开发过程质量,

而最终产品度量标准反映了软件产品的特征。根据产品度量系统的不同方面,产品度量可以进一步分为静态度量和动态度量。静态度量在软件非运行状态下进行度量,是对软件的内部质量的描述,大多数现有的质量度量以静态形式出现,如代码行度量和开发复杂度度量。动态度量只能在运行时对软件的属性进行度量,是对软件运行状态的描述。

2 研究方法

本研究采用的可测试性数据集源自1186个大大小小的、不同类别的Java开源项目^[3-6]。数据包含多个不同类的度量(继承、大小、复杂性、内聚性、封装性和耦合性等)、6个测试工作量度量和3个测试质量度量。在文献[3-6]中提到了收集数据信息和分析计算指标的工具,本文采用了其中介绍的方法。表1为本研究所涉及的软件度量类别。

本研究所选实验数据集源自1186个Java开源项目的9861个实例(类)^[4]。考虑到本实验的研究重点是测试有效性问题,本研究将突变分值一项作为研究重点。突变分值通过组合测试百分比计算被消除的突变体。除了通过突变分值来评估测试执行中发现缺陷的能力外,这些分值还表示了不同测试活动发现缺陷的能力的差异。突变分值范围为0~1。

本研究根据突变分值将数据结合的测试分为有效和无效两类,利用突变分值四分位数将第1个四分位数内的测试分配到非四分位数集,而第4个四分位数内的测试被分配到有效集。任何落在这两个四分位数之间的测试用例都将被丢弃。基于响应离散化噪声的背离分配给特定类的数据点。由于本研究重点在于探索有效测试特征和无效测试特征,因此,为了突出效果的差异,实验去掉了具有平均有效值的测试类。

在计算该数据集的四分位数后,研究发现突变分数为1的测试用例被认为是第4个四分位数的一部分,0.4及以下的测试用例被认为是第1个四分位数的一部分。因为具有平均行,实验删除了其余四分位数中的所有实例。数据集集中的其余实例为5022。本研究旨在评估Java面向对象类的测试有效性(可测试性)。

表 1 软件度量分类列表

度量类别	度量属性	度量指标
代码度量	规模	代码行(LOC)
		字节码指令数(NBI)
		注释行数(LOCCOM)
		公共方法数(NPM)
		静态方法数(NSTAM)
		字段数(NOF)
		静态字段数(NSTAF)
		方法调用数(NMC)
		内部方法调用数(NMCI)
		外部方法调用数(NMCE)
	复杂性	每类加权方法(WMC)
		平均方法复杂度(AMC)
		类的响应(RFC)
	继承	继承树深度(DIT)
		子类数(NOC)
		功能抽象的测量(MFA)
	耦合	对象类之间的耦合(CBO)
		继承耦合(IC)
		方法间耦合(CBM)
		传入耦合(Ca)
		传出耦合(Ce)
	内聚	方法中内聚性不足(LCOM)
		方法的内聚性不足(LCOM3)
		类中方法间内聚(CAM)
	封装	数据访问度量(DAM)
		私有字段数量(NPRIF)
		私有方法数量(NPRIM)
		受保护方法数(NPROM)
测试工作量 度量		测试代码行(T-LOC)
		测试数(T-NOT)
		测试断言数(T-NOA)
		测试方法调用数量(T-NMC)
		每个类的测试加权方法(T-WMC)
测试质量 度量		测试方法平均复杂度(T-AMC)
		代码行测试覆盖率(L)
		分支覆盖(B)
		突变分值(M)

2.1 因变量和自变量

以突变得分作为因变量,即被消除的突变体除以产生的突变体数量的百分比,因变量也是开发人员使用的可测试性关键指标之一。自变量除了两个主要维度(测试工作量和代码度量)外,还考虑了34个因素(见表1)。这些度量为测试人员提供了验证测试是否具有有效性以及是

否与这些元素相关联^[5]。本研究目标之一是确定仅依赖于静态代码质量特征的轻量级估计模型,这些特征可以很容易地计算到当前版本的测试类。由28个特性组成的代码度量集,用于评估继承、复杂性、耦合、大小、封装和内聚的效果。对测试类工作量的计算参考了测试用例的复杂性和数量的相关的测试工作量的度量^[6]。

2.2 研究问题

本研究的主要目标之一是对影响测试用例有效性(突变分值)的因素进行分析,例如检测缺陷的能力,以便在开发测试用例时将有效性指标更好地付诸实施。为此,本研究设计了三个研究问题。第一个研究问题针对所选的34个自变量和测试用例有效性之间是否存在关联的初步评估,其目的是揭示具有高或低突变分值的测试用例的自变量分布是否变化以及变化到何种程度;第二个研究问题关于在机器学习技术的基础上制定自动化策略时,如何根据突变的分数来检查测试的有效性;最后一个研究问题是找到最主要的可测试性指标是什么。本研究将三个研究问题概括如下:

研究问题1:静态度量和测试用例有效性之间有关系吗?

研究问题2:静态代码和测试度量在多大程度上有助于预测测试用例的有效性?

研究问题3:最具影响力的可测试性指标是什么?

2.3 方法

针对研究问题1,本研究引入了Spearman的秩相关系数用于评估静态度量和测试用例有效性两个指标的统计依赖性,Spearman方法准确度高而且不需要正态数据分布^[7]。针对研究问题2,有不同的机器学习算法可供选择,由于分类器使用的预测目的可能会对模型的性能产生较大影响,因此在选出最优的估算模型的分类器之前,本研究将对多个分类器分别进行测试。因此,本研究将比较不同机器学习算法的效果来确定可测试性预测模型。

决策树利用高度信息性的特性来划分含有不同类的测试实例。本研究采用了C4.5决策树版本,它使用了基础信息熵的概念为每个子树选择根节点,决策树是一种易于理解和解释的

分类器^[8]。

随机森林其实是一种协作学习方法，其可在引导数据上创建决策树的多元变量。随机森林中的每棵树都是基于真实数据的子集上构建的，并在一个高鲁棒性的分类器中展示特征结果，一个预测类的选出将取决于选票数量。随机森林的開箱即用的便捷性能是一大特色^[9]。

多层感知器是一种嵌入在深层架构中的前馈神经网络，该架构包括在有向无环图中相互连接的不同计算单元层，他们具有将输入映射到输出的特性，并且在回归测试项目中采用较多^[10]。

为区分不同的类标签，本研究中的静态指标排名根据它们的贡献来判定，排名算法采用预分配的一个分数来评估单个特征，并根据该分数对它们进行排名。针对研究问题3，本研究采用了四种不同的算法来分析对可测性影响最大的指标，即增益比、信息增益、对称不确定性和OneR算法^[11]。

3 实验结果

本研究采用了Spearman等级相关系数并对实验结果进行分析，出于验证测试有效性的目的，本实验结果只包含系数达到0.5以上的正相关或负相关的结果。表2显示的相关值反映出大小和耦合与突变分值呈高度负相关。针对研究问题1的答案可以得出：有效测试用例与被测类的复杂性和大小度量之间存在明显的负相关关系，也就是说，减少类的大小并降低类的复杂性将增加类的可测试性。

表 2 带有突变分数的静态度量的 Spearman 相关系数

静态度量	相关性系数
节码指令数	-0.61652431
类的响应	-0.603928734
代码行	-0.58334397
每类加权方法	-0.56027992
方法调用数	-0.55492337
外部方法调用数	-0.512557727
注释行数	-0.50032704
类中方法的内聚	-0.521969871

针对研究问题2，探究静态代码和测试度量在多大程度上能够影响对测试用例有效性的预

测精准度。表3反映了所选分类器在五个评价标准下的性能。如上所述，使用测试指标和静态代码的模型不仅依赖于F测量(87%)，并且在AUC分析期间也表现出了较高性能。三个选定的分类器中，随机森林机制在所有五个评价标准上得分最高。实验结果表明，机器学习方法可有效评估测试用例的有效性。本实验选择了多个领域的测试项目，实验结果证明本评估模型对于测试用例有效性的评估是有效的。

表 3 预测模型的分类结果

分类器	准确度	精确度	调用	F测量	AUC
随机森林	0.877	0.879	0.877	0.877	0.942
多层感知器	0.830	0.830	0.830	0.830	0.895
决定树	0.825	0.827	0.826	0.826	0.870

针对研究问题3，未找到对可测试性影响最大的因素，本实验比较了不同的排名算法。表4显示了不同排名算法的重要度量，这些度量是针对度量类可测试性的易用指标。表4反映出表内所有算法都能找到排名靠前的指标。这些重要的指标包括：NBI、RFC、LOC、LOCCOM、WMC和NMC。为深入研究这些指标，本研究通过对这些指标进行分类来设计属性。字节码指令数(NBI)、代码行数(LOC)、注释行数(LOCCOM)和方法调用数(NMC)是代码量大小的度量。而对一个类的响应(RFC)和每个类的加权方法(WMC)被认为是复杂度度量。这表明，在软件开发过程中，为确保系统可测试性，需要观察和监视这些度量值。

表 4 使用不同排名算法的重要度量

	增益率	信息增益	对称不确定性	OneR算法
排名 1	NBI	NBI	NBI	RFC
排名 2	RFC	RFC	RFC	NBI
排名 3	LOC	LOC	LOC	LOC
排名 4	LOCCOM	NMC	LOCCOM	NMC
排名 5	WMC	WMC	NMC	LOCCOM
排名 6	NMC	LOCCOM	WMC	WMC
排名 7	NMCE	CAM	NMCE	CAM
排名 8	CAM	NMCE	CAM	NMCE
排名 9	T-NMC	NPM	NPM	NPM
排名 10	NPM	LCOM	LCOM	LCOM

4 论证

本研究通过实验证明了源代码质量在解决测试用例有效性方面的作用,软件测试人员参考本研究成果可有效避免无效测试的发生。软件开发人员在软件设计和开发过程中参看本研究提出的模型可从设计和编码阶段的源头避免无效测试情况的出现。

本研究推荐的评估模型可检查测试过程中的质量。对于开发人员来说,通过突变测试方法来评估测试有效性问题能够做到更精准和更高效,这种方法可帮助测试人员分析找出影响精准高效测试的负面原因。通过对测试代码有效性的初步评估可帮助测试人员理解影响测试有效性的主要因素。

5 结语

本研究分析了测试指标和静态代码方面的28个因素与测试用例有效性和可测试性之间的关系。通过建立预测模型,利用已知因素区分无效测试和有效测试。此外,还阐明了相关的基本代码度量对面向对象类的可测试性的影响。本研究还将在未来通过进一步的实验,探索更多其它的度量因素对测试用例有效性的影响,例如:对面向过程类应用的测试,并将研究成果应用于测试过程中,如:测试用例的设计与选择、优先级划分和自动化测试。

参考文献:

[1] 王孝梅. 测试驱动开发的研究[J]. 电脑编程技巧与

维护,2021(2):47-49.

- [2] 王丽娟,刘全周,李占旗,等. 汽车电子嵌入式软件单元测试用例设计方法研究[J]. 中国汽车,2022(8):14-18,23.
- [3] 唐佳丽. 面向对象软件测试性度量及应用研究[D]. 石家庄:河北师范大学,2021.
- [4] TERRAGNI V, SALZA P, PEZZÈ M. Measuring software testability modulo test quality[C]//Proceedings of the 28th International Conference on Program Comprehension,2020:241-251.
- [5] 胡仁青. 基于深度学习算法的机器自动翻译质量评估模型[J]. 电子设计工程,2021,29(21):33-37,42.
- [6] COLES H, LAURENT T, HENARD C, et al. PIT: a practical mutation testing tool for java[C]//Proceedings of the 25th International Symposium on Software Testing and Analysis, Saarbrücken, Germany,2016:449-452.
- [7] 苏丽敏,何慧爽. 基于区间数的Spearman秩相关系数的多属性决策方法[J]. 统计与决策,2019,35(6):51-53.
- [8] 张朋涛. 基于决策树生成软件测试方案的研究与实现[D]. 西安:西安电子科技大学,2017.
- [9] 胡金昊. 基于随机森林和SVM的软件缺陷预测方法研究[D]. 杭州:浙江工业大学,2020.
- [10] 尤志宁,浦云明. 随机样本遗传MLP模型算法[J]. 计算机工程与应用,2015,51(21):121-127,143.
- [11] MALHOTRA R, SHARMA A. Threshold benchmarking for feature ranking techniques[J]. Bulletin of Electrical Engineering and Informatics, 2021, 10(2):1063-1070.

The review of software testability and test cases effectiveness

Wang Dong*

(Big Data and Computing Institute, Guangdong Baiyun University, Guangzhou 510450, China)

Abstract: To develop robust and maintainable software systems, the term software measurement a tool to be employed. By utilizing software metrics, the complexity of software design and development can be reduced significantly. Ensuring testability for a software system, maintainability attribute is one of important concerns which is also a vital part of quality assurance. This paper studies interactive relationship among static code, test metrics, testability and test cases effectiveness by finding out answers for three research questions. As a result, this paper shows that the index of size and complexity metrics are suitable to be utilized to predict the testability of a software system.

Keywords: software engineering; software quality; testability; test case

文章编号: 1007-1423(2023)11-0043-05

DOI: 10.3969/j.issn.1007-1423.2023.11.008

基于 WOA-RBF 的温室温度预测研究

李其操, 董自健*

(江苏海洋大学电子工程学院, 连云港 222005)

摘要: 温度是温室内作物生长的关键环境因素, 准确预测出温室内温度的变化对温室管理有着重要意义。为了提高预测精度, 提出了基于鲸鱼优化算法(WOA)优化 RBF 神经网络的温室温度预测模型, 通过鲸鱼优化算法对 RBF 神经网络的输出层的神经元线性权重、隐藏层的神经元中心点和隐藏层的高斯核宽度进行寻优, 得到最优参数, 最终得到优化后的输出结果。实验证明, 与未优化的 RBF 神经网络预测模型相比, WOA-RBF 神经网络温室温度预测模型得到的结果误差更小, 平均绝对误差(MAE), 均方误差(MSE)和平均绝对百分比误差(MAPE)分别为 0.478、0.346 和 0.588, 与未优化的 RBF 神经网络预测模型相比分别提升了 9.1%、32.7% 和 18.0%。

关键词: 鲸鱼优化算法; RBF 神经网络; 温室温度; 预测模型

0 引言

随着农业技术的发展, 国内温室面积已经排在了世界前列。温室是一个复杂的动态环境, 各种环境因子之间具有很强的耦合作用, 一种环境因子往往随着另一种环境因子的变化产生较大波动^[1]。温度是温室内作物生长的关键影响因子, 能够精准地预测出温度对温室的调控有着重要作用。

近年来, 许多学者提出了针对温度预测的方法。左志宇等^[2]提出采用时序分析法建立温度预测模型的方法; 徐意等^[3]构建了基于 RBF 神经网络的温室温度预测模型; 徐宇等^[4]构建了基于复数神经网络的温室温度预测模型; 为了提高精度, 夏爽等^[5]使用 PSO-RBF 神经网络设计了针对温室温度的预测模型; 王媛媛^[6]使用改进 PSO-RBF 神经网络算法, 提升了网络模型的预测精度, 但该模型并未对隐藏层节点个数进行优化, 导致模型的学习时间延长。

为了提高模型的预测精度, 加快模型的学习速度, 本文利用鲸鱼优化算法(WOA)对 RBF

神经网络输出层的神经元线性权重、隐藏层的神经元中心点和隐藏层的高斯核宽度进行优化, 提高模型的预测精度, 对温室的智能调控有着一定的参考价值。

1 模型构建

1.1 鲸鱼优化算法

鲸鱼优化算法(whale optimization algorithm, WOA)是模仿自然界中座头鲸狩猎捕食的行为而提出的群体优化算法。在鲸鱼群的狩猎过程中, 当某条鲸鱼发现猎物时, 鲸鱼群中的其他鲸鱼便会靠近该鲸鱼的位置一起狩猎该猎物。将这种狩猎行为应用到求解问题时, 可以将整个鲸鱼群中的每个鲸鱼个体位置都看作每个解, WOA 就是将搜索问题解的过程看作每个鲸鱼个体不断更新位置, 直到得到最优解的算法。

座头鲸的狩猎行为有三种类型, 分别是包围猎物、气泡网攻击和搜索猎物。

1.1.1 包围猎物

座头鲸在狩猎时要包围猎物, 这种行为可

收稿日期: 2023-02-21 修稿日期: 2023-03-08

作者简介: 李其操(1997—), 男, 浙江绍兴人, 硕士研究生, 研究方向为农业物联网; *通信作者: 董自健(1973—), 男, 江苏连云港人, 博士, 教授, 研究方向为检测与控制、通信技术, E-mail: dzjian@126.com

以由下面的数学模型表示:

$$D_k = |C_1 \cdot X_k^* - X_k^j| \quad (1)$$

$$X_k^{j+1} = X_k^* - A_1 \cdot D_k \quad (2)$$

式中: k 是当前迭代次数; A_1 和 C_1 为表示系数, X_k^* 表示目前为止最佳的鲸鱼位置; X_k^j 表示当前鲸鱼的位置。 A_1 和 C_1 可以由下列公式得出:

$$A_1 = 2a \cdot r_1 - a \quad (3)$$

$$C_1 = 2 \cdot r_2 \quad (4)$$

式中: a 是随着迭代次数的增加, 从 2 到 0 线性递减, r_1 和 r_2 都是 $[0, 1]$ 之间的随机数。

1.1.2 气泡网攻击

气泡网攻击是指座头鲸通过转圈吐泡泡将猎物包围起来, 并且不断缩小包围圈, 由于猎物也在移动, 所以座头鲸是呈螺旋状前进的。这种行为可以由下面的数学模型表示:

$$X_k^{j+1} = \begin{cases} X_k^* - A_1 \cdot D_k & p < 0.5 \\ X_k^* + D_k \cdot e^{bl} \cdot \cos(2\pi l) & p \geq 0.5 \end{cases} \quad (5)$$

式中: b 为对数螺旋形状常数; l 为 $[-1, 1]$ 之间的随机数; p 为 $[0, 1]$ 之间的随机数。

1.1.3 搜索猎物

搜索猎物是指座头鲸随机搜索猎物的行为, 在这种行为中, 座头鲸个体会向着某个随机的座头鲸个体移动。包围猎物的数学模型中的 A_1 可以表示出座头鲸是否朝着最佳鲸鱼个体移动, 若 A_1 在 $[-1, 1]$ 范围内, 则表示座头鲸在向最佳鲸鱼个体移动; 若 A_1 不在 $[-1, 1]$ 范围内, 则表示座头鲸在向其他鲸鱼个体移动。这种行为虽然会使鲸鱼偏离目标猎物方向, 但其优点是会增强群体的全局搜索能力。这种行为可以由下面的数学模型表示:

$$D_k = |C_1 \cdot X_k^{\text{rand}} - X_k^j| \quad (6)$$

$$X_k^{j+1} = X_k^{\text{rand}} - A_1 \cdot D_k \quad (7)$$

$$A_1 = 2a \cdot r_1 - a \quad (8)$$

$$C_1 = 2 \cdot r_2 \quad (9)$$

式中: X_k^{rand} 表示鲸鱼群体中的随机座头鲸个体。

1.2 RBF 神经网络

RBF 神经网络是一种由输入层、隐藏层和输出层组成的三层前馈神经网络。其基本结构图如图 1 所示。

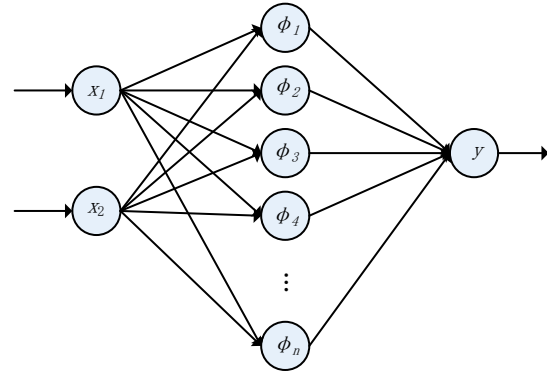


图 1 RBF 神经网络结构

RBF 神经网络的隐藏层使用的激活函数为径向基函数, 最常见的径向基函数为高斯核函数, 其定义如下所示:

$$\varphi(x_i, c_i) = e^{-\frac{\|x_i - c_i\|^2}{2\sigma^2}} \quad (10)$$

式中: c_i 为第 i 个神经元的中心点, σ 为高斯核的宽度, $\|x_i - c_i\|$ 为样本 x_i 到中心点 c_i 的欧式距离。

根据激活函数, 可以得出 RBF 神经网络的输出函数如下所示:

$$y(x) = \sum_{i=1}^q w_i \cdot \varphi(x, c_i) \quad (11)$$

式中: q 为隐藏层神经元的个数; w_i 为第 i 个隐藏层神经元与输出层之间的权重; c_i 为第 i 个隐藏层神经元的中心点。

1.3 WOA-RBF 神经网络预测模型

RBF 神经网络预测模型的训练速度、精度与输出层的神经元线性权重、隐藏层的神经元中心点和隐藏层的高斯核宽度有关, 引入鲸鱼优化算法优化这三个参数可以加快预测模型的训练速度并且提高预测精度。WOA-RBF 神经网络预测模型的流程如图 2 所示。

算法流程:

(1) 初始化参数: 即鲸鱼种群规模大小, 最大迭代次数;

(2) 将 RBF 神经网络输出层的神经元线性权重、隐藏层的神经元中心点和隐藏层的高斯核宽度转换为鲸鱼种群的位置坐标;

(3) 计算每一头鲸鱼相应的适应度值, 根据适应度值的大小排序, 并选取设定适应度范围内的鲸鱼个体作为初始种群, 适应度函数的计算公式如下:

$$F_i = k \left(\sum_{i=1}^n \text{abs}(y_i - o_i) \right) \quad (12)$$

式中: k 为系数; n 为神经网络输出节点数量; y_i 为神经网络第 i 个节点的期望输出; o_i 为神经网络第 i 个节点的预测输出。

(4) 更新下一代的位置;

(5) 若达到终止条件, 则输出最优个体, 即算法找到最优解; 否则, 返回步骤(4)。

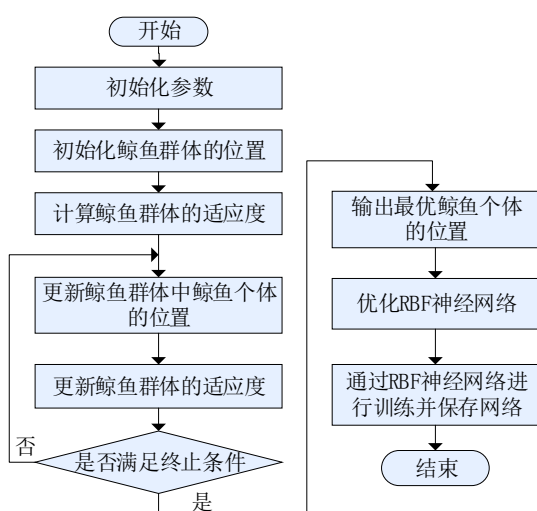


图2 WOA-RBF神经网络预测模型的流程

2 实验与结果分析

2.1 样本数据采集

本文实验数据采集自连云港葡萄园温室,

选用温度、湿度、二氧化碳浓度、土壤氮含量、土壤磷含量和土壤钾含量作为样本数据, 每15分钟采集一次数据, 共采集了2292组样本数据。为了实验测试更方便, 本文选用其中2000组数据, 以前80%的样本数据作为训练样本, 剩余的20%样本数据作为测试样本。表1为部分样本数据。

2.2 模型参数设定

根据所获得的样本数据, 将RBF神经网络输入层节点设定为5个, 即5个特征, 分别为湿度、二氧化碳浓度、土壤氮含量、土壤磷含量和土壤钾含量数据; 输出层节点为1个, 特征为温度数据; 通过试凑法确定隐藏层节点为10个。因此, RBF网络的结构为5-10-1。

2.3 模型评价标准

为了评定预测模型的性能, 本文以平均绝对误差(MAE)、均方误差(MSE)和平均绝对百分比误差(MAPE)作为评判预测模型性能优劣的标准。各评估误差指标的计算公式如下所示:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (13)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (14)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (15)$$

式中: $\hat{y}_i$ 为模型的预测值; y_i 为真实值; n 为样本数。所得的值越小, 则模型的性能越优异。

2.4 预测结果及分析

通过 Matlab 软件对 WOA-RBF 网络预测模型和未优化的 RBF 网络预测模型进行验证, 得到的预测结果对比如图3所示。

表1 部分样本数据

日期	时间	温度/℃	湿度/%	二氧化碳浓度/ppm	土壤氮含量/mg	土壤磷含量/mg	土壤钾含量/mg
2022/8/28	22:02:29	21.7	20.6	326	152	62	112
2022/8/28	22:17:43	21.7	20.7	326	152	62	112
2022/8/28	22:32:58	21.6	21.6	335	152	62	112
2022/8/28	22:48:12	21.5	22.6	338	153	62	112
2022/8/28	23:03:27	21.5	24.8	345	153	62	113
2022/8/28	23:18:42	21.4	25.2	346	153	62	113

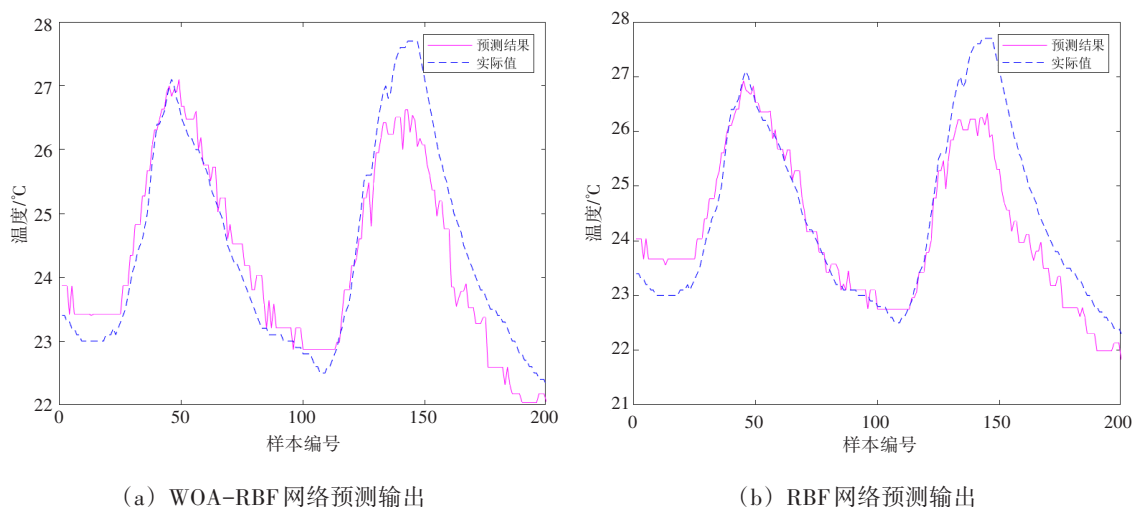


图3 WOA-RBF与RBF训练效果对比

由图3可知,WOA-RBF神经网络预测模型与未优化的RBF网络预测模型相比,WOA-RBF的预测效果更优,预测结果更贴近实际值。

两个模型的评价指标如表2所示,可以看出,WOA-RBF预测模型的各项误差指标均小于未优化的RBF预测模型。实验证明,与未优化的RBF网络预测模型相比,WOA-RBF神经网络预测模型具有更好的预测效果。

表2 模型的评价指标对比

评价指标	WOA-RBF	RBF
MAE	0.478	0.526
MSE	0.346	0.514
RMSE	0.588	0.717

3 结语

本文选择湿度、二氧化碳浓度、土壤氮含量、土壤磷含量和土壤钾含量这五种影响温室温度的影响因子作为预测模型的输入特征,通过鲸鱼优化算法对RBF神经网络的线性权重、神经元中心和高斯核宽度进行优化,建立了WOA-RBF神经网络温室温度预测模型。实验结

果表明,与未优化的RBF神经网络预测模型相比,WOA-RBF神经网络温室温度预测模型的预测误差更小、精度更高,对温室的管理有一定的参考价值。

参考文献:

- [1] 王定成. 温室环境的支持向量机回归建模[J]. 农业机械学报,2004(5):106-109.
- [2] 左志宇,毛罕平,张晓东,等. 基于时序分析法的温室温度预测模型[J]. 农业机械学报,2010,41(11):173-177,182.
- [3] 徐意,项美晶. 基于RBF神经网络的温室温度调控研究[J]. 农机化研究,2010,32(3):74-76.
- [4] 徐宇,冀荣华. 基于复数神经网络的智能温室温度预测研究[J]. 中国农机化学报,2019,40(4):174-178.
- [5] 夏爽,李丽宏. 基于PSO-RBF神经网络在温室温度预测中的应用[J]. 计算机工程与设计,2017,38(3):744-748.
- [6] 王媛媛. 基于改进PSO优化RBF神经网络的温室温度预测研究[J]. 计算机与数字工程,2016,44(7):1210-1215.

(下转第51页)

文章编号: 1007-1423(2023)11-0047-05

DOI: 10.3969/j.issn.1007-1423.2023.11.009

基于 EEMD-KELM 的淮安市二河溶氧量预测研究

陈肖*, 张强, 姚年春

(江苏财经职业技术学院智能工程技术学院, 淮安 223300)

摘要: 为提高淮安市二河溶氧量的预测精度, 设计了集合经验模态分解与核极限学习机结合的溶氧量预测模型。在极限学习机模型中添加核函数, 解决其预测稳定性较差的问题; 引入正则化项以避免系统性风险。使用集合经验模态分解解耦二河溶氧量序列, 对解耦后的各分量进行预测并叠加作为最终预测结果。将该模型与其他预测模型对比, 结果表明: 所提模型能够更精确地预测二河的溶氧量, 稳定性高。

关键词: 集合经验模态分解; 核函数; 极限学习机; 水质预测

0 引言

淮安市二河位于洪泽湖东侧, 其水量主要来自于洪泽湖的补给。作为淮安市重要的河流型水源地, 已经被纳入了水利部公布的第二批《全国重要饮用水水源地名录》^[1]。二河支撑着淮安市城南区域灌溉、养殖、供水等活动, 近年来, 洪泽湖透明度降低, 湖体富营养化严重, 因二河主要由洪泽湖通过二河闸供水, 二河水质也不断恶化^[2]。溶氧量(dissolved oxygen, DO)作为一项重要的水质参数, 随着季节的变化有规律的波动。DO 可以指示河流的受污染程度, 也能从中分析得出鱼、虾等水生生物的生存条件^[3]。因此, 对于 DO 发展趋势的精确预测, 对水源地的水质管理以及淮安市的经济及社会可持续发展具有重要意义, 同时也能为二河的饮用水源地保护提供一种新的思路。

基于上述分析, 本文将集合经验模态分解(ensemble empirical mode decomposition, EEMD)与核极限学习机(kernel extreme learning machine, KELM)组合, 运用在淮安市二河溶氧量

的预测中。将 EEMD 分解得出每个分量的 KELM 预测结果叠加, 作为整个组合模型的最终预测结果。与其他预测模型相比, 本文提出的模型预测精度高、泛化性强, 可以实现对二河水质的高精度预测。

1 EEMD-KELM 预测模型的构建

1.1 EEMD

EEMD 基于经验模态分解(empirical mode decomposition, EMD), 在 EMD 的基础上将数个高斯白噪声加入到待分解的信号中, 再将所有的分解结果相加取均值以得出最终的本征模态函数(intrinsic mode function, IMF), 以此优化 EMD 中的模态混叠问题, 具有良好的分解效果^[4]。IMF 分量利用三次样条插值法, 可以实现对原始序列的筛分, 第一个被筛分出来的 IMF 分量通常频率最高, 之后的 IMF 分量频率依次降低。当分解达到终止条件时, 剩余未被筛分的序列称之为残差。EEMD 的分解步骤如下:

(1) 将数个高斯白噪声信号 $\varepsilon^i(t)$ 加入到原

收稿日期: 2023-02-14 修稿日期: 2023-03-04

基金项目: 江苏财经职业技术学院校级教改课题(2022421053): 基于 HHT 与极限学习机的淮安市二河水质预测研究

作者简介: *通信作者: 陈肖(1996—), 男, 江苏淮安人, 硕士, 助教, 研究方向为数据处理、机器学习, E-mail: shawn-chen96@foxmail.com; 张强(1993—), 男, 江苏徐州人, 硕士, 助教, 研究方向为嵌入式系统、智能算法; 姚年春(1975—), 男, 山东青岛人, 硕士, 副教授, 研究方向为嵌入式系统

始信号序列 $x(t)$ 中,得到加噪信号 $x^i(t)$:

$$x^i(t) = x(t) \pm \varepsilon^i(t) \quad (1)$$

式中: i 表示加入的高斯白噪声的个数。

(2) 将加噪信号 $x^i(t)$ 按照EMD分解的步骤分解,分解过程参考文献[5],得到 i 个加噪信号的IMF分量 $c_{n,i}(t)$,然后对 $c_{n,i}(t)$ 进行整体平均,得出的第 n 个IMF分量可表示为

$$c_n(t) = \frac{1}{i} \sum_{i=1}^i c_{n,i}(t) \quad (2)$$

式中, n 为IMF分量的个数。对残差的处理同样参照上式,此处不再赘述。

对于白噪声的添加,根据EEMD算法提出者的要求,应满足:

$$\varepsilon_n = \frac{\varepsilon}{\sqrt{i}} \quad (3)$$

式中, ε_n 为IMF重构信号与初始信号的误差,经过查阅其他文献与实验验证,本文将 i 的取值设置为100, ε 的取值为0.1时有较好的分解效果,可以有效避免模态混叠的发生。

1.2 KELM模型

1.2.1 极限学习机

ELM算法是Huang等[5]提出的一种单隐层前馈神经网络算法,其输入层与隐含层的权值阈值随机产生,只需要设置好隐含层节点数即可直接得到隐含层到输出层的权值,其数学模型为

$$H\beta = Y \quad (4)$$

式中: H 为隐含层输出矩阵; Y 为输出矩阵; β 为隐含到输出层的连接权值矩阵。求解式(4)的最小二乘解:

$$\min_{\beta} \|H\beta - Y\| \quad (5)$$

解出 $\beta = H^+Y$ 。 H^+ 为隐含层输出矩阵的摩尔彭罗斯广义逆矩阵。

1.2.2 核极限学习机

KELM是基于ELM发展而来,传统的ELM算法为了实现高效快速的训练效果牺牲了一定的稳定性与泛化性。因此,Huang等[5]提出采用核函数 $h(x)$ 来代替随机映射 H ,由此将样本从复杂的低维度空间问题映射到高维空间转化为内积问题[6]。在此基础上,引入正则化系数 C 以提高模型的稳定性与泛化性,解决模型的系统性风险,进一步提升了模型的预测能力[7]。引入正则化系数 C 后,ELM学习过程由最小二乘最优解转变为求解二次规划问题,利用拉格朗日乘数法求解得出:

$$\beta^* = H^+T = H^T \left(HH^T + \frac{I}{C} \right)^{-1} T \quad (6)$$

式中, C 为正则化系数,根据Mercer's条件[8],将核矩阵定义为

$$\begin{cases} \Omega = HH^T \\ \Omega_{i,j} = h(x_i)h(x_j) = K(x_i, x_j) \end{cases} \quad (7)$$

式中, $K(*)$ 为核函数,本文选择高斯核函数,因此:

$$K(x_i, x_j) = \exp \left(-\frac{\|x_i - x_j\|^2}{g^2} \right) \quad (8)$$

式中, g 为核参数,KELM的预测输出为

$$\begin{aligned} \hat{y} &= h(x)\beta \\ &= [K(x, x_1), K(x, x_2), \dots, K(x, x_n)]^T \left(\frac{I}{C} + \Omega \right)^{-1} \cdot y \end{aligned} \quad (9)$$

可以看出,核参数 g 以及正则化系数 C 的选择十分重要,可以决定KELM的预测性能。

1.3 EEMD-KELM组合模型

EEMD-KELM二河水质DO值组合预测模型的整体预测流程如图1所示。

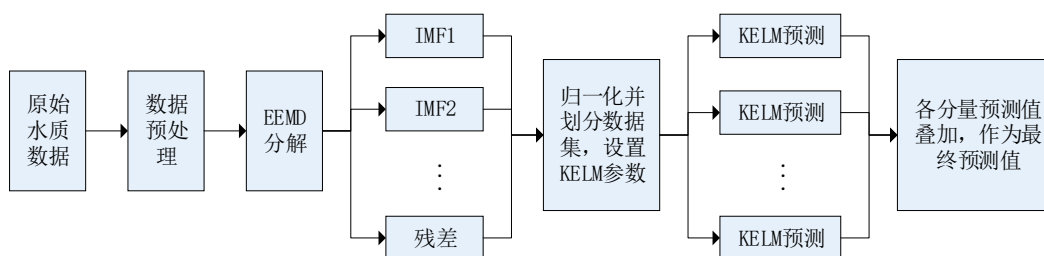


图1 EEMD-KELM预测模型整体流程

2 二河水质 DO 值的组合模型预测

2.1 水质数据处理

数据集使用淮安市环境局提供的2010年第1周至2020年第50周共537周的水质DO数据。利用箱线图法对水质DO值数据集中的异常值进行剔除。利用均值平滑法对异常值和缺省值进行修复，如果需要修复的数据连续分布，则基于相似日的方法，通过查询缺失数据所在周的气象信息与水文信息，寻找相似周数的水质数据的均值作为最终的修复值^[9]。

利用EEMD算法对二河的DO序列进行分解，利用1.1的数据设置好EEMD的相关参数，分解的结果如图2所示。

由图2可知，二河的DO序列被分解成7个IMF分量和一个残差，这些分量反映出了二河DO值的变化情况，使得原来难以预测的DO序列有规律可循。在这些分量中，IMF1与IMF2的频率最高，表现出的特征类似噪声，该分量反映出了二河DO值的随机变化，可以称之为随机分量。IMF3、IMF4与IMF5反映的是二河DO序列有规律的变化细节，易于拟合与掌握，统称为细节分量。IMF6、IMF7以及残差Res反映的是二河DO序列的整体变化趋势，称之为趋势分量。由于IMF6与IMF7的量纲较小，为了提升算法的运行效率，本文选择叠加趋势分量，将叠加后的分量作为一个新的分量带入模型预测，

简化模型预测过程。

这里以IMF1为例，利用分布式取值法构建数据集，即每10个为一组，一组内的前9个作为输入，第10个作为输出。因此，543个数据可以创建534组数据，取前510组作为训练集，剩下的24组为测试集。剩余的其它分量都以此方式创建数据集并按统一的标准设定训练集和测试集。

2.2 二河DO值预测对比分析

依然以IMF1为例，本文参照文献[10]，利用麻雀搜索算法对KELM的 g 、 C 、以及隐含层节点数进行寻优。设定麻雀搜索算法的最大迭代次数是100，种群大小是30，其中隐含层节点数的寻优范围设定为 $[10, 250]$ ，正则化系数 C 的范围设定在 $[0.001, 10]$ ，核参数 g 的范围设定在 $[0.001, 10]$ 。最终将IMF1分量的KELM参数设定为：107个隐含层节点，核参数 g 设定为2.054，正则化系数 C 设定为1.685。其它分量的设置方法与IMF1分量设置方法类似，此处不再赘述。使用KELM对各分量进行预测，为降低偶然性因数的影响，取预测结果30次的平均值作为最终预测结果，将各分量预测结果叠加，作为最终的预测结果。

为验证EEMD-KELM模型预测效果，将其与支持向量机(support vector machine, SVM)、KELM以及长短时记忆网络(long short time

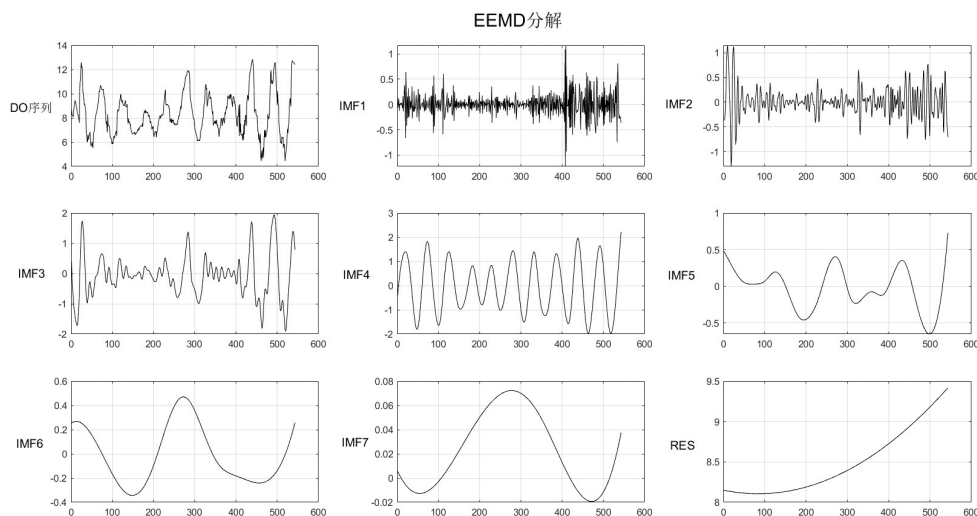


图2 二河DO序列EEMD分解结果

memory, LSTM)进行对比。为了保证结果的客观,在计算预测结果时同样采用30次实验的平均值作为最终结果。本次研究的评价指标选择均方根误差(root mean square error, RMSE)与平均绝对百分误差(mean absolute percentage error, MAPE)。EEMD-KELM与其他模型的预测数据比较结果如图3所示。为了直观地比较和分析各模型的预测性能,各个预测模型的RMSE及MAPE如表1所示。

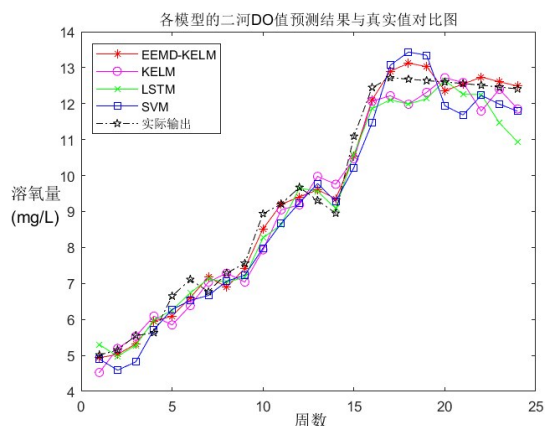


图3 二河DO值各预测模型预测值与真实值对比

表1 二河DO值各预测模型参数评价对比

模型	RMSE	MAPE/%
EEMD-KELM	0.3243	3.2914
KELM	0.5226	4.9581
LSTM	0.5341	4.6506
SVM	0.5780	5.5556

从图3可以看出, EEMD-KELM模型预测性能良好,能够以较高拟合度与精度对二河DO值进行预测。因此,“分解—预测—结合”的预测思路是十分有效的, EEMD有效地解耦了二河的水质DO值数据,使得KELM能够高效地实现对每个分量的预测,组合预测模型极大地提高了二河DO值的预测精度。

由表1可知, EEMD-KELM模型具有最高的预测精度。在与其他模型的比较中, EEMD-KELM模型的RMSE比SVM降低了43.89%;比KELM降低了37.94%;比LSTM降低了39.28%。其MAPE比SVM降低了40.76%;比KELM降低

了33.62%;比LSTM降低了29.23%。由此得出, EEMD-KELM模型预测精度高、泛化性强。能够较为准确地预测二河水质DO值的变化,可以为二河的水环境保护与治理等活动提供决策依据。

3 结语

本文将核函数与正则化系数引入极限学习机,提高了极限学习机的泛化性与预测稳定性。利用EEMD解耦二河水质数据,并将解耦所得分量分别代入KELM模型进行预测,将各分量预测结果叠加作为最终预测结果。结果表明,与其他预测模型相比,本文所提模型能够对二河DO值进行准确预测,泛化性强、预测精度高。

在水环境保护呼声渐涨的当下,如何实现对流域环境的高效治理与保护正逐渐成为一项重要的研究工作,而水质预测作为上述工作的基础,应具有较好的预测能力和稳定性,以此为从业人员提供参考。本文提出的模型能够实现上述目标,因此具有一定的使用前景。

在后续的研究中,可以将其他的水质参数与天气因素考虑在内,实现多因素的预测方法。另外,还可对解耦算法进行改进,提高随机分量的预测能力,从而进一步提高水质预测模型的预测精度和稳定性。

参考文献:

- [1] 崔璨,董增川,童建,等. 洪泽湖主要入湖河道水质评价及变化规律[J]. 江苏水利, 2021(9):27-34, 49
- [2] 石晴宜,董增川,罗赞,等. 基于机器学习方法的洪泽湖入湖水质评价及预测研究[J]. 中国农村水利水电, 2021(12):53-59.
- [3] 张腾龙,陈琬姪. 金坑岭水库与通济桥水库溶解氧含量及其对水质的影响研究[J]. 城镇供水, 2022(5):74-78, 8.
- [4] 刘扬,王立虎,杨礼波,等. 改进EEMD-GRU混合模型在径流预报中的应用[J]. 智能系统学报, 2022, 17(3):480-487.
- [5] HUANG Y, YU J, DAI X, et al. Air-quality prediction based on the EMD-IPSO-LSTM combination model[J]. Sustainability, 2022, 14(9):1-18.
- [6] 商立群,李洪波,侯亚东,等. 基于VMD-ISSA-KELM的短期光伏发电功率预测[J]. 电力系统保护与控制, 2022, 50(21):138-148.
- [7] 马良玉,程善珍,王永军. 正弦算法优化正则化

- ELM在NO_x排放量建模中的应用[J]. 华北电力大学学报, 2022, 49(3): 112-118.
- [8] GUO B, JUN L. Kernel extreme learning machine for indoor positioning in location fingerprinting [J]. Computer Engineering & Applications, 2016, 52(9): 78-83.
- [9] 王瑞, 闫方, 逯静, 等. 运用相似日和LSTM的短期负荷双向组合预测[J]. 电力系统及其自动化学报, 2022, 34(1): 93-99.
- [10] 马飞燕, 李向新. 基于改进麻雀搜索算法: 核极限学习机耦合算法的滑坡位移预测模型[J]. 科学技术与工程, 2022, 22(5): 1786-1793.

Prediction of dissolved oxygen in Huaian Erhe River based on EEMD-KELM

Chen Xiao*, Zhang Qiang, Yao Nianchun

(School of Intelligent Engineering, Jiangsu Vocational and Technical College of Finance and Economics, Huai'an 223300, China)

Abstract: In order to improve the prediction accuracy of dissolved oxygen in Huaian Erhe River, a prediction model of dissolved oxygen based on ensemble empirical mode decomposition and nuclear extreme learning machine was designed. Kernel function is added to the extreme learning machine model to solve the problem of poor predictive stability. Regularization terms are introduced to avoid systemic risk. The ensemble empirical mode was used to decompose the decoupled oxygen sequence, and each component after decoupling was predicted and superimposed as the final prediction result. Compared with other prediction models, the results show that the proposed model can predict the dissolved oxygen of the two Rivers more accurately and has high stability.

Keywords: ensemble empirical mode decomposition; Kernel function; extreme learning machine; water quality prediction

(上接第46页)

Research on greenhouse temperature prediction based on WOA-RBF

Li Qicao, Dong Zijian*

(School of Electronic Engineering, Jiangsu Ocean University, Lianyungang 222005, China)

Abstract: Temperature is the key environmental factor for crop growth in greenhouse, and it is of great significance to accurately predict the temperature change in greenhouse for greenhouse management. In order to improve the prediction accuracy, a greenhouse temperature prediction model based on the whale optimization algorithm (WOA) optimized RBF neural network is proposed. The whale optimization algorithm is used to optimize the neural linear weight of the RBF neural network output layer, the neural center of the hidden layer and the Gaussian kernel width of the hidden layer to obtain the optimal parameters, and finally obtain the optimized output results. The experiment shows that compared with the unoptimized RBF neural network prediction model, the result error of the WOA-RBF neural network greenhouse temperature prediction model is smaller. The mean absolute error (MAE), mean square error (MSE) and mean absolute percentage error (MAPE) are 0.478, 0.346 and 0.588, respectively, which are 9.1%, 32.7% and 18.0% higher than the unoptimized RBF neural network prediction model.

Keywords: whale optimization algorithm; RBF neural network; greenhouse temperature; prediction model

文章编号: 1007-1423(2023)11-0052-04

DOI: 10.3969/j.issn.1007-1423.2023.11.010

基于极限学习机算法的农业绿色智慧评价系统

王 琦^{1,2*}

(1. 郑州大学图书馆, 郑州 450001; 2. 郑州大学科学技术信息研究所, 郑州 450001)

摘要: 通过整理相关的研究数据, 建立了基于 2 个层次、4 个一级指标和 20 个二级指标的农业绿色智慧发展水平评价体系和神经网络样本集。基于该样本集对极限学习机算法进行训练, 建立了基于极限学习机算法的农业绿色智慧评价系统模型。并使用测试数据集对极限学习机的预测性能进行测试。测试结果表明, 该评价模型可以对农业绿色智慧发展水平进行有效的预测, 具有一定的应用价值和产业化前景。

关键词: 极限学习机; 机器学习; 评价系统

0 引言

2019 年, 中共中央办公厅、国务院办公厅印发《数字乡村发展战略纲要》^[1], 提出要大力发展农村数字经济和建设智慧绿色新农村。2021 年 9 月农业部编制《“十四五”全国农业农村信息化发展规划》, 提出要进一步细化推进数字农业建设思想、内容和重点任务, 其中, 包括大力推进物联网、大数据、人工智能、机器人等新一代信息技术在农业生产加工各环节融合应用。

人工智能技术(artificial intelligence, AI)是二十一世纪最重要的技术之一, 世界各国政府的工作报告中都多次提到人工智能技术。人工智能技术的研究目标是使机器像人一样思考, 完成需要人类智能才能完成的复杂工作流程, 使应用对象具有“智慧”能力。人工智能技术包含机器学习、机器视觉、语音分析等多个研究领域。人工神经网络(artificial neural network, ANN)从信息处理角度, 模拟人类脑神经元, 具有黑箱功能, 可以模拟自然界的数学规律, 作为机器学习的载体, 已经大规模应用于人类社会的各个领域。

近几年来, 关于人工智能和神经网络技术

在农业绿色智慧的应用研究, 正受到国内学者的广泛关注^[2-7]。李扬^[3]对机器学习技术在农业领域的应用现状进行了详细介绍。王肖芳^[4]使用反向传播(back propagation, BP)神经网络对河南省生态农业综合效益进行评价, 促进了生态农业的发展。吴晓柯等^[5]对区块链技术在农业发展中的应用展开了相关研究, 分析了区块链技术在农业应用中所面临的挑战。王翠^[6]对农业高新技术进行了综合评价, 采用 BP 神经网络构建了系统的评价模型, 并采用机器学习和自然语义处理技术, 实现了评价报告的自动生成。汝刚等^[7]针对使用人工智能对我国传统农业进行改造所存在的问题和改造方法进行了理论探讨研究。

绿色智慧农业(green and smart agriculture, GSA)是物联网技术、计算机技术、人工智能技术和农业的有机结合。绿色智慧农业就是为农民、政府和企业提供了科学的农业生产解决方案, 可以解决传统农业管理粗犷、效率低下的问题。通过物联网技术, 可以实时观测农作物的生长状况; 通过大数据和人工智能技术, 可以精确匹配农作物的生长环境, 精准地执行农业生产活动, 提高农业生产效率, 提升农业生产竞争力, 使经济效益最大化。通过利用物联

收稿日期: 2023-04-17 修稿日期: 2023-05-16

作者简介: *通信作者: 王琦(1986—), 男, 河南新乡人, 馆员, 研究方向为智慧图书馆建设、机器学习与大数据分析, E-mail: qwang@zzu.edu.cn

网技术、机器学习技术、大数据技术和人工智能技术，可以有效提升农产品质量和生产效率，绿色智慧农业是未来农业发展的必然方向。

1 农业绿色智慧发展水平评价体系的建立

采用层次分析法(analytic hierarchy process, AHP)建立了基于 2 个层次、4 个一级指标和 20 个二级指标的农业绿色智慧发展水平评价体系。

1.1 评价指标的选取

根据绿色智慧农业的本质特点，以河南省为例，通过对河南省下辖 17 个地级市进行调查走访研究，并查阅大量的文献研究资料，筛选出和河南省农业绿色智慧发展水平最相关的高频词汇，借鉴评价体系建立的方法建立了资源利用、生态效益、智慧程度和经济效益 4 个一级指标和 20 个二级指标。其中， $A_5 \sim A_7$ 为负向指标，其他均为正向指标。本研究样本所用指标数据来源于《河南省统计年鉴(2019)》和部分其他网络统计年鉴数据。评价指标的建立如表 1 所示。

表 1 评价指标及权重分配值

一级指标	二级指标	权重
资源利用 A	土地复种指数 A_1	0.0026
	有效灌溉指数 A_2	0.0014
	电量消耗指数 A_3	0.0016
	农用机械 A_4	0.0025
	农膜使用强度 A_5	0.0072
	化肥使用强度 A_6	0.0119
	农药使用强度 A_7	0.0140
生态效益 B	森林覆盖率(%) B_1	0.0111
	农村自来水普及率(%) B_2	0.0085
	农村卫生厕所普及率(%) B_3	0.0085
	绿色农产品认证数 B_4	0.0519
	绿色食品企业认证数量 B_5	0.0465
智慧程度 C	物联网终端用户数 C_1	0.2116
	移动电话交换机容量 C_2	0.3275
	铁路里程数 C_3	0.0696
	大气成分观测站数量 C_4	0.0656
经济效益 D	农村居民生产总值 D_1	0.0430
	农村居民消费水平 D_2	0.0403
	农林牧渔业就业人员平均工资 D_3	0.0175
	农林牧渔业增加值 D_4	0.0572

影响农业绿色智慧发展水平评价的因素有很多，总体评价结果由 20 个二级指标以及它们各自所占的权重共同决定。其中，权重是决定指标重要程度的重要数值。用 K 和 L 分别表示所有二级指标的数值矩阵和权重矩阵，用 R 表示评价结果，那么矩阵 K 与矩阵 L 的转置乘法运算后即可得到 R ，如公式(1)~(3)所示：

$$K = [A_1, A_2, A_3, \cdots, D_4] \tag{1}$$

$$L = [\omega_{A_1}, \omega_{A_2}, \cdots, \omega_{D_4}] \tag{2}$$

$$R = K \cdot L^T = A_1\omega_{A_1} + A_2\omega_{A_2} + A_3\omega_{A_3} \cdots D_4\omega_{D_4} \tag{3}$$

采用层次分析法构建农业绿色智慧发展水平评价体系需要构造权重矩阵 L 。首先，需要构建判断矩阵来确定各指标之间的相对重要性。以表 2 为例，引入“标度”的概念，来定量地衡量不同指标间相互之间的关系。其中， $P_{A_1A_2}$ 取 2，代表 A_1 (土地复种指数)与 A_2 (有效灌溉指数)相比， A_1 比 A_2 稍微重要。 $P_{A_1A_2}$ 表示的是指标 A_2 相对于 A_1 的比较结果，且与 $P_{A_1A_2}$ 存在倒数关系，因此取值 0.5。此外，同一个指标相对于自身具有同等的重要性，因此 $P_{A_1A_1}$ 和 $P_{A_2A_2}$ 皆为 1。

表 2 标度关系表

标度	指标 A_1	指标 A_2	
指标 A_1	$P_{A_1A_1} = 1$	$P_{A_1A_2} = 2$	
指标 A_2	$P_{A_2A_1} = 0.5$	$P_{A_2A_2} = 1$	
.....			

求得标度之后，通过 yaahp 软件进行一致性检验分析， CR 值小于 0.1，通过一致性检验后，可得指标的权重矩阵 L 。各指标权重如表 1 所示。

1.2 评价结果计算

根据前文所述，选取 2005—2018 年对应的二级指标数值矩阵 K 与权重矩阵 L 代入公式(3)，可得到 2005—2018 年农业绿色智慧发展水平的评价结果，具体结果如表 3 所示，由于二级指标数值矩阵中的数据不在一个量纲，因此需要对所有的数据进行归一化处理。

通过选取农业绿色智慧发展水平评价体系的指标变量和相应指标权重分配矩阵，构建了农业绿色智慧发展水平评价体系，并得到了每一年最

终的评价结果,该评价体系的建立,为下一部分基于极限学习机算法的农业绿色智慧发展水平评价模型的构建提供了相关的理论支持。

表 3 评价结果

时间(年份)	评价结果	时间(年份)	评价结果
2005	0.004	2012	0.357
2006	0.046	2013	0.385
2007	0.092	2014	0.477
2008	0.170	2015	0.565
2009	0.226	2016	0.607
2010	0.269	2017	0.735
2011	0.306	2018	0.942

2 基于极限学习机算法的农业绿色智慧发展水平评价模型构建

2.1 极限学习机算法

极限学习机(extreme learning machines, ELM)算法是新加坡南洋理工大学的 GuangBin Huang 教授于 2006 年首次提出的^[8]。极限学习机算法基于单隐含层前馈神经网络结构,该算法随机产生输入层和隐藏层之间的权值矩阵及隐藏层神经元阈值,在训练过程中,不需要每次反向调整权值矩阵及隐藏层神经元阈值,且隐藏层和输出层之间的权值矩阵是通过求解方程组而确定的。因此,相比于传统 BP 神经网络,具有训练速度快、泛化能力强等优点,使用极限学习机算法建立农业绿色智慧发展水平评价模型可以有效地对河南省农业绿色智慧的发展水平进行预测。单隐含层前馈神经网络结构如图 1 所示。其中,该神经网络包含一个输入层,节点数为 n ; 一个隐藏层,节点数为 j ; 一个输出层,节点数为 m 。其中, $\mathbf{X} = [x_1(k), x_2(k), \dots, x_n(k)]$ 为输入变量矩阵, $\mathbf{Y} = [y_1(k), y_2(k), \dots, y_m(k)]$ 为输出变量矩阵, $\mathbf{W}\mathbf{a}_1$ 为输入层和隐藏层之间的权值矩阵, $\mathbf{W}\mathbf{a}_2$ 为隐藏层和输出层之间的权值矩阵。

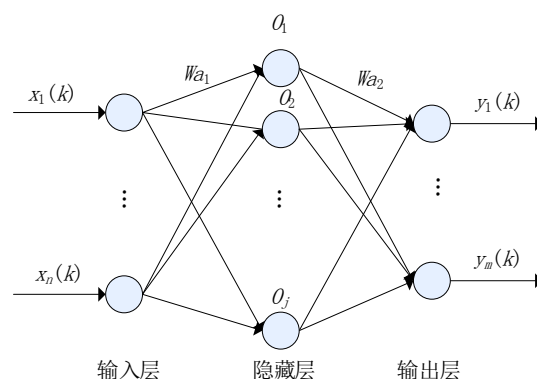


图 1 单隐含层前馈神经网络结构

2.2 基于极限学习机农业绿色智慧评价系统

根据因子分析后的样本集,引入极限学习机算法进行学习和训练,建立基于 ELM 的农业绿色智慧发展水平预测模型。ELM 评价模型建立步骤如图 2 所示。

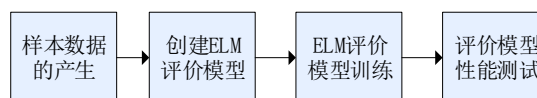


图 2 基于 ELM 的农业绿色智慧发展水平评价模型建立步骤

算法基于 Matlab2021a 仿真环境进行编程实现。首先,对样本数据进行处理。训练数据采用样本数据中 2005—2014 年的数据。由于样本数据量较小,因此,需要对训练数据集进行扩充。为了方便模型的学习和训练,将数据进行归一化处理。随后,建立 ELM 神经网络模型,这里采用 Matlab2021a 自带的神经网络工具箱软件建立 ELM 神经网络模型。建立模型后,使用训练数据集对极限学习机进行训练。训练停止的条件为满足设定的训练次数。设定训练次数完成后,极限学习机算法停止训练。下一步,需要使用测试数据集,测试基于极限学习机算法建立的农业绿色智慧发展水平评价模型的准确性。

2.3 评价模型的验证

训练完成后,使用 2015—2018 年的数据对极限学习机的预测性能进行测试。将测试数据

输入训练好的极限学习机，并得到极限学习机输出的预测结果，并和表3所得到的评价结果进行对比，结果如图3所示，可以看出，极限学习机的预测结果和基于层次分析法计算的结果基本重合，误差较小。

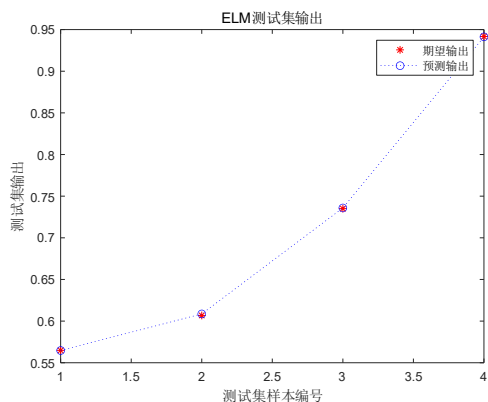


图3 极限学习机测试结果

3 结语

本课题以河南省为例，对河南省下辖17个地级市进行调查走访研究，查阅相关资料和文献，筛选出能反映河南省农业绿色智慧发展水平的变量指标，建立了具有2个层次、4个一级指标和20个二级指标的农业绿色智慧发展水平评价体系。其次，通过引入极限学习机算法，建立了基于极限学习机算法的农业绿色智慧发

展水平预测模型。最后，在Matlab2021a环境下对预测模型进行学习和训练，并使用测试数据集对极限学习机的预测性能进行测试。测试结果表明，该评价模型可以对农业绿色智慧发展水平进行有效的预测，具有一定的应用价值和产业化前景。

参考文献：

- [1] 中共中央办公厅, 国务院办公厅. 数字乡村发展战略纲要[R/OL]. (2019-05-16). https://www.gov.cn/gongbao/content/2019/content_5395476.htm.
- [2] 张治玲. 人工智能在农业领域的应用现状与发展趋势[J]. 农业科技与信息, 2020(19): 93, 97.
- [3] 李扬. 机器学习在农业领域应用现状与前景[J]. 安徽农学通报, 2021, 27(1): 164-165, 174.
- [4] 王肖芳. 基于因子分析-BP神经网络的河南省生态农业效益评价[J]. 经济经纬, 2015(4): 37-42.
- [5] 吴晓柯, 管孝锋, 黄海龙, 等. 基于区块链技术的智慧农业发展研究[J]. 南方农机, 2020, 51(22): 10-11.
- [6] 王翠. 农业高新技术综合评价理论与智能方法研究[D]. 北京: 中国农业科学院, 2020.
- [7] 汝刚, 刘慧, 沈桂龙. 用人工智能改造中国农业: 理论阐释与制度创新[J]. 经济学家, 2020(4): 110-118.
- [8] HUANG G, ZHU Q, SIEW C. Extreme learning machine: theory and applications[J]. Neurocomputing, 2006, 70(1-3): 489-501.

Agricultural green intelligent evaluation system based on extreme learning machine algorithm

Wang Qi^{1,2*}

(1. Library of Zhengzhou University, Zhengzhou 450001, China;

2. Institute of Scientific and Technical Information, Zhengzhou University, Zhengzhou 450001, China)

Abstract: The evaluation system of agricultural green intelligent development level based on 2 levels, 4 first-level indicators and 20 second-level indicators and neural network sample set were established by collating relevant research data. Based on this sample set, the extreme learning machine algorithm is trained, and the agricultural green intelligent evaluation system model based on the extreme learning machine algorithm is established. The predictive performance of the extreme learning machine is tested with the test data set. The test results show that the evaluation model can predict the development level of agriculture green intelligent effectively, and has certain application value and industrialization prospect.

Keywords: extreme learning machine; machine learning; evaluation system

文章编号: 1007-1423(2023)11-0056-04

DOI: 10.3969/j.issn.1007-1423.2023.11.011

列生成算法的改进策略研究

罗凤娥, 张 鑫*, 赵 强, 杨思瀚

(中国民用航空飞行学院空中交通管理学院, 广汉 618307)

摘要: 列生成算法是一种用于解决线性规划问题的高效算法, 其核心思想是通过逐步添加列来构建松弛问题的最优解。然而, 随着问题规模的增加, 算法的计算复杂度也会急剧增加, 导致算法的效率下降。为了优化这一问题, 经过对算法改进策略的研究, 分析其优缺点, 并给出基于机器学习的列生成算法改进模型。该策略的应用可以大大减少算法的计算时间和空间消耗, 提高算法的求解效率和精度。

关键词: 列生成; 改进策略; 图神经网络

0 引言

列生成算法是一种用于解决大规模线性规划问题的迭代算法。该算法的核心思想是通过逐步添加列来构建松弛问题的最优解, 从而逼近原问题的最优解。在算法的初期阶段, 只考虑少量的变量和约束条件, 随着计算的进行, 逐渐添加新的变量和约束条件, 直到得到原问题的最优解。在每次迭代中, 它解决了一个受限的主问题(restricted master problem, RMP)和一个定价问题(pricing issues, PP)。受限的主问题是原始线性规划(linear programming, LP), 被限制在其变量的子集中, 并由原始线性规划求解器求解, 例如原始单纯形算法。与传统的线性规划算法相比, 列生成算法在处理大规模问题时具有更高的效率和精度, 尤其适用于包含大量稀疏约束条件的问题。此外, 该算法还具有较好的可扩展性和灵活性, 可以通过调整算法参数和优化策略来满足不同的求解需求。

1 列生成算法的基本原理

列生成算法通过求解子问题(sub problem,

SP)来找到可以进基的非基变量, 该非基变量在模型中并没有显性地写出来^[1](可以看成是生成了一个变量, 每个变量其实等价于一列, 所以该方法被称为列生成算法)。如果找不到一个可以进基的非基变量, 那么就意味着所有的非基变量的检验数(reduced cost, RC)都满足最优解的条件, 也就是说, 该线性规划的最优解已被找到。其思路如下:

(1) 先把原问题(master problem, MP)转换到一个规模更小(即变量数比原问题少)的问题上, 这个只使用部分变量的模型被称为原问题的RMP问题。在RMP上用单纯形法求最优解, 注意此时求得的最优解只是RMP上的, 并不是MP的最优解。

(2) 然后需要通过一个子问题去检测在那些未被考虑的变量中是否有使得RC小于零的情况, 如果有, 那么就把这个变量的相关系数加入到RMP的系数矩阵中, 返回第一步。经过反复迭代, 直到子问题中的RC都大于等于零, 此时就找到了MP的最优解。流程如图1所示。

收稿日期: 2023-03-04 修稿日期: 2023-03-14

作者简介: 罗凤娥(1972—), 女, 重庆人, 硕士, 教授, 研究方向为航空公司运行控制; *通信作者: 张鑫(1996—), 男, 山东淄博人, 硕士研究生, 研究方向为航空公司运行控制, E-mail: 877566395@qq.com; 赵强(1998—), 男, 山东临沂人, 硕士研究生, 研究方向为航空公司运行控制; 杨思瀚(1999—), 男, 浙江绍兴人, 硕士研究生, 研究方向为航空公司运行控制

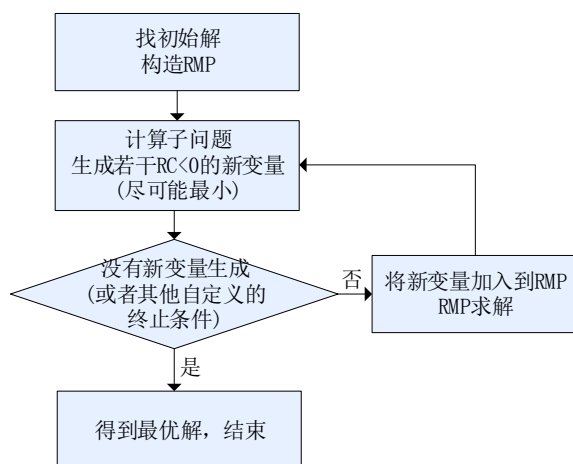


图1 列生成算法流程

2 列生成算法的改进思路

列生成算法的改进思路主要有以下四个方面。

2.1 加速松弛问题的求解

松弛问题的求解是列生成算法中的瓶颈，因此加速松弛问题的求解可以提高整个算法的效率。可以通过优化线性规划求解器、改进算法的分支策略和剪枝技术、引入启发式算法等方法来加速松弛问题的求解。

优点：通过优化线性规划求解器、改进算法的分支策略和剪枝技术、引入启发式算法等方法来加速松弛问题的求解，可以大幅提高算法的效率和精度。

缺点：一些优化方法可能会增加算法的复杂度，导致运算时间和空间开销增加。

2.2 改进列选择策略

列选择策略是列生成算法中的关键因素，决定了列的质量和数量，因此改进列选择策略可以提高算法的精度和效率。可以通过引入先进的列选择策略、基于贪心策略的启发式算法等方法来改进列选择策略^[2]。

优点：通过引入先进的列选择策略、基于贪心策略的启发式算法等方法来改进列选择策略，可以提高算法的精度和效率。

缺点：列选择策略的改进需要对问题本身有深入的理解和分析，可能需要大量的实验和测试，才能找到合适的列选择策略。

2.3 优化列集合的管理

列集合是列生成算法中的关键数据结构，对算法的效率和精度有着重要的影响。可以通过优化列集合的更新策略、管理策略、排序策略等方法来提高算法的效率和精度。

优点：通过优化列集合的更新策略、管理策略、排序策略等方法来提高算法的效率和精度，可以改善列生成算法的求解速度和质量。

缺点：列集合管理的优化可能会增加算法的复杂度，导致运算时间和空间开销增加。并且需要仔细选择优化策略，以避免出现不稳定的情况。

2.4 结合其他方法

列生成算法可以结合其他方法来进一步提高算法的效率和精度，如将列生成算法与分支定界算法、混合整数规划算法、模拟退火算法等方法相结合，可以得到更加高效和精确的求解方法。

优点：通过结合其他方法，如将列生成算法与分支定界算法、混合整数规划算法、模拟退火算法等方法相结合，可以得到更加高效和精确的求解方法。

缺点：不同方法的组合需要进行合适的调整 and 实验，才能找到最优的求解策略。而且，不同方法的结合可能会导致算法的复杂度增加，导致运算时间和空间开销增加。

总体来说，不同的思路和方法在不同的问题和情况下具有不同的优缺点。为了提高算法的效率和精度，需要根据具体问题的特点和求解需求来选择合适的改进方法，并对改进方法进行适当的组合和调整，同时进行大量的实验和测试，以验证改进方法的有效性和性能。

3 基于机器学习的列生成算法改进策略

在用列生成求解时，一些大规模模型容易出现高度简并，导致收敛速度较慢。对于这些问题，可以在每次迭代中生成许多列并添加到受限的主问题中。然而，这样做只会增加处理简并的难度，从而产生反效果。在过去的几年里，研究人员对使用机器学习算法来解决组合优化问题或加速现有的优化方法越来越感兴趣。Bengio等^[3]对组合优化的机器学习技术进行了很

好的概述。

3.1 列生成算法的基本模型

在每次列生成迭代中,优化受限的主问题并求解定价问题后,得到一组列 β 。目标是选择要添加到受限的主问题的列 $\beta^s \subseteq \beta$ 的子集。这种列的选择可以被认为是一个二元分类问题,其中我们尝试将每个生成的列分为两个类 $y \in \{0, 1\}$,即如果要选择列,则为1;否则为0。设 l 为列生成迭代号, α_l 为迭代开始时受限的主问题中出现的列集, β_l 为此迭代中生成的列。对于每一列 $p \in \beta_l$,我们定义了一个决策变量 y_p ,如果选择列 p 则取值1;否则取0。由此得到用于列分类任务的混合整数线性规划模型如下:

$$\min_{\theta, y} \sum_{p \in \alpha_l \cup \beta_l} c_p \theta_p + \sum_{p \in \beta_l} y_p \quad (1)$$

$$\text{s.t.} \sum_{p \in \alpha_l \cup \beta_l} a_p \theta_p = b \quad (2)$$

$$\theta_p \geq 0, \forall p \in \alpha_l \cup \beta_l \quad (3)$$

$$\theta_p \leq y_p, \forall p \in \beta_l \quad (4)$$

$$y_p \in \{0, 1\}, \forall p \in \beta_l \quad (5)$$

其中: p 是变量 θ_p 的指数集; $c_p \in \mathbf{R}$, $a_p \in \mathbf{R}^m$ 分别为 θ_p 的代价系数向量和约束系数向量; $b \in \mathbf{R}^m$ 是约束条件(2)的右边向量。注:这些约束条件中的一些或全部也可以是不等式。如果没有 y_p 变量和约束(4)和(5),模型(1)~(5)将恰好是迭代 $l+1$ 时的受限的主问题。假设足够小,这些变量和约束只考虑最小化所选列的集合。这些列对应于 $y_p = 1$ 的列,即 y_p 列与正值 θ_p 变量相关(约束(8))。因此,迭代 l 时受限的主问题中要添加的列子集为

$$\beta_l^s = \{p \in \beta_l: y_p = 1\} \quad (6)$$

3.2 图神经网络概述

图神经网络(GNNs)为图结构化数据的学习提供了一个有效的框架。它们可以用于节点分类、链接预测、图预测、文档分类甚至组合优化等任务。它们最初由Gori等^[4]提出,并由Scarselli等^[5]进一步发展。GNN的目标是以迭代的方式聚合来自邻居节点的信息来学习节点表示。然后,使用节点表示形式为节点分类任务生成输出标签。

GNN概述:给定一个图 $G = (V, E)$,其中 V

是节点的集合, E 是边的集合,每个节点 $v \in V$ 由其特征向量 x_v 来表示。目标是通过聚合来自邻居节点的信息,迭代地更新节点的表示。设 $h_v^{(k)}$ 是节点 $v \in V$ 在迭代 k 处的表示向量 $k = 0, 1, \dots, K$ 。设 $N(v)$ 为 $v \in V$ 作为相邻节点集合,如图2所示,通过聚合邻居节点的表示来迭代更新表示向量。在迭代 $k > 0$ 时,该聚合函数($aggr$)首先用于计算每个节点 $v \in V$ 的聚合信息向量 $a_v^{(k)}$:

$$a_v^{(k)} = aggr\left(\left\{\varnothing^{(k)}(h_u^{(k-1)}, h_v^{(k-1)}): u \in N(v)\right\}\right), \forall v \in V \quad (7)$$

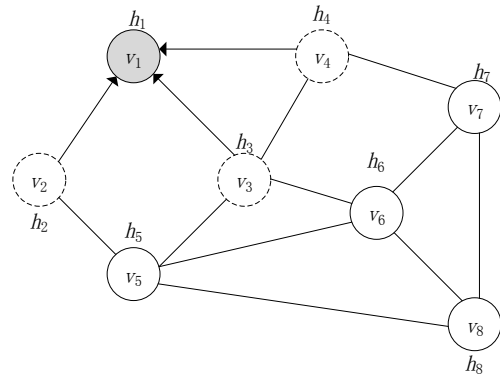


图2 图神经网络

$h_v^{(0)} = x_v$, $\varnothing^{(k)}$ 是一个学习函数。函数 $aggr$ 对于节点顺序的排列应该是不变的。然后,使用一个表示 $comb$ 的函数,我们将聚合的信息与给定节点的当前状态结合起来,以获得更新的节点表示向量:

$$h_v^{(k)} = \varphi^k\left(comb(h_v^{(k-1)}, a_v^{(k)})\right), \forall v \in V \quad (8)$$

其中 $\varphi^{(k)}$ 是另一个学习函数。

随着迭代的进行,节点从较远的节点收集信息。在最后的迭代 K 中,节点 $v \in V$ 的表示 $h_v^{(K)}$ 可以用来预测它的标签,记为 f_v ,通过应用最终习得的变换,记为 out ,即:

$$f_v = out(h_v^{(K)}), \forall v \in V \quad (9)$$

学习的函数 $\varnothing^{(k)}$, $\varphi^{(k)}$, out 通过使用多层感知器前馈神经网络实现。

3.3 算法模型

上述内容中存在明显的架构是用一个节点表示每一列,如果两个节点对应的列至少有一个共同的约束,就用一条边连接它们。但是,

在这种情况下，边的数量会非常大，模型中很难包含对偶值信息。另一种架构使用具有两种节点类型的二部图：列节点 V 和约束节点 C 。如果列 v 对约束 c 有贡献，则在节点 $v \in V$ 和节点 $c \in C$ 之间存在边 (v, c) 。这种表示的优点是可以将特征向量附加到约束节点上，这将允许我们轻松地包含双重信息。

因为有两种节点类型，所以每次迭代都分两个阶段进行，且在一次迭代之后，每个列节点都包含一些关于对相同约束有贡献的其他列的信息。然而，通过执行额外的迭代可以传递更好的信息。

列表示 $\mathbf{h}_v^{(K)}$, $v \in V$ ，然后用于预测标签 $y_v \in \{0, 1\}$ 。以下为算法示例：

1. $\mathbf{h}_v^{(0)} = x_v, \forall v \in V$
2. $\mathbf{h}_c^{(0)} = x_c, \forall c \in C$
3. for $k = 1, \dots, K$ do
4. $\mathbf{a}_c^{(k)} = \sum_{u \in N(c)} \phi_c^{(k)}(\mathbf{h}_c^{(k-1)}, \mathbf{h}_u^{(k-1)}), \forall c \in C$
5. $\mathbf{h}_c^{(k)} = \phi_c^{(k)}([\mathbf{h}_c^{(k-1)}, \mathbf{a}_c^{(k)}]), \forall c \in C$
6. $\mathbf{a}_v^{(k)} = \sum_{u \in N(v)} \phi_v^{(k)}(\mathbf{h}_v^{(k-1)}, \mathbf{h}_u^{(k-1)}), \forall v \in V$
7. $\mathbf{h}_v^{(k)} = \phi_v^{(k)}([\mathbf{h}_v^{(k-1)}, \mathbf{a}_v^{(k)}]), \forall v \in V$
8. end for
9. $y_v = out(\mathbf{h}_v^{(K)}), \forall v \in V$

4 结语

本文通过对列生成算法进行探索，针对列生成算法求解子问题的速度，进行相应的分析研究并提出对应的改进策略，同时通过相应的实验验证了本文提出内容的可行性。本文提出的改进策略对于提高列生成算法的计算速度具有很大的增益，也为各个应用领域运用列生成算法提供了一些改进思路。

参考文献：

- [1] 田倩南, 张成勇, 李杰, 等. 基于多因素不确定环境的受扰航班恢复建模与仿真[J]. 物流工程与管理, 2022, 44(5): 1-6.
- [2] 吴刚, 严俊. 不正常航班恢复的一种改进的列生成算法[J]. 南京航空航天大学学报, 2014, 46(2): 329-334.
- [3] BENGIO Y, LODI A, PROUVOST A. Machine learning for combinatorial optimization: a methodological tour d'horizon[J]. European Journal of Operational Research, 2021, 290(2): 405-421.
- [4] GORI M, MONFARDINI G, SCARSELLI F. A new model for learning in graph domains[C]//Proceedings of the IEEE International Joint Conference on Neural Networks, Montreal, Canada, 2005: 729-734.
- [5] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model[J]. IEEE Transactions on Neural Networks, 2009, 20(1): 61-80.

Research on improvement strategies for column generation algorithm

Luo Feng'e, Zhang Xin*, Zhao Qiang, Yang Sihan

(School of Air Traffic Management, Civil Aviation Flight University of China, Guanghan 618307, China)

Abstract: The column generation algorithm is an efficient algorithm for solving linear programming problems, whose core idea is to construct the optimal solution of the relaxed problem by gradually adding columns. However, as the size of the problem increases, the computational complexity of the algorithm also increases sharply, leading to a decrease in its efficiency. To optimize this problem, after researching the improvement strategies of the algorithm, analyzing its advantages and disadvantages, a column generation algorithm improvement model based on machine learning is proposed. The application of this strategy can greatly reduce the computation time and space consumption of the algorithm, improve its solving efficiency and accuracy.

Keywords: column generation; improvement strategies; graph neural network

实践与经验

文章编号: 1007-1423(2023)11-0060-04

DOI: 10.3969/j.issn.1007-1423.2023.11.012

一种基于功能点的项目资源日历计算方法

杨帆^{1,2}, 刘雅莹^{1,2*}, 唐海涛^{1,2,3}

(1. 中国电建集团华东勘测设计研究院有限公司, 杭州 311122;

2. 浙江华东工程数字技术有限公司, 杭州 311122;

3. 数字城市 CIM 技术浙江省工程研究中心, 杭州 311122)

摘要: 介绍了一种软件项目的资源日历计算方法。首先准备资源日历计算所需的数据, 包括相关需求文档、开发的功能清单、投入人员等, 然后根据数据进行项目人力资源统计和工作量、任务完成进度等指标的计算, 最后, 根据各人力资源和相关指标, 生成项目的资源日历。

关键词: 资源日历; 项目管理; 软件工程; 功能点

0 引言

在项目管理领域, 项目经理大多使用资源日历来协调人力资源、派发工作任务。项目资源日历展示了项目资源的可利用时间, 本文的项目资源指人力资源。在项目管理过程中, 如果项目经理对项目组内员工的工作任务、工作量和可利用时间不够了解, 极易出现工作任务分配不均的情况, 这就导致不同员工的工作量饱和程度有较大差别。所以资源日历的存在是极为必要的, 它可以展示人力资源的工作任务、工作量和可利用时间, 为管理者后续进行资源配置和任务分发提供了参考。它可以起到两个主要的作用, 首先保证每个员工能够有适当的工作饱和度, 其次是保证每个员工的工作都是有效率的。

近年来, 项目管理领域的研究者探究了多种人力资源调度、配置的方法, 如基于改进 CNN 的人力资源配置方法^[1]、针对多技能人员项目调度问题提出的改进布谷鸟搜索算法^[2]、在紧急状态下协同配置护理人员的网格化管理

方法^[3], 还讨论了资源受限情况下的多任务调度问题^[4]。但是对资源日历看板的研究较少, 目前项目的资源日历一般采用人工编辑的方式, 这种方法存在如下不足:

(1) 人工编辑资源日历的工作量较大, 而且随着项目任务的变化, 需要不断调整, 资源日历无法自动更新。

(2) 人工编辑资源日历只能确定短期内的资源投入, 无法预测多个功能模块或者整个项目的资源投入。

(3) 人工编辑资源日历时一般会对任务所需的工作量进行人工估算, 而人工估算工作量一般不够精准, 仍然会导致工作任务分配不均的情况。

(4) 人工编辑的资源日历无法定量计算任务的完成进度。

基于以上不足, 本文提出了一个基于功能点的自动生成项目资源日历的方法, 适用于软件工程行业。该方法无需人工编辑资源日历, 只需要拆分项目的功能, 并且批量设置每个功能的投入人员, 即可自动生成项目的资源日历。

收稿日期: 2023-02-16 修稿日期: 2023-03-06

作者简介: 杨帆(1982—), 男, 湖南益阳人, 高级工程师, 本科生, 研究方向为项目管理、软件工程; *通信作者: 刘雅莹(1998—), 女, 河北保定人, 本科生, 研究方向为软件工程、项目管理、质量管理, E-mail: liu_yx22@hdec.com; 唐海涛(1983—), 男, 江苏连云港人, 高级工程师, 硕士研究生, 研究方向为工程数字化

1 项目资源日历计算方法

1.1 数据准备

1.1.1 识别范围和边界

根据招标文件、用户需求说明书、需求规格说明书、原型图、详细设计说明书或数据库设计说明书等项目文档，识别所开发应用程序的范围，即包括在功能点统计范围内的功能需求/规格说明的集合。功能点是用来表示项目规模的计量单位。功能点计数可以对项目的规模进行测量，在确定项目的工作量投入和成本时，功能点数是一个必要的参数^[5]。

一旦确定了范围就可以定义边界，边界是应用程序和其用户的概念接口。其中，用户包括操作当前应用程序的人和与当前应用程序有交互的其他软件或硬件。

1.1.2 识别功能

在需求范围内，识别应用程序的数据功能和事务功能，形成项目各功能模块的功能清单。数据功能是指从用户的视角看到的数据的一个逻辑分组，包括内部逻辑文件(internal logical file)和外部接口文件(external interface file)两类。事务功能是指被用户看作属于一个工作单元的一连串动作，包括外部输入(external input)、外部输出(external output)和外部查询(external inquiry)三类^[5]。

1.1.3 设置投入人员和开发状态

为功能清单中的每个功能设置投入人员，并选定开发状态，即开发完成或未开发完成。投入人员具有角色属性，角色包括需求分析、设计开发、系统测试和运维实施四类。设置功能清单中各功能的投入人员和开发状态不仅包括初始化设置，还包括随着项目进展更新各功能的投入人员和开发状态。

1.2 资源统计

基于上述功能清单及投入人员等数据，统计出项目的人力资源列表。首先，筛选项目下所有的投入人员。然后根据不同的投入人员，筛选某投入人员在该项目下参与的所有功能模块。根据员工的角色属性和员工参与的功能模

块，生成该员工负责的任务的名称。如员工A的角色属性是“设计开发”，他参与了“成本管理”和“计划管理”模块的工作，那么员工A有两个任务，分别是“设计开发-成本管理模块”和“设计开发-计划管理模块”。

1.3 指标计算

计算生成资源日历所需的数据指标，包括每个员工的工作任务所需工作量、计划完成日期、计划开始日期和任务完成进度。

1.3.1 所需工作量

根据不同的功能模块，筛选某投入人员在功能模块下参与的功能清单，并计算完成各个功能模块所需的工作量。具体包括：

(1) 根据不同的功能模块 j ，筛选某投入人员 i 在该功能模块参与的功能清单 $List_{ij}$ 。

(2) 功能清单 $List_{ij}$ 中，检查每一个功能的功能类别(EI、EO、EQ、ILF、EIF)，以及该功能下与该投入人员具有相同角色的投入人员数量 N 。

(3) 根据该功能的功能类别，得到该功能的功能点数 A 。假设该功能的功能类别为内部逻辑文件，则该功能的功能点数 A 为7^[5]。

(4) 计算出该功能在该投入人员的角色职责下的平均功能点，例如该投入人员的角色为设计开发，该功能共投入了3个角色为设计开发的人员，则该投入人员在该功能所消耗的功能点为 $\frac{A}{3}$ 。

(5) 计算得出该投入人员 i 在功能模块 j 下所有功能投入的功能点之和 U_{ij} 。

(6) 计算得出该投入人员 i 完成功能模块 j 所分配功能的工作量 WL_{ij} ^[6]。

$$WL_{ij} = U_{ij} \times VAF \div C \quad (1)$$

其中， VAF 和 C 分别为项目设置的规模调整因子和各个角色的生产率。

1.3.2 计划开始时间和计划完成时间

其中，任务的计划完成时间(PF)需要人工填写。计划开始时间(PS)可以人工填写，也可以根据填写的计划完成日期和求得的工作量得到计划开始日期。

$$PS = PF - WL_{ij} + 1 \quad (2)$$

1.3.3 任务完成进度

根据功能清单中各功能开发状态, 计算实际完成进度(AC), 包括单个任务 j 的完成进度 AC_{ij} 和员工的总体任务完成进度(AC_i)。

首先, 计算 i 员工功能清单 $List_{ij}$ 中的功能点总数 FP_{ij} ,

$$FP_{ij} = 7 \times ILF + 5 \times EIF + 4 \times EI + 5 \times EO + 4 \times EQ \quad (3)$$

其次, 计算 i 员工功能清单 $List_{ij}$ 中的已完成的功能点总数 $FP_{ij_finished}$,

$$FP_{ij_finished} = 7 \times ILF_finished + 5 \times EIF_finished + 4 \times EI_finished + 5 \times EO_finished + 4 \times EQ_finished \quad (4)$$

员工 i 的单个任务 j 完成进度 AC_{ij} 计算如下:

$$AC_{ij} = FP_{ij_finished} \div FP_{ij} \quad (5)$$

统计员工 i 任务进度为100%的任务数量 $Task_finished$ 和任务总数 $Task$, 则员工 i 的总体任务完成进度 AC_i 计算如下:

$$AC_i = Task_finished \div Task \quad (6)$$

1.4 资源日历生成

根据上述公式所得到的投入人员、功能模块、计划开始日期、计划结束日期、工作量信

息、任务完成进度等信息生成该项目的资源日历。资源日历生成之后, 用户可对自动生成的资源日历进行检查, 根据实际需求重新调整计划开始日期和计划结束日期。

2 方法实现

使用本文提出的项目资源日历计算方法在 K 项目中应用, 现有5名人员共同完成一个软件项目, 包括A、B、C、D、E五个功能模块。表1截取了部分数据。

表1 示例数据

功能模块	功能名称	类别	投入人员	开发状态
A	功能1	内部逻辑文件	User1、User2	未完成
A	功能2	内部逻辑文件	User1、User2	未完成
A	功能3	外部接口文件	User1、User2	未完成
A	功能4	外部输入	User1、User2	未完成
A	功能5	外部输出	User2、User2	未完成
A	功能6	外部查询	User2、User2	未完成
A	功能7	外部输入	User2、User2	未完成
A	功能8	外部输出	User2、User2	已完成
.....				

根据上述项目资源日历计算方法, 得到该项目的资源日历, 如图1所示。

项目资源日历看板

序号	责任人	任务名称	计划开始日期	计划结束日期	工作量(天)	任务完成进度
1	√ User1	设计开发	2022-03-01	2022-03-23	23	1/3
1.1		设计开发—A模块	2022-03-01	2022-03-12	12	100%
1.2		设计开发—B模块	2022-03-13	2022-03-18	5	8%
1.3		设计开发—C模块	2022-03-16	2022-03-23	6	0%
2	> User2	设计开发	2022-03-01	2022-03-23	23	1/3
3	√ User3	需求分析	2022-03-02	2022-03-24	23	0/2
3.1		需求分析-A模块	2022-03-02	2022-03-16	15	95%
3.2		需求分析-B模块	2022-03-17	2022-03-24	8	0%
4	> User4	系统测试	2022-03-02	2022-03-24	23	1/3
5	> User5	运维实施	2022-03-15	2022-03-27	13	0/1

图1 项目资源日历看板

3 结语

本文重点介绍了一种软件项目的资源日历计算方法, 包括数据准备、资源统计、指标计

算、资源日历生成。该方法可以很好地支持项目管理者查看项目人力资源的可利用时间, 帮助项目管理者更好地进行任务分配和人员调度。

参考文献：

- [1] 雷江平,许涛,王志勇. 基于改进CNN算法的人力资源自动配置方法[J]. 自动化技术与应用, 2022, 41(12):142-146.
- [2] 段鹏飞,余杰,聂慧,等. 求解广义优先关系下多技能人员项目调度问题的改进布谷鸟搜索算法[J]. 计算机应用研究, 2018, 35(5):1315-1319.
- [3] 林陶玉,方鹏骞. 疫情防控紧急状态下医院护理人员资源配置与动员管理策略[J]. 中国卫生事业管理, 2020, 37(5):332-334.
- [4] 齐小刚,陈玲琳,宋卫星,等. 面向多维修中心的资源受限任务调度问题研究[J]. 智能系统学报, 2022, 17(4):661-669.
- [5] 上海宝信软件股份有限公司,上海旋思科技有限公司,上海翰纬信息技术有限公司,等. SJ/T 11619—2016,软件工程 功能规模测量 NESMA方法[S]. 北京:中华人民共和国工业和信息化部, 2016.
- [6] 中关村智联软件服务业质量创新联盟,中国电子技术标准化研究院,广州赛宝认证中心服务有限公司,等. GB/T 36964—2018,软件工程 软件开发成本度量规范[S]. 北京:国家市场监督管理总局,中国国家标准化管理委员会, 2018.

A method for calculating project resource calendar based on function points

Yang Fan^{1,2}, Liu Yaxuan^{1,2*}, Tang Haitao^{1,2,3}

(1. Powerchina Huadong Engineering Co., Ltd., Hangzhou 311122, China;

2. Zhejiang Huadong Engineering Digital Technology Co., Ltd., Hangzhou 311122, China;

3. Zhejiang Digital City CIM Technology Research Center, Hangzhou 311122, China)

Abstract: This paper proposed a resource calendar calculation method for software projects. To begin with, prepared the data required for the resource calendar calculation, including the relevant requirements documents, the developed function list, the input personnel, etc., and then calculated the project human resources statistics and workload, task completion progress and other indicators according to the data. Finally, generated the project resource calendar according to the human resources and related indicators.

Keywords: resource calendar; project management; software engineering; function point

文章编号: 1007-1423(2023)11-0064-04

DOI: 10.3969/j.issn.1007-1423.2023.11.013

基于区块链的电子病历访问控制方法研究

刘荣管¹, 王兆栋¹, 邓兰华², 陈长源¹, 陈佩江^{3*}(1. 临沂亿通软件有限公司产品开发部, 临沂 276000; 2. 临沂市大数据协会, 临沂 276000;
3. 临沂大学机械与车辆工程学院, 临沂 276000)

摘要: 电子病历等医疗数据是宝贵资源, 其共享和使用为开展科研和制定决策提供支持, 但在这过程中可能会侵犯患者隐私权。为了共享电子病历和保护患者隐私, 利用区块链技术研究了电子病历的访问控制方法。分析了区块链的基本结构, 设计了电子病历访问控制模型, 采用区块链和密码技术, 防止电子病历数据被篡改和泄露。实验结果表明, 本访问控制方法能够抵抗身份伪装攻击, 防止隐私泄露, 实现医疗数据的安全共享和高效访问。

关键词: 区块链; 电子病历; 访问控制; 共享模型; 医疗数据安全

0 引言

近年来, 随着计算机和信息技术的快速发展, 电子病历已基本取代纸质文本, 给患者就医、医院管理都带来了极大便利。病历的电子化既解决了纸质文本的存放和丢失问题, 也为科学研究提供了宝贵资源。然而, 随着智慧医疗和医疗大数据应用的快速发展, 仍存在一些障碍。一是医疗数据的安全存储, 电子病历一般由医疗机构集中存储在单位数据库, 容易受到单点攻击, 导致患者数据丢失和隐私泄露^[1]; 二是医疗数据共享, 各医疗机构的信息系统不兼容, 沟通不畅, 担心信息泄露, 不愿分享信息, 形成数据孤岛; 三是数据所有权, 患者是电子病历的所有者, 希望有权管理个人数据, 并可以根据请求者身份和内容决定是否授予相应权限。

医疗数据的共享和访问新技术主要有两种方法: 云计算和区块链。云计算是指把医疗数据上传到云端实现存储和共享, 有效减少本地

存储成本, 方便用户访问数据, 但这种方法的隐患是如果云端服务器出现故障或受到恶意攻击, 会引起数据丢失或泄露^[2]。区块链是一种由许多节点通过一致性协议进行维护的分布式账本, 具备去中心化、可追溯、不可篡改、集体维护等特点, 医疗数据可以在互不信任的环境下保证安全和使用的可追溯性, 电子病历等数据可以在众多医疗机构之间安全传输和共享, 实现医疗体系的互联互通, 提高医疗服务质量和准确的远程会诊^[3], 为医疗数据共享所面临的问题提供了新的解决方法。

电子病历等医疗数据的共享与应用具有极其重要的意义, 本文基于区块链技术, 提出一个电子病历的访问控制方案, 实现医疗数据的共享和隐私保护, 促进医疗大数据的应用。

1 区块链技术

区块链是一种链式数据结构, 以时间为顺序把数据区块相连接, 其核心是分布在 P2P 网络中的账本, 用于存储交易公共记录, 对所有

收稿日期: 2023-02-17 修稿日期: 2023-03-05

作者简介: 刘荣管(1977—), 男, 山东临沂人, 本科, 工程师, 研究方向为医疗信息化; 王兆栋(1977—), 男, 山东临沂人, 本科, 工程师, 研究方向为医疗信息化; 邓兰华(1972—), 男, 山东临沂人, 本科, 高级工程师, 研究方向为大数据分析; 陈长源(1989—), 男, 山东临沂人, 本科, 助理工程师, 研究方向为医疗信息化; *通信作者: 陈佩江(1977—), 男, 山东临沂人, 研究生导师, 博士, 教授, 研究方向为医疗信息化, E-mail: chenpeijiang@163.com

用户共享以及验证交易。

1.1 区块链基本结构

作为一种数据结构, 区块由区块内结构与区块间链式结构组成。一个区块结构含有头信息和体信息。头信息是元数据, 包括时间戳、前区块哈希值、随机数、Merkle 根值等, 用于验证区块, 并把其前驱和后继区块联系起来; 体信息用于交易序列^[4]。哈希值作为区块的唯一标识, 把区块链接起来, 前区块哈希值决定了所链接的唯一区块, 最终构成一种链状数据结构。

1.2 共识机制

作为分布式系统, 为了让所有主机都能达到一致性, 区块链利用了共识机制, 常用算法包括: 权益证明(PoS)、工作量证明(PoW)、委托权益证明(DPoS)和实用拜占庭容错算法(PBFT)等。

在委托权益证明区块链中, 节点能根据其权益来选择证人^[5], 获得最高票数的前 N 个节点具有记账权利, 至少有 50% 的投票利益相关者才有足够的权力下放。DPoS 的每个客户端都有权利选择可信任节点^[6], 被选中节点可轮流记账, 有效地减少了验证节点的个数, 既能保证网络安全, 又提高共识验证效率, 降低运行成本, 更高效、节能和去中心化^[7]。

1.3 智能合约

智能合约被定义为是可以自动执行的电子版合同, 其实智能合约就是代码程序, 它被部署在区块链区块中, 当一个预先编制好的条件被触发时, 智能合约自动执行该工作, 目的是用于验证和执行合约。智能合约协商签订后, 如果有用户发起交易, 会向区块链网络提交智能合约, 经网络传输和矿工验证存储在特定区块, 向用户返回地址和接口等信息, 可调用合约^[8]。合约代码检查信息、判断是否满足触发条件, 并执行响应、更新状态。

智能合约适用于各种不同的区块链网络, 并可应用在电子商务、供应链优化等具体场景中, 智能合约执行的自动化和智能性提高了交易的安全性和效率。

2 基于区块链的电子病历访问控制模型

2.1 电子病历基本数据

在基于区块链的电子病历系统中, 产生新区块的医院账号、患者号码和加密的病历内容写入区域数据库。如果某患者在多个机构就诊, 区域数据库将存储若干条记录。电子病历类型多样, 文本内容丰富, 涉及的数据既有健康数据, 也有非健康数据, 前者与个人身体状况相关。有些数据涉及患者隐私, 比如个人属性数据、健康状况数据、医疗应用数据、医疗资金和支付数据等, 这些数据的隐私情况不同, 利用价值也有区别。

为有效控制电子病历访问, 按数据敏感程度把患者隐私信息分为三个等级: 一级敏感度最高, 对应权值最大, 二级、三级的相应权值依次减小。一级隐私信息主要是病人的身份信息, 比如: 姓名、性别、身份证号、家庭住址等; 二级隐私信息主要为患者医疗记录; 三级隐私信息主要为患者诊断辅助数据。将患者的各级隐私逐条分块处理, 加密存储。

2.2 基于区块链的电子病历访问控制模型

为了实现电子病历的共享和访问控制安全, 基于区块链设计一个电子病历共享模型, 其中主要有四个实体。

(1) 权威中心, 即密钥生成中心。由区块链中的医院、医学研究院等机构联合组成, 负责系统主密钥的生成, 并和各机构协商生成对应私钥^[9]。

(2) 数据拥有者, 即数据的所有者。患者拥有自己的电子病历数据, 一般授权医生生成病历, 加密存储在区块链中。在数据共享过程中, 负责生成代理重加密密钥, 并发送给代理者^[10]。

(3) 数据用户, 即各医疗机构。医疗机构如有需求, 经授权后发送请求, 获取重新加密电子病历文本, 并可利用私钥来解密阅读。

(4) 代理者, 即执行重加密的节点。根据委托权益证明共识算法所推举出来的矿工节点作为代理节点, 对数据拥有者发送的密文重新加密。

2.3 基于角色的电子病历访问控制流程

医疗信息化系统拥有大量用户，各有其特点，可利用用户特点制定访问控制模型，更好地满足医疗电子病历系统的要求。此模型主要有如下几个基本要素：用户、角色、权限和会话；为用户分配角色，角色具备相应权限，在系统应用中，角色把用户与权限之间联系起来，用户通过其所处角色具有了相应权限。

为做好电子病历的隐私信息保护，基于用户角色实现电子病历的访问控制，主要过程如下：①根据要求的安全参数，进行系统初始化；②根据系统初始化结果，生成主密钥，得到主公钥和主私钥；③根据主密钥，生成用户私钥；④对电子病历文本进行加密；⑤对电子病历文本进行解密。

将制定的访问控制流程用智能合约来实现，降低了管理员的工作量，提高了隐私数据安全性^[11]。本智能合约定义了一系列角色与相关访问权限，不同角色有不同权限，不同权限用户可操作不同的数据检索指令。当数据访问者的角色满足权限要求，经身份认证后，方能满足医疗信息访问请求。

基于角色的电子病历访问控制流程如图1所示，主要有六个步骤：身份验证、处理访问信息、访问权限验证、信息量计算、访问控制策略和访问记录修改等。

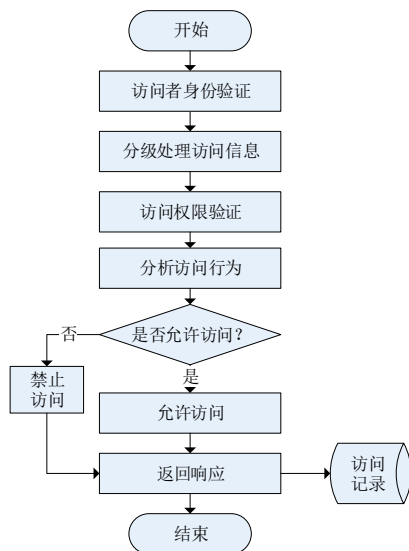


图1 基于角色的EMR访问控制流程

2.4 电子病历访问控制性能测试

电子病历访问控制模型对不同用户进行了角色和权限的划分，利用区块链技术，电子病历数据分块存储，具有文件存储安全、防止数据篡改和盗窃，制定动态访问策略，可以实现对医疗数据的完全控制，有效保护患者隐私。

为测试电子病历访问控制模型的性能，利用访问行为的准确率来验证安全性，主要指标有正常访问行为准确识别率、非法访问行为准确识别率和总访问行为准确识别率。

测试数据为公司所用电子病历数据库，在实验中，选取120次电子病历访问行为进行测试。在每次测试中，设置为不同角色，既包括本科室的合法访问，也有跨部门的非法访问，其中，符合访问控制策略的行为，即正常访问数量为40；不符合控制策略，即非法访问数量为80。电子病历访问控制的测试结果如图2所示。

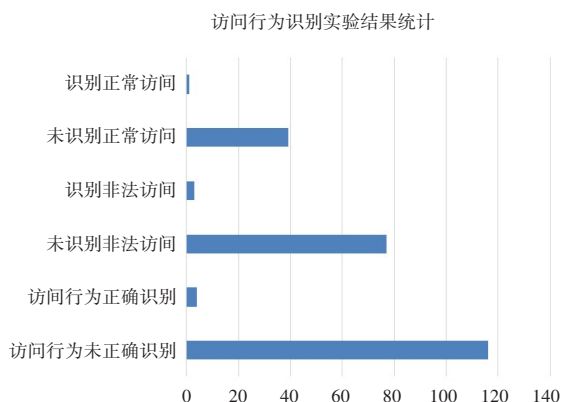


图2 电子病历访问行为识别结果

根据图2测试结果，经分析可知：正常访问、非法访问和总访问的行为识别准确率分别为97.5%、96.25%、96.7%。同时可发现，尽管电子病历访问行为正确识别率不是100%，但由于对每次访问都有记录，可以有效约束访问行为，防止医疗数据隐私泄露。

3 结语

鉴于医疗大数据应用的意义，电子病历等医疗数据需要在各医疗机构中传输和共享，医疗数据的安全性和患者隐私保护是重要的研究热点。考虑到区块链的去中心化和防篡改等特点，基于用户角色，提出了一种电子病历访问

控制方案,提高医疗信息的隐私保护。

区块链的信息安全以私钥安全为前提,患者私钥一旦丢失,就需要重新考虑隐私安全性;本方案是以DPoS算法为基础实现的,其效率和应用效果有待进一步提高。此外,由于性能测试采用的公司测试数据库,电子病历规模比较小,测试结果难免会受到一定影响,在后面研究中,需要改进算法和模型,增加测试数据集,提高访问控制效果。

参考文献:

- [1] 周正强. 基于区块链的医疗数据安全共享模型研究[D]. 贵州:贵州大学,2021.
- [2] 周正强,陈玉玲,李涛,等. 基于联盟链的医疗数据安全共享方案[J]. 应用科学学报,2021,39(1):123-134.
- [3] 吴思桦. 基于区块链的医疗隐私保护应用的研究[D]. 重庆:重庆邮电大学,2019.
- [4] 谢辉,王健. 区块链技术及其应用研究[J]. 信息安全,2016(9):192-195.
- [5] 崔梦. 基于区块链的电子病历访问控制研究[D]. 石家庄:石家庄铁道大学,2020.
- [6] 余攀. 基于区块链的电子病历隐私数据保护共享研究[D]. 赣州:江西理工大学,2020.
- [7] 汪一景. 基于区块链的电子病历隐私保护与共享机制研究[D]. 西安:西安邮电大学,2021.
- [8] 刘芬,朱壮友,许勇. 基于区块链的电子病历数据共享模型研究[J]. 安徽师范大学学报(自然科学版),2021,44(6):536-544.
- [9] 罗文俊,闻胜莲,程雨. 基于区块链的电子医疗病历共享方案[J]. 计算机应用,2020,40(1):157-161.
- [10] 闻胜莲. 基于区块链的医疗数据共享方案研究[D]. 重庆:重庆邮电大学,2020.
- [11] 张琳浩. 基于信息熵的电子病历隐私保护访问控制研究[D]. 石家庄:石家庄铁道大学,2019.

Study on the access control method of electronic medical records based on blockchain

Liu Rongguan¹, Wang Zhaodong¹, Deng Lanhua², Chen Changyuan¹, Chen Peijiang^{3*}

(1. Product Development Department, Linyi Yitong Software Co., Ltd., Linyi 276000, China;

2. Linyi Big Data Association, Linyi 276000, China;

3. School of Mechanical and Vehicle Engineering, Linyi University, Linyi 276000, China)

Abstract: Medical data such as electronic medical records are valuable resources, and their sharing and use provide support for scientific research and decision-making, but in this process, patients' privacy may be violated. In order to share electronic medical records and protect patient privacy, the access control method of electronic medical records is studied based on blockchain technology. The basic structure of blockchain is analyzed, and the access control model of electronic medical record is designed. The blockchain and cryptography technologies are used to prevent the electronic medical record data from being tampered with and leaked. The experimental results show that this access control method can resist identity camouflage attacks, prevent privacy disclosure, and achieve secure medical data sharing and efficient access.

Keywords: blockchain; electronic medical record; access control; shared model; medical data security

文章编号: 1007-1423(2023)11-0068-04

DOI: 10.3969/j.issn.1007-1423.2023.11.014

BP 神经网络在测控设备自动调平系统中的应用

薛缠明*, 赵翔宇

(太原卫星发射中心, 太原 036301)

摘要: Ka 频段车载测控设备在航天领域的快速发展, 促进了电子水平传感器自动调平系统在车载测控设备的广泛应用。目前设备都是在跟踪目标前利用调平系统传感器反馈的信号进行调平, 并计算出补偿误差, 而在天线运转过程中重心变化又引入了一定的系统误差, 降低了设备的外测精度。利用 BP 神经网络在天线实时跟踪目标时电子水平传感器反馈的信号预测误差, 进行动态补偿。通过实际验证, 证明可以减少设备测量误差和达到提高测量精度的目的。

关键词: 测控设备; 自动调平系统; BP 神经网络

0 引言

随着 Ka 频段车载测控设备在航天领域的快速发展, 技术部门对设备测量数据的精度要求越来越高, 车载测控设备都配置了电子水平传感器自动调平系统^[1]。此前测控设备都是利用电子水平传感器反馈的水平误差信号, 在跟踪目标前进行一次调平, 来保证设备的测量精度。但在跟踪过程中由于设备自身重心的变化使得设备水平度在一定范围内变化, 导致设备测量数据的精度较低。尤其是 Ka 频段设备天线波束较窄, 跟踪精度高, 天线水平度的较小变化, 都会导致测量的外测俯仰角度有一定误差。鉴于 BP 神经网络具有非线性的预测特点, 而且在预测方面具有高时效和高精度的能力^[2], 针对上述出现的问题, 本文提出以 BP 神经网络为基础, 在天线跟踪目标过程中利用电子水平传感器反馈的信号实时预测补偿误差, 动态进行补偿, 进一步有效减小重心变化导致的测量误差。

1 电子水平传感器自动调平系统简介

电子水平传感器自动调平系统如图 1 所示,

电子水平传感器安装在天线底座中心位置, 自动检测底座水平度, 利用 PID 控制原理控制设备四个支撑腿的电机进行调平。为了保证测量的俯仰角等外测数据在要求的误差范围内, 目前设备不能进行跟踪过程的动态调平计算补偿误差, 只能在跟踪目标前将天线俯仰角固定在 45° , 方位角从 $0^\circ \sim 360^\circ$ 每 30° 进行一次调平, 并将所得的 12 组数据取平均值得到补偿误差。最后天线在跟踪目标过程中将测量的俯仰角和补偿误差值送到中心进行处理分析。

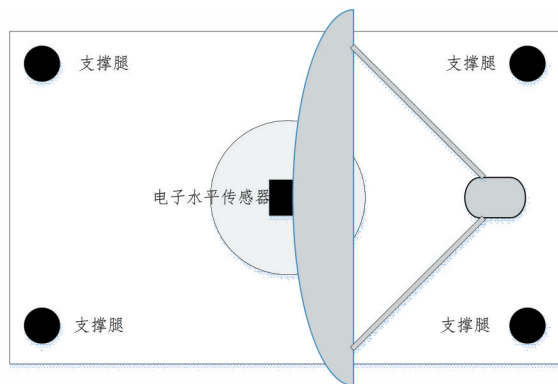


图 1 自动调平系统

收稿日期: 2023-02-21 修稿日期: 2023-03-05

作者简介: *通信作者: 薛缠明(1988—), 男, 山西吕梁人, 硕士, 工程师, 研究方向为无线电, E-mail: 1036853231@qq.com; 赵翔宇(1983—), 男, 山西右玉人, 硕士, 工程师, 研究方向为无线电

2 BP神经网络算法预测

目前在数据预测中BP神经网络算法凭借它非线性映射能力、泛化能力和时效性强等特点,已经成为数据预测的重要工具。BP神经网络是由多个神经元组成的广泛互连的神经网络模型,并通过数据样本对神经网络模型的训练,使其具有完成各种数据处理和非线性转化的能力。对于任意一组随机的、正态的数据,都可以利用BP神经网络进行统计分析,做出拟合和预测。在利用BP神经网络算法对数据进行预测的过程中,为提高预测数据的精度,模型的建立和训练成为关键环节。为此,提出一种以BP神经网络算法为基础预测补偿误差的方法,其可分为三个步骤。

(1) BP神经网络建模。BP神经网络的模型结构主要由输入层、隐藏层和输出层组成,层之间采用全连接方式,层内无连接^[3]。这样的神经网络模型通过反向传播算法来求解,中间隐藏层的主要作用是更好地对数据进行表示以及提高函数的拟合能力。

(2) 模型训练。网络模型的训练一般采用反向传播法,在反向传播过程中主要是采用梯度下降公式进行有效的反向迭代,根据预测值与实际值的偏差量,改变参数矩阵,以达到更好的训练效果。

(3) 补偿误差数据预测。BP神经网络模型建立和训练后,对实时补偿误差数据进行预测。通过实际验证,预测的补偿误差可以有效地提高测量精度。

2.1 BP神经网络建模

在Ka频段车载测控设备实时跟踪目标过程中,由于时效性强,要求电子水平传感器自动调平系统建立的BP神经网络模型能够快速预测出补偿误差,确保测量的精度。通常BP神经网络的模型如图2所示,由输入层、隐藏层和输出层组成。

其中隐藏层各神经元的BP结构如图3所示,其基本的单元公式为

$$o = g \cdot \left(\sum_{i=0}^n (w_i \cdot X + b_i) \right) \quad (1)$$

式中: x 为输入值, g 为激活函数, w 为权重, b

为偏置, n 为隐藏神经元个数, o 为输出值,引入偏置 b 的原因在于让其模型适用于更多情况。各层之间的工作流程为:从输入端开始,沿着箭头指向乘以相关权重系数后加和(还需加偏置),加和后得到的结果为激活函数的输入,再次计算得到的结果作为此节点的输入,依次计算,直到得到最终的输出。

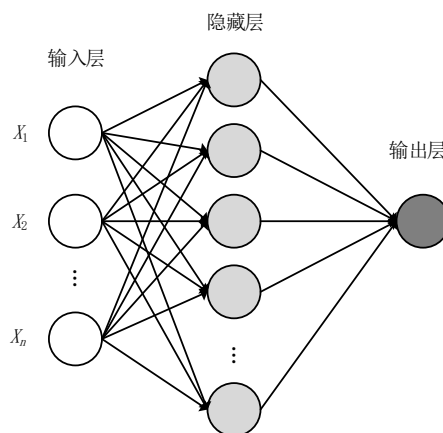


图2 BP神经网络模型

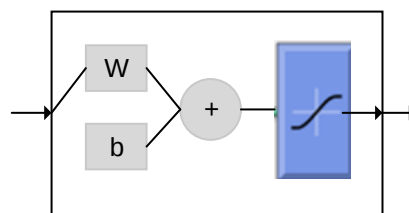


图3 隐藏层各神经元BP结构

模型建立时,增加隐藏层可以更好地对数据进行表示和提高函数的拟合能力,在三层的基础上,再引入更多的隐藏层,就变为深度神经网络。但是每增加一层,模型的参数数量急剧增加,对计算机资源的要求增高,在实际使用中,模型训练的时间非常久。针对电子水平传感器自动调平系统神经网络模型的建立,考虑到实时性以及计算机资源的限制,所以本系统选择只有一层隐藏层的BP神经网络模型。

2.2 模型训练

BP神经网络模型训练的过程,包括正向传播和反向传播^[4]。其中正向传播是把样本从输入层输入,经过隐藏层处理后,从输出层输出。

这样得到的实际值没有达到期望值后,就可以采用反向传播。反向传播是将预测值与实际值的偏差量以某种形式通过隐藏层向输入层逐层反传,并将误差分摊给所有神经元,获得各层的误差信号,用它们可以对各层的神经元的权值进行调整,循环不断地利用输入输出样本集进行权值调整,以使所有输入样本的输出误差都减小到满意的精度。

在反向传播中核心问题是如何度量预测值与实际值的误差,传统的度量方法是采用预测值与实际值的均方误差。本系统在度量中采用梯度下降方法^[5],对每个权重做偏导,乘以学习率 η ,公式为

$$\Delta\omega = -\eta \frac{\partial \left(\frac{1}{m} \sum_{i=1}^m (y_{\text{pre}} - y_{\text{tru}})^2 \right)}{\omega_i} \quad (2)$$

式中: y_{pre} 为预测值, y_{tru} 为实际值, η 为学习率。

2.3 补偿误差数据预测

为验证模型,采用某Ka频段车载测控设备跟踪无人机定点调平的数据集进行验证。数据集采集了电子水平传感器在设备模拟跟踪无人机不同运行轨迹下的几组补偿误差数据,每组采集了12个定点的数据,通过多次模拟跟踪定点数据采集,得到数据样本,截取一种运动轨迹下部分样本数据如表1所示。

表1 电子水平传感器定点采集的原始数据

S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
-0.702	-0.700	-0.697	-0.698	-0.696	-0.698	-0.700	-0.703	-0.704	-0.706	-0.707	-0.706
-0.701	-0.698	-0.696	-0.697	-0.697	-0.698	-0.700	-0.703	-0.705	-0.705	-0.706	-0.705
-0.700	-0.698	-0.696	-0.698	-0.698	-0.700	-0.702	-0.703	-0.704	-0.704	-0.704	-0.703
-0.699	-0.697	-0.698	-0.700	-0.700	-0.701	-0.704	-0.703	-0.703	-0.703	-0.702	-0.701
-0.701	-0.698	-0.697	-0.698	-0.697	-0.698	-0.701	-0.703	-0.704	-0.705	-0.705	-0.703

将上述不同定点采集的数据作为训练数据,通过模型学习训练后,得到预测的补偿误差和真实补偿误差情况如图4所示。

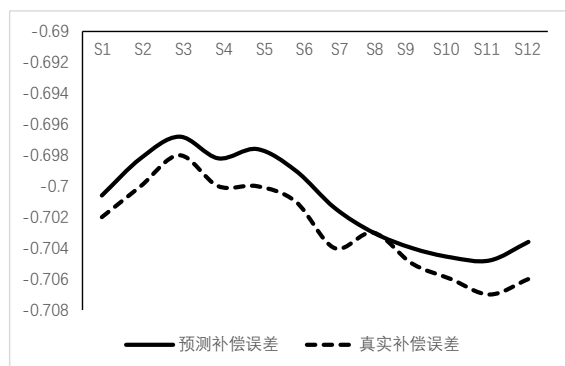


图4 补偿误差对比情况

最后通过训练后的模型模拟跟踪无人机,实时将电子水平传感器反馈的信号作为输入,进行动态预测补偿误差比较。通过动态补偿误差测量的俯仰角数据比跟踪前补偿固定误差测量的俯仰角数据精度更接近于真实值,极大地提高了设备的精度。

3 结语

本文将BP神经网络算法应用到Ka频段车载测控设备自动调平系统中,经过实际试验使用,对天线跟踪过程中重心变化导致的补偿误差实时预测补偿。结果表明,BP神经网络算法可以高效精确地预测补偿误差,提高了设备的测量精度。

参考文献:

- [1] 陈万军.全自动调平控制系统设计及调平策略研究[D].河北:河北科技大学,2015.
- [2] 刘杰,黄亚楼.基于BP神经网络的非线性网络流量预测[J].计算机应用,2007,7:1770-1772.
- [3] 黄丽.BP神经网络算法改进及应用[D].重庆:重庆师范大学,2012.
- [4] 李萍,曾令可,税安泽,等.基于MATLAB的BP神经网络预测系统的设计[J].计算机应用与软件,2008,4:149-184.
- [5] 刘颖超,张纪元.梯度下降法[J].南京理工大学学报(自然科学版),1993,2:98-101.

(下转第83页)

开发案例

文章编号: 1007-1423(2023)11-0071-05

DOI: 10.3969/j.issn.1007-1423.2023.11.015

研究生业务费预算与支出管理系统设计与实现

张建宇, 徐怡彤, 张盈盈, 吕震宇*

(华北理工大学管理学院, 唐山 063210)

摘要: 针对研究生业务费管理工作中的公共费用管理、联合预算报销、实际报销金额与预算不一致等问题, 确保在不修改现有财务系统的前提下, 实现与预算支出管理的无缝对接, 设计并实现了本研究生业务费预算与支出管理系统。导师结组功能解决了校外导师无法校内报销的问题, 实现了导师间预算打通使用; 多人联合预算功能允许对票面金额较大的发票联合多人预算予以报销; 基于二维码的预算来源单管理一方面确保了单据防伪和不可篡改, 另一方面实现了和现有财务系统的无侵入式对接; 研究生业务费自动分配功能减轻了管理人员的工作量, 降低了出错率。

关键词: 研究生业务费; 信息系统; 系统分析与设计

0 引言

研究生业务费是指学校为研究生培养工作提供的经费支持, 直接用于研究生开展科学研究与实践、学术交流、学位论文开题报告、中期检查及评审答辩等工作。《教育部 财政部关于普通高等学校“研究生业务费”开支问题的试行规定》(教计财字[1982]029号)对研究生业务费开支范围、开支标准等问题做了明确规定。《教育部 国家发展改革委 财政部关于加快新时代研究生教育改革发展的意见》(教研[2020]9号)指出, 要切实做好研究生经费保障, 完善研究生教育投入体系, 研究建立差异化生均拨款机制。

各高校分别结合各自研究生培养实际, 制定了相应的研究生业务费管理办法。不同高校对研究生业务费的管理方式各不相同, 通常都是按照一定计算原则将研究生业务费划拨至研究生, 但具体使用由导师统筹安排。随着研究生人数的增加, 研究生业务费在实际使用中面

临的最大问题就是如何对资金额度进行有效监管, 避免学生超额度报销或无法足额按期报销。

1 研究生业务费管理现状

传统研究生业务费管理多采用手工记账方式, 学校为每一位研究生发放一经费账本, 每次下拨经费或经费报销均需到账本上登记并由导师和主管领导签字。这种手工记账的管理模式难以对研究生业务费公共支出(例如查重、论文打印装订等由相关专业统一办理的业务)部分进行有效管理; 此外, 可能出现在账本上登记了支出, 但实际报销金额与账本登记不一致的情况; 再有, 手工记账方式难以监管资金支出进度, 不利于研究生业务费的统筹使用。

很早就有高校采用信息化手段对研究生业务费进行管理, 但管理模式主要围绕财务报销流程展开, 对预算的管理相对较弱。马静等^[1]通过设计基于 Web 三层架构的科研经费精细化管理系统解决了管理部门对课题预算无法有效管理、审查和科研人员对预算实时动态难以查

收稿日期: 2023-02-25 修稿日期: 2023-03-21

基金项目: 华北理工大学校级大学生创新创业训练计划项目(X2022208)

作者简介: 张建宇(2001—), 男, 河北沧州人, 在读本科, 研究方向为计算机软件、信息管理; 徐怡彤(2002—), 女, 河北石家庄人, 在读本科, 研究方向为计算机软件、信息管理; 张盈盈(2002—), 女, 湖北荆州人, 在读本科, 研究方向为计算机软件、信息管理; *通信作者: 吕震宇(1976—), 男, 山东安丘人, 硕士, 教授, 研究方向为信息管理与信息系统、数据分析与挖掘, E-mail: zhenyulu@163.com

看的问题。刘海峰等^[2]设计的科研业务费项目管理信息系统解决了手工录入信息出错、信息孤岛和工作效率低的问题。刘沐等^[3]通过对现有研究生业务费管理系统的分析解决了科研管理与财务管理脱节、项目立项后实施过程缺乏监管、项目经费入账手续繁杂、预算编制与实际经费使用不符、统计决策无数据支持的问题。

上述系统设计均以研究生业务费报销管理为核心,在预算分配和利用方面考虑相对较少。特别是在多人联合预算报销、公共预算分配与支出、预计报销金额与实际报销金额不一致等问题上没有给出有效解决方案。此外,目前高校早已实施了信息化财务报销管理,如何在入侵入现有财务系统的同时,有效实现对研究生业务费使用的监管是摆在面前的重要问题。

2 系统设计

2.1 研究生业务费预算与支出管理方案设计

针对上述情况,所设计的研究生业务费预算与支出管理系统必须能与现有财务报销系统无缝对接,必须不影响现有财务报销系统的运行,另外还要能够有效解决公共支出分配、多人联合报销、实际报销金额与预算登记支出不一致等问题。为此设计如下管理方案:

2.1.1 与现有财务报销系统对接方案设计

为了不影响现有财务报销流程,本《研究生业务费预算与支出管理系统》只负责研究生业务费的预算分配与支出管理。在报销研究生业务费时,报销人除向财务处提供相关发票及佐证材料外,还需要提供一张由本系统出具的《研究生业务费预算来源单》,该单据上有一个由系统生成的,可用于校验预算来源真伪的二维码,财务处在报销时(或主管领导签字时)通过扫码确认该笔报销是否有足够的研究生业务费预算来源做支撑,以此判断是否予以报销。同时,该二维码一旦被扫描确认,《研究生业务费预算与支出管理系统》将把这笔经费记为已支出,以此实现与现有财务系统的无侵入式对接。

2.1.2 公共研究生业务费分配方案设计

财务部门在将总研究生业务费下达至学校

或二级学院后,该笔经费将首先登记至系统。研究生秘书可先将当年需要支出的公共经费(含查重费、外审费、打印装订费、答辩费等)单独划拨出来纳入公共支出预算,剩余部分再按照指定比例分配至导师或具体研究生,作为个人预算用于论文发表、实验材料采购等支出。

2.1.3 多人联合报销方案设计

报销研究生业务费时,若发票金额大于研究生本人剩余预算额度时,可采取以下三种方式予以处理:①到财务处按照本人剩余预算额度进行报销,超出预算额度的发票金额将不予报销;②剩余预算额度内的费用从本人研究生业务费中报销,其余部分从本人或导师的科研费中报销;③从系统中联合多人的预算生成与发票金额等额的《研究生业务费预算来源单》,并依此到财务处按发票金额足额报销。报销出来的费用按照联合预算的构成分给其他研究生或导师。

2.1.4 实际报销与预算登记支出不一致问题的解决

若实际报销金额与预算登记支出不一致(通常是报销金额小于预算登记金额),研究生或导师可凭有效证明材料找主管领导,从系统中将未能报销的部分从“已支出预算”部分转移至“未支出预算”,以确保用户下次还能使用该预算额度。系统将记录所有操作过程,以备随时调阅查看。

2.2 用例与核心业务活动图分析

2.2.1 用例图

《研究生业务费预算与支出管理系统》以二级学院研究生业务费管理为原型,用例图如图1所示。系统的主要参与者包括管理员、院长、科研副院长、研究生秘书与研究生导师。管理员主要负责基础信息维护与系统管理工作;院长负责年度研究生业务经费来源登记、报销前预算审核以及未报销金额登记等;科研副院长可负责查询统计所有报销预算单和未报销登记记录等;研究生秘书主要负责研究生业务费分配(包括公共经费和研究生经费分配)、导师分

组等；导师主要通过系统生成报销预算来源单。

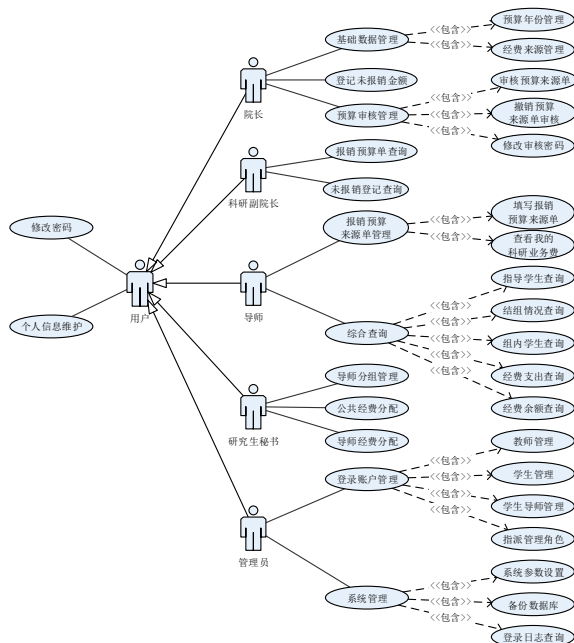


图1 研究生业务费管理系统用例

2.2.2 核心业务活动图

《研究生业务费预算与支出管理系统》核心业务分为研究生业务分配和研究生业务费预算与来源单的创建及报销两部分，如图2所示。

研究生业务费预算分配前，首先由院长在系统登记经费来源，随后研究生秘书在此基础上进行查重费、外审费、打印装订费、答辩费等公共经费的分配，剩余部分根据预先设定的规则，有系统自动根据导师指导的学生人数以及所在年级自动分配至导师个人账户。经费分配完成后由院长审核确认。

研究生报销业务费时，首先委托导师通过系统创建预算来源单。预算来源单创建支持以下场景：①导师结组报销。结组后，同组导师间研究生业务费预算可打通使用。②预算额度内报销。若某研究生报销时，导师名下预算额度小于票面金额，可只报销预算额度内金额。③联合预算报销。联合报销就是在经其他导师同意的情况下借用其他导师预算额度，确保票面金额与联合预算额度一致，全额按票面金额报销。

报销前，研究生将发票、相关附件、系统生成的研究生业务费预算来源单一同提交给主管院长审核。预算来源单中除了记录报销人、预算来源以及金额外还会生成一个唯一的二维码。用户对预算来源单的任何修改动作都会导致该二维码发生变化。院长在审核签字前首先通过手机扫描预算来源单上的二维码，系统自动比对纸质预算单二维码与系统现有登记的数据是否一致，如果一致且预算来源单的金额小于等于发票金额时，院长在手机上输入预算来源单审核密码并确认，该预算来源单即从后台锁定并标记为已支出，报销单院长签字后即可到财务处报销（线上报销流程类似，不再赘述）。

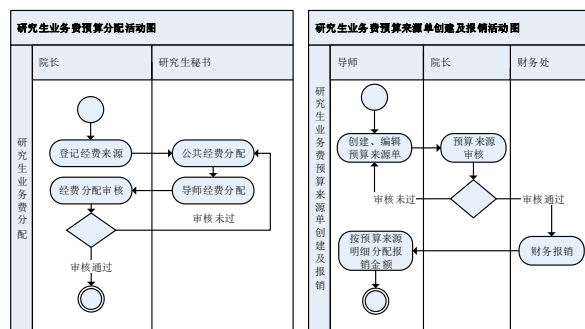


图2 研究生业务费管理系统核心业务活动

2.3 功能模块设计

研究生业务费预算与支出管理系统主要功能模块包括：登录账户管理、基础数据管理、综合查询管理、业务费报销管理等功能模块如图3所示。

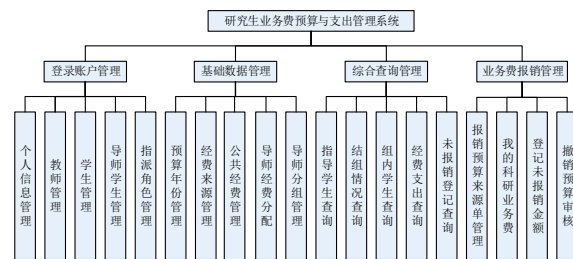


图3 研究生业务费管理系统功能模块

2.4 数据库设计

系统数据库概念模型设计如图4所示。

如果该预算来源单状态为未审核,对钩颜色为橙色,且下方出现审核确认按钮。在正确输入了审核确认密码后可对该预算单予以报销确认。审核后的预算单再次扫码时将呈现绿色对钩,且不再出现审核确认按钮。如图7所示。



图7 预算来源单扫码后的三种审核状态

3.2.3 研究生业务费自动分配

系统会根据待分配研究生业务费总额度、导师所带的研究生人数与年级信息实现业务费的自动分配。系统允许设置研一至研三各年级金额分配系数,然后系统会自动计算各导师分配的研究生业务费预算额度。如图8所示。

学段	金额	分配系数	分配金额	分配比例	分配比例	分配比例
研一	10000	1.0	10000	100%	100%	100%
研二	10000	1.0	10000	100%	100%	100%
研三	10000	1.0	10000	100%	100%	100%

图8 系统自动按照指导的研究生人数分配研究生业务费

4 结语

以研究生业务费预算管理为核心的《研究生业务费预算与支出管理系统》设计有效解决了研究生业务费公共预算分配与支出、导师(研究生)业务费自动分配、多人联合预算报销、预计报销金额与实际报销金额不一致等研究生业务费管理问题;基于二维码的预算来源单设计,一方面确保了预算来源单的防伪和不可篡改,另外一方面实现了和现有财务系统无侵入的无缝对接,保证了该信息系统的顺利实施,有效提高了研究生业务费管理工作效率和预算管理水平。未来系统将在导师动态结组、小额预算余款管理、研究生业务费支出进度管理等方面做进一步优化改进。

参考文献:

- [1] 马静,李永革.基于Web三层架构的科研经费精细化管理系统设计[J].中国管理信息化,2016,19(17):93-97.
- [2] 刘海峰,刘春,廖冠琳,等.中央高校基本科研业务费项目管理信息系统的设计与实现[J].中国教育信息化,2020(5):57-60.
- [3] 刘沐,谢素萍,皮新玲,等.从科研项目管理现状谈高校项目管理系统设计[J].中国管理信息化,2011,14(2):47-50.

Design and implementation of budget and expense system for graduate research-related subsidy management

Zhang Jianyu, Xu Yitong, Zhang Yingying, Lyu Zhenyu*

(School of Management, North China University of Science and Technology, Tangshan 063210, China)

Abstract: In graduate research-related subsidy management, public spending, joint budget reimbursement, and handling inconsistency between actual reimbursement and budget source are critical for business process, and it is also necessary to collaborate with the existing financial system without modifying it. In this graduate research-related subsidy budget and expense system, tutor grouping solves the problem that external tutors cannot reimburse in school; joint budgeting allows tutors joint their budgets to reimburse larger invoices; QR code based budget source sheet ensures the documents immutable and anti-counterfeiting, and can collaborate with current financial system seamlessly; automatic subsidy allocation reduces management workload and lower the error rate.

Keywords: postgraduate business fees; information system; system analysis and design

文章编号: 1007-1423(2023)11-0076-08

DOI: 10.3969/j.issn.1007-1423.2023.11.016

基于 SpringBoot 的物联网智能家居方案设计

陈 欣, 阮锦新*, 邬依林, 张 谦, 陈勇昌, 熊 燕

(广东第二师范学院计算机学院, 广州 510303)

摘要: 随着现代信息技术的不断发展和突破, 智能家居逐渐走进我们的日常生活。针对用户对家居设备的个性化智能控制需求, 提出采用传感器、无线通信、云端服务等技术手段, 基于“单片机+Web 服务”的系统架构设计出具有环境数据采集、云端远程智能控制及可视化数据展示的智能家居解决方案。方案可同时接入 6 个家居设备并进行智能控制, 其最大的特点是用户通过云端配置传感器的阈值实现对设备的智能控制。通过功能验证, 方案可运用于实际生活中, 其设计思路和系统架构具有一定的参考价值。

关键词: 智能家居; STC89C52 单片机; 传感器; 物联网; SpringBoot

0 引言

快速发展的信息技术革命正在推动着不同行业的变革, 尤其是以物联网为代表的技术正快速渗透到工业生产和产品生命周期中, 促进了工业智能化和数字化^[1]。物联网本质上就是让各种物品相互连接形成一个庞大的网络, 通过传感器与互联网提供通信, 最终实现万物互联^[2-3]。智能家居是物联网的典型应用, 通过使用多种传感器对家居环境数据进行采集, 并根据这些数据对家居设备做出相应的定制化控制, 在这个过程中用户完全不需要对家居设备进行干预, 而且智能家居系统通常具有 24 小时不间断工作的优点^[4]。

普通的家居设备之所以可以变得人性化和智能化, 是因为其背后依靠强大的物联网进行

通信和数据交换。传感器就像计算机的眼睛, 不仅可以有效检测出相应的信息, 同时还能够将检测到的信息以其他信号形式对外输出, 这就意味着传感器可以帮助我们准确、快速地获得信息, 甚至有些传感器的灵敏程度远超人的感官^[5]。为了让智能家居系统更具定制化和智能化, 可通过使用多种传感器以及后端智能控制策略, 让家居设备信息上传到云端, 实现基于云端的家居设备智能控制^[6]。本文针对用户对家居设备智能控制的需求, 选用 STC89C52 单片机作为硬件主控, 采用 SpringBoot 后端框架开发云端, 设计出贴近生活的智能家居系统。系统利用弱电控强电, 具有高度灵活的设备绑定功能, 可以让传感器作用于任意预留的设备控制端口并实现对相应家居设备的控制, 传感器采集的数据还可以实时上传云端并进行可视化展示。

收稿日期: 2023-04-12 修稿日期: 2023-04-25

基金项目: 广东省自然科学基金(2022A1515010485); 广东省高等学校教学质量与教学改革工程项目(2019 年, No.18); 广东省 2020 年高等教育教学改革项目(2020 年, No.440); 广东第二师范学院高等教育教学改革项目(2021jxgg07); 广东第二师范学院重点学科项目(ZD2017004)

作者简介: 陈欣(1999—), 男, 广东汕头人, 本科, 研究方向为嵌入式软件开发、网络设计; *通信作者: 阮锦新(1971—), 男, 广东广州人, 博士, 讲师, 研究方向为无线通信、模式识别、图像处理, E-mail: ruanjinxin@163.com; 邬依林(1970—), 男, 四川仪陇人, 博士, 教授, 研究方向为复杂系统建模、图像处理、网络控制、智能控制、教育信息化; 张谦(1979—), 男, 广西桂林人, 硕士, 副教授, 研究方向为信息化教学; 陈勇昌(1983—), 男, 广东广州人, 博士, 讲师, 研究方向为信息安全、图像处理; 熊燕(1976—), 女, 江西南昌人, 硕士, 讲师, 研究方向为无线通信、信号处理

1 智能家居系统设计

智能家居系统设计包括硬件选型、硬件电路设计以及软件设计。硬件选型是根据应用场景的实际需求选择相应的主控芯片、传感器和通信模块。硬件电路设计是对这些模块进行引脚定义和连接，使其构成一个完整的电路系统。软件设计包括单片机控制系统设计和云端系统设计。

1.1 硬件选型

主控芯片采用通用性高、性能稳定、价格低廉的 STC89C52 单片机^[7-9]，同时选用温湿度传感器、光敏传感器、人体红外射频传感器、超声波传感器，选用无线通信模块实现单片机之间的通信。主要使用的器件如表 1 所示。

表 1 主要使用的器件

分类	模块	优点
主控芯片	STC89C52 系列单片机	功耗低、性能高、抗干扰能力超强，可在线编程，完全兼容传统 8051 的指令代码
	DHT11 温湿度传感器	良好的灵敏感应特性、防水性、稳定性、高精度、低漂移
传感器	HC-SR04 超声波传感器	性能稳定，测度距离精确，模块高精度，盲区小
	HC-SR501 人体红外感应模块	全自动感应、灵敏度高、可靠性强，人进入其感应范围则输出高电平、人离开感应范围自动延时关闭高电平，输出低电平
	可调光敏模块	可通过调节电阻值改变光敏电阻的灵敏度
	NRF24L01 无线通信模块	内置功率放大器、频率合成器、晶体振荡器、调制器等模块，可通过程序对输出功率和通信频道进行配置

1.2 硬件电路设计

硬件电路设计基于单片机最小系统板，通过传感器、继电器模块以及家居设备的连接，组成完整的电路控制系统。控制系统的核心控制部件为 STC89C52 系列单片机，通过使用最小系统板引出单片机的所有引脚，最小系统板的电路原理如图 1 所示。

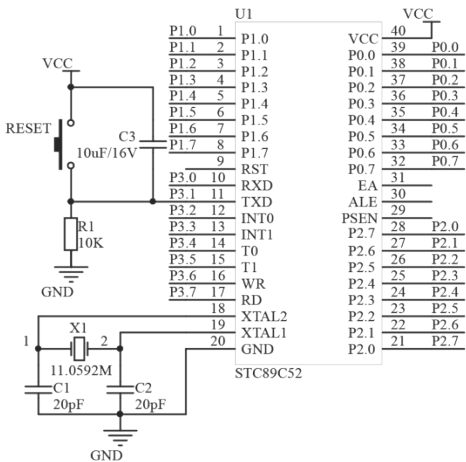


图 1 最小系统板原理

温湿度传感器模块、人体红外感应模块及可调光敏模块均为三引脚，其中两个为正负极，一个为数据端，数据端分别连接在单片机的 P2.0、P2.4、P2.6 引脚。超声波传感器模块共四个引脚，控制引脚分别为 Trig 触发引脚和 Echo 回响引脚，分别接在单片机的 P2.1 和 P2.2 引脚。通过对单片机电源模块的正负极进行拓展，将这些传感器模块的正负极接入到单片机的正负极中，实现模块供电。无线通信模块为八引脚，供电电压为 3.3V，因此需要额外的供电电路进行供电，剩余的六个引脚为 IRQ、MIS、MSO、SCK、CS、RST，分别接在单片机的 P3.2、P3.3、P3.4、P3.5、P3.6、P3.7 引脚。上述所有模块的接线原理如图 2 所示。

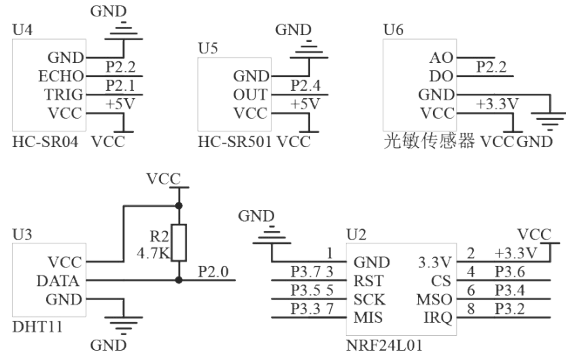


图 2 传感器接线原理

使用了 6 个光耦继电器模块实现弱电控制强电，这些模块的信号脚接在单片机的 P1.2~P1.7 引脚，继电器模块组接线原理如图 3 所示。

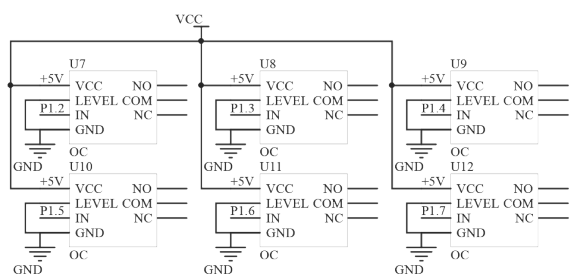


图3 继电器模块接线原理

1.3 软件设计

单片机控制系统主要实现传感器数据采集、指令识别以及端口控制和控制反馈。云端系统实现人机交互，为用户提供可视化的家居设备控制功能，同时将单片机采集到的数据入库并进行数据展示。

1.3.1 单片机控制系统设计

单片机控制系统设计包括时钟程序、端口控制程序、单总线通信、无线通信和串口通信等程序设计。

(1) 时钟程序设计。单片机的时钟通过设置定时器T0溢出值来触发CPU中断，利用中断进行延时计数并以此划分出一条时间轴作为单片机运行的基准时间，其余的程序都会在这个设计好的“节拍”中有序运行。

(2) 端口控制程序设计。根据用户的设定或控制改变单片机引脚的输出电平，所有对电平的控制都是通过这个程序进行，任何操作最终都将转换为“端口号+控制码”的形式进行相应的位运算并做出最终的电平控制操作。

(3) 单总线通信设计。温湿度传感器、人体红外传感器均采用这种通信模式，分别使用一根数据线完成数据的交换和控制，通过系统时间轴有序触发单次数据读取操作。温湿度模块读取流程和超声波距离传感器测量距离流程分别如图4、图5所示。

(4) 无线通信程序设计。在通信上采用“单片机指令分发器+单片机设备控制器”的架构，单片机指令分发器对控制指令进行中转，真正实现控制的是单片机设备控制器，而两个单片机之间则是采用NRF24L01无线模块进行数据传

输。无线模块在收到数据后改变状态寄存器的电平，单片机通过轮询检测寄存器状态变化后调用无线通讯解析函数，对收到的消息进行格式化、解析和重新封装等处理。

(5) 串口通信程序设计。单片机服务器与TCP服务器通过串口通信方式进行数据传输，设置定时器1产生19200的波特率用于通信，后端服务将用户在云端的操作指令通过TCP客户端发送给TCP服务端，TCP服务端以串口通信方式将指令传送给单片机服务器，进而转发给控制器实现控制。

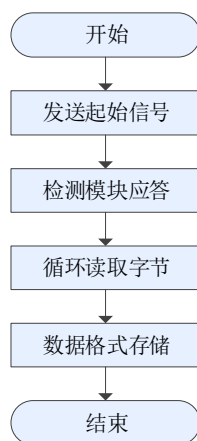


图4 温湿度模块读取数据流程

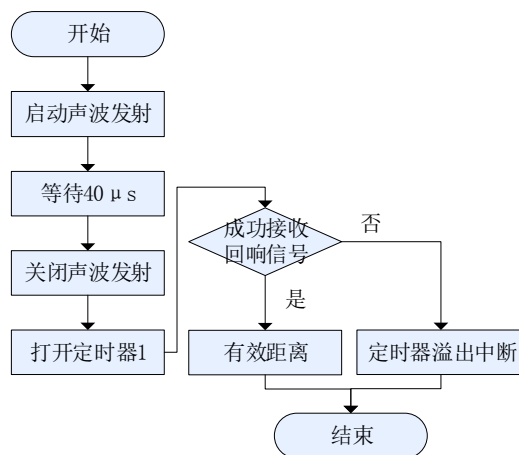


图5 超声波距离传感器测量距离流程

1.3.2 云端系统设计

云端系统设计包括TCP服务、JavaWeb服务端程序以及数据库等设计。

(1) TCP服务。采用Python开发TCP服务

端，TCP 服务端以串口通信方式连接单片机服务器，并定时发送温湿度查询指令及传感器状态查询指令获取返回数据，数据经过处理后进行入库操作，用户对家居设备的控制也通过这程序进行转发。

(2) JavaWeb 服务。采用目前主流的 Spring-Boot 框架^[10]搭建交互式的用户访问平台，平台具有设备映射、设备控制、控制策略设置等功能，支持用户在平台上对端口进行绑定和控制，同时还可查看传感器相关的可视化数据。其中，设备控制的核心是在 SpringBoot 中注入一个 TCP 客户端对象，用户的控制指令通过 TCP 客户端发送给 TCP 服务端并进行后续的控制和反馈。

(3) 数据库设计。数据库设计包括端口映射表、端口状态日志表、传感器采集日志表、温湿度采集日志表、自动控制策略智能硬件表和自动控制策略服务下发表的设计。其中，最为核心的是端口映射表，记录了设备与端口的绑定信息，也是远程控制的关键。传感器数据的更新是通过 TCP 服务定时向控制器发出查询请求来获得数据并按照时间顺序入库，这些数据可用于前端展示，并且用户可根据这些数据对家居设备进行准确可靠的控制。

2 智能家居场景模型搭建

本文设计的智能家居解决方案可以提供 6 个控制端口给用户接入家居设备，为了验证方案的可行性，对智能房间进行模拟。

2.1 家居设备布局

智能房间有卧室、阳台和卫生间。在房间内，配置 4 个照明设备以及 1 个加湿器和 1 个换气装置。其中照明设备分别安装在卧室、卧室书柜、卫生间和阳台，加湿器安装在卧室，换气装置安装在卫生间，具体的家居设备布局如图 6 所示。

2.2 传感器规划

智能房间安装了温湿度传感器、人体红外感应器、超声波距离传感器、可调光敏模块，分别安装在卧室、卧室衣柜、卫生间和阳台，传感器功能与布局如图 7 所示。

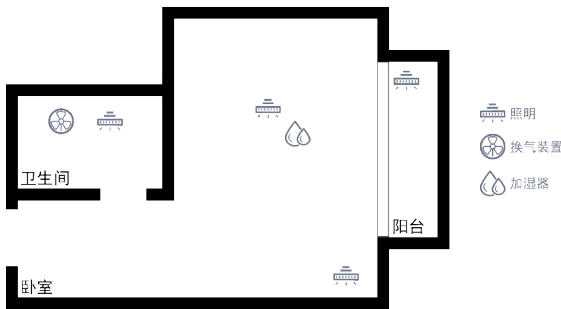


图 6 家居设备布局图

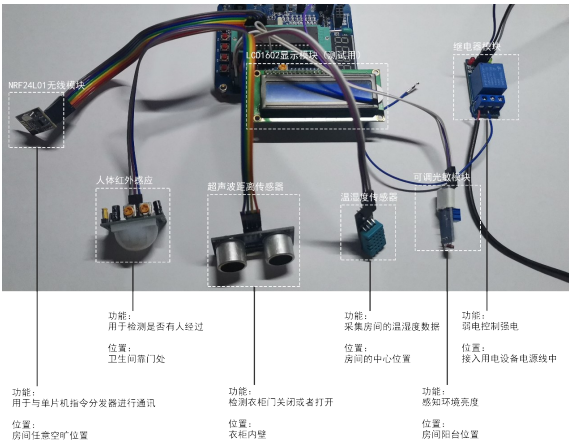


图 7 房间传感器规划图

2.3 云端配置

硬件规划完成后，需要在云端进行设备映射绑定，登录到系统后通过设备映射功能，录入对应的设备到系统中即可完成绑定，新增绑定设备如图 8 所示，所有家居设备绑定完成之后的设备列表如图 9 所示。

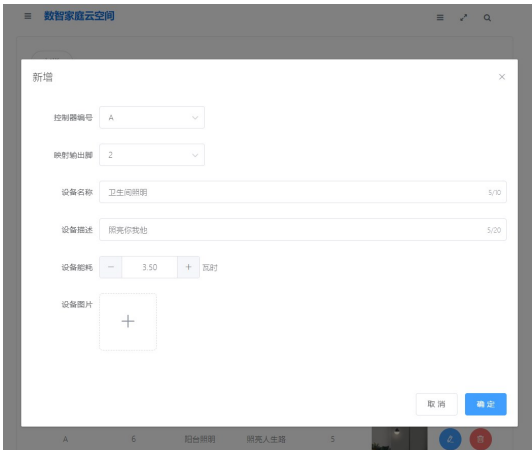


图 8 新增绑定设备

控制图编号	映射输出群	设备名称	设备描述	设备能耗	设备图片	操作
A	2	卫生间照明	照亮你旅途	3.5		
A	3	衣柜照明	照亮你世界	3.5		
A	4	换气装置	通风更便捷	15		
A	5	加湿器	干净又卫生	350		
A	6	阳台照明	照亮人生路	5		
A	7	景观照明	照亮你视线	110		

图 9 所有家居设备绑定后的设备列表

在“智能房间”场景中，人体红外感应模块用来控制卫生间的照明设备，当有人经过时，照明设备会自动打开。当一段时间没有感应到人经过时，照明设备会自动关闭，这些功能可在系统上进行设置。对于阳台的照明设备，则是利用可调光敏感应器控制，亮度低于一定值的时候会自动打开阳台照明设备。温湿度感应器则是用来收集房间的温湿度信息并在系统中展示，用户也可通过系统，利用温湿度传感器的数据来自动控制房间中的家居设备，例如控制加湿器、换气装置等。

2.4 家居设备的接入

家居设备接入系统需要通过继电器模块。下面是照明设备接入系统的过程：首先继电器模块电源接入+5V，接着将单片机控制信号线接入到继电器的信号接收端IN，同时在继电器上将跳线帽接入到低电平触发引脚，最后按照图10接线图将照明设备接入。

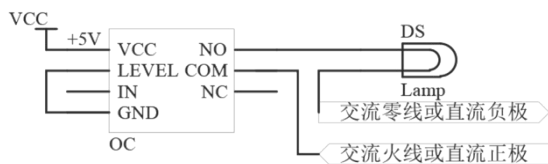


图 10 继电器模块用电器接线图

3 系统测试

系统测试包括硬件测试、软件测试和功能测试。

3.1 硬件测试

硬件测试包括单片机外围电路连接检查，单片机复位后控制端口输出电平以及温湿度传感器等模块的采集情况。复位后采集的温湿度可以正常显示在LCD屏幕上，同时控制口输出电平为高电平，P1.2~P1.7的板载LED灯熄灭，测试结果如图11所示。

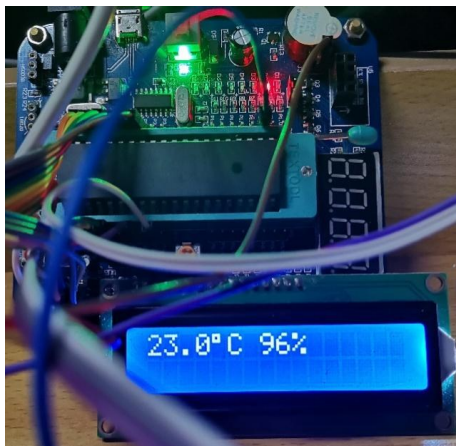


图 11 单片机复位

3.2 软件测试

软件测试包括单片机控制系统测试和云端系统测试。

单片机控制系统实现收集数据、接收指令、控制端口和控制反馈，因此最关键的是查看串口反馈的数据是否符合预期。在单片机控制系统测试中，将程序写入单片机后，使用串口调试助手按照192000的波特率连接单片机并发送相关指令，查看返回的数据，如果可以正常返回数据同时符合预期，则证明功能正常。在串口调试助手中向串口发送指令“{AD}”，对链路进行简单测试，测试结果如图12所示，能正常接收到串口回传的温湿度数据，说明串口通信和单片机之间无线通信链路正常。

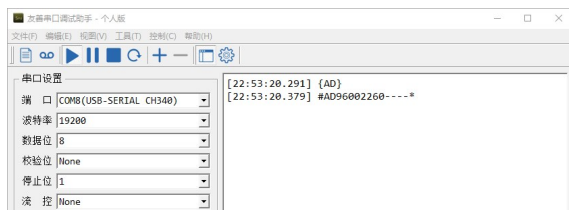


图 12 串口通信及单片机之间无线通信链路测试

在云端系统测试中，首先测试TCP服务是否可以正常连接和通信，在TCP客户端测试程序中向TCP服务端发送测试数据可以收到反馈，功能正常，测试结果如图13所示。接着对用户交互界面进行功能测试，确保可以从数据库中正常获取到相应的数据，这里以首页为例，温湿度及传感器状态能正常获取并显示，测试结果如图14所示。

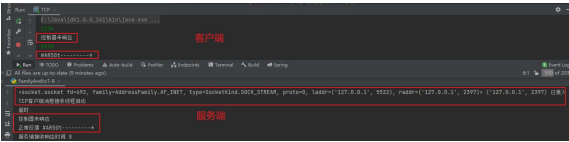


图 13 TCP连接及通信



图 14 首页温湿度及其他传感器状态

3.3 功能测试

下面以照明灯控制为例对系统的硬件和软件进行整体测试。在照明灯关闭的状态下，点击开关，照明灯会被打开，同时云端页面会收到打开的反馈提示，如图15和图16所示。再次点击开关，照明灯会被关闭，同时页面会收到关闭的反馈提示，如图17和图18所示。



图 15 打开照明灯实物图



图 16 打开照明灯系统反馈

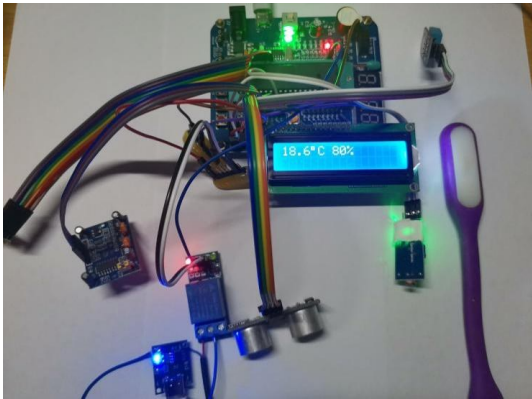


图 17 关闭照明灯实物图



图 18 关闭照明灯系统反馈

在自动控制策略测试中，以智能控制卫生间照明灯为例进行测试。首先将可调亮度模块

绑定到卫生间照明灯上,同时设置临界控制及边缘值,具体配置如图 19 所示。



图 19 智能亮度传感器配置

配置完成后,打开使能开关,对环境亮度进行调节,分别模拟白天和黑夜,测试效果如图 20、图 21 所示。

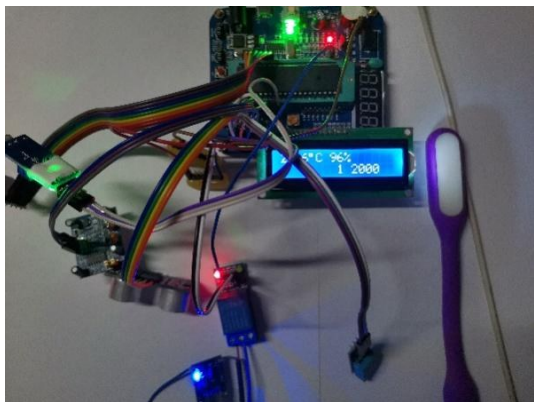


图 20 模拟白天(照明灯熄灭)

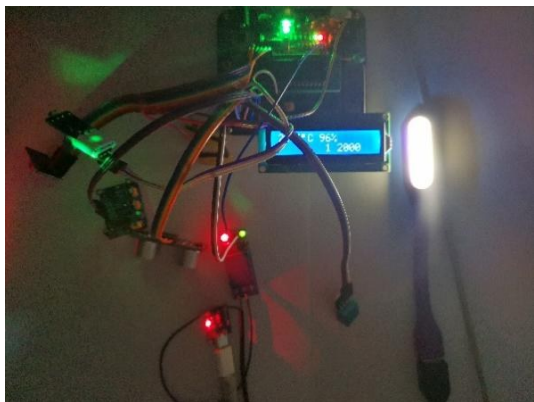


图 21 模拟黑夜(照明灯自动点亮)

4 结语

随着物联网技术的不断发展,智能家居技术给我们生活带来越来越多的便利。本文提出的智能家居方案通过将单片机与 JavaWeb 云端进行有机结合,具有操作便捷、数据直观等特点,为用户提供了在线和离线两种控制方式,用户可通过云端系统自定义规则实现灵活的家居设备自动控制,大大提升了智能家居的使用体验。

参考文献:

- [1] 钱肇钧,杨森,李伟. 工业互联网概念研究及频率规划研究建议[J]. 中国无线电,2016(4):40-43.
- [2] OSCAR N. Blockchain meets IoT: an architecture for scalable access management in IoT[J]. IEEE Internet of Things Journal,2018,5(2):1184-1195.
- [3] Oracle. What is the Internet of Things (IoT)? [EB/OL]. [2023-04-10]. <https://www.oracle.com/internet-of-things/what-is-iot/#link0>.
- [4] 应宇. 基于物联网的智能家居系统人性化设计体现[J]. 产业创新研究,2022(6):60-62.
- [5] 曹雅丽. 数字政府建设发展加速工业互联网取得积极进展[J]. 智慧工厂,2022(1):18-19.
- [6] 邓昀,程小辉. 面向物联网的智能家居系统设计[J]. 桂林理工大学学报,2012,32(2):259-264.
- [7] 臧克家,李宝营,杨剑飞. 基于 STC89C52 和 TC35 的智能家居系统设计[J]. 大连工业大学学报,2015,34(5):383-386.
- [8] 王涛. 基于 51 单片机的 LED 可见光通信系统[J]. 电视技术,2013,37(9):61-63.
- [9] 颜增显,闫秀燕. 基于单片机 STC89C52 智能小车设计[J]. 火力与指挥控制,2011,36(12):198-201.
- [10] 马聪,华亮,羌予践. 一种基于 SpringBoot 架构下的水质监测系统的设计[J]. 电子器件,2021,44(5):1109-1114.

Design of IoT smart home solution based on SpringBoot

Chen Xin, Ruan Jinxin*, Wu Yilin, Zhang Qian, Chen Yongchang, Xiong Yan

(School of Computer Science, Guangdong University of Education, Guangzhou 510303, China)

Abstract: With the continuous development and breakthroughs of modern information technology, smart homes are gradually entering our daily lives. To meet the personalized intelligent control needs of users for home devices, a smart home solution with environmental data collection, cloud remote intelligent control, and visual data display is designed based on the system architecture of “microcontroller + web service” using technologies such as sensors, wireless communication, and cloud services. The solution can simultaneously connect 6 home devices and carry out intelligent control, and its biggest feature is that user can achieve intelligent control of the devices through cloud configured sensor thresholds. Through functional verification, the solution can be used in real life, and its design idea and system architecture have certain reference value.

Keywords: smart home; STC89C52 microcontroller; sensors; Internet of Things; SpringBoot

(上接第 70 页)

The application of BP neural network in the automatic leveling system of measurement and control equipment

Xue Chanming*, Zhao Xiangyu

(Taiyuan Satellite Launch Center, Taiyuan 036301, China)

Abstract: With the rapid development of the Ka frequency-band vehicle test and control equipment in the aerospace field, it has promoted the widespread application of electronic level sensor automatic leveling systems in vehicle measurement and control equipment. At present, the device uses the signal of the leveling system sensor feedback before tracking the target, and calculates the compensation error. During the antenna operation, the center of gravity changes a certain system error, which reduces the external test accuracy of the device. The BP neural network is used to dynamically compensate the signal prediction error fed back by the electronic level sensor when the antenna is tracking the target in real time. Through actual verification, it proves that the equipment measurement error can be reduced and the purpose of improving measurement accuracy can be achieved.

Keywords: measurement and control equipment; automatic leveling system; BP neural network

文章编号: 1007-1423(2023)11-0084-06

DOI: 10.3969/j.issn.1007-1423.2023.11.017

基于区块链的公益捐赠系统研究

李保东*, 高洪美

(上海计算机软件技术开发中心软件平台服务部, 上海 200443)

摘要: 在现行的公益捐赠体系下, 捐赠数据保存在慈善机构内部, 信息披露的及时性与真实性难以保障。区块链技术是一种去中心化的分布式数据库, 具有信息追溯、数据透明、无法篡改等特征, 能够有效化解慈善行业的信任危机。建立以区块链技术为基础的公益捐赠平台, 可以有效连接捐赠者、慈善机构、监管机构和受赠者, 使得公益捐赠安全可信、有据可查, 对构建健康发展的公益生态体系具有促进作用。

关键词: 区块链; 公益捐赠; 捐赠溯源; 信息共享

0 引言

近年来, 随着社会生产生活的日益向好, 公益事业发展取得了长足的进步。在面临诸如新冠疫情、地质灾害、应急救援等多种突发事件前, 公益慈善机构作为其中重要的资金及物质支撑来源, 发挥了重要的作用。然而, 公益慈善组织等中心化的运营机构对资金的来源、分配以及流向处理往往不够透明, 公众对慈善机构产生质疑在所难免。同时, 随着移动互联网的兴起, 信息的流通程度不断增强, 作为公益机构的慈善组织时常爆出负面新闻, 这更进一步加剧了慈善机构的运营现状^[1]。公益捐赠概括起来主要有以下几个问题: ①捐赠数据不够透明。现行的公益机构体系下, 接受多方捐赠的物资及善款很难公开透明地列出收入和支出明细, 拉低慈善机构的公信力, 降低公众捐赠的热情。②相关法律法规难落实。虽然慈善机构受到多个部门的监督, 但对于慈善工作中涉及的业务明细很难详细记录, 很容易给一些不法活动提供可趁之机。③公众无法参与监督。对于公众参与的慈善事业, 应当获取捐赠善款

的使用情况, 但对于捐助资金与物资的使用方式, 公众在现行条件下很难参与其中。

针对这种状况, 目前已出现相关研究。徐度康^[2]分析了中心化模式下捐助数据真实验证的问题, 提供了在分布式存储数据情况下的数据真实性验证方法。温直锋^[3]基于区块链智能合约、加密算法等技术, 设计了善款追踪系统。王芳等^[4]基于区块链技术, 研究了在重大疫情下医疗物资的供应管控体系, 分析了医疗物资供应链与区块链技术的耦合机理, 对促进物资有效配送具有一定效果。崔军等^[5]从区块链技术赋能捐赠的角度, 阐述在慈善事业中的协同治理框架及应用模式, 通过对节点的调整、智慧合约的设计, 以及对漏洞的修复能够实现捐赠系统的不断迭代升级。许靖杰^[6]利用跨链技术, 促使慈善项目数据在不同的区块链平台之间转移, 同时构建联盟链来实现数据在公益平台上的信息共享。陈东利等^[7]阐述了通过区块链技术, 能够赋能三次分配, 解决慈善治理面临的公平及效率问题。通过研究分析可知, 目前的研究多从建立信息系统的角度, 来促进

收稿日期: 2023-03-16 修稿日期: 2023-03-24

基金项目: 高效安全商密区块链服务平台研发及应用推广(SKY2022005); 基于可信计算的区块链跨链适配及评估工具研发(21511101504); 上海市软件技术创新服务平台(20DZ2291700)

作者简介: *通信作者: 李保东(1992—), 男, 安徽阜阳人, 硕士研究生, 工程师, 研究方向为区块链、大数据、数据治理, E-mail: lbd@sscenter.sh.cn; 高洪美(1986—), 女, 山东临沂人, 硕士研究生, 工程师, 研究方向为大数据、数据治理、区块链等

捐赠数据的追溯,但对于捐赠者的捐赠证明、捐赠数据核验、添加监管机构作为公证节点等方面的研究还有所欠缺。

本文基于公益捐赠面临的难题,结合区块链技术,研究慈善捐赠模式,构建链接捐赠者、慈善机构、监管机构和受赠者的区块链公益捐赠系统,促使捐赠数据可信、可查、可验证,为捐赠人提供捐赠证明,进而来促进良性的公益生态体系。

1 相关技术

区块链作为比特币的底层技术,为比特币的发行、交易、安全验证及存储提供了有力支撑^[8]。比特币的出现,促进了一大批诸如以太坊、莱代币、EOS等加密数字货币的发展。比特币自2009年发行至今,以其稳定的运行能力被大众熟知,并吸引大批的专家学者参与到对区块链的研究中来。除了数字货币,对区块链的研究已延伸至金融、物联网、智能制造、供应链管理、社会治理等多个领域。作为一种分布式的数据库,区块链技术可以用于构建价值互联网,是具备多层次和多个类型应用的价值传输技术^[9]。区块链采用非对称加密技术来保障数据信息的安全性,利用共识机制保证数据无法篡改,添加时间戳能够让交易可追溯,通过分布式存储确保数据的安全存储和公开透明。

1.1 共识机制

共识机制是实现区块链去中心化的基础,也是维护信任的机制关键。目前,现存的共识机制有30多种,比如常见的工作量证明机制(POW)、权益证明机制(POS)、股份授权证明机制(DPOS)、实用拜占庭容错(PBFT)等^[10]。不同的共识机制达成共识的方式不同,其应用场景也有所差异,如表1所示。

本文根据公益捐赠系统这种多中心化的特点,认为实用拜占庭容错共识机制(PBFT)更符合慈善捐赠区块链平台,更符合实际场景。虽然其最大恶意节点数占比略低,但在实用过程中不需要使用token,契合国内不允许发行代币的政策要求。同时,PBFT以其高效率的操作更加符合慈善捐赠的真实情况,特别是应对诸如

新冠肺炎这样的突发状况,PBFT机制才能高效率地提供高并发量的系统服务。

表1 常见的共识机制类型

共识机制	中文名称	区块链类型	最大恶意节点数占比/%	是否需要token
POW	工作量证明	公有链	51	需要
POS	权益证明	公有链	51	需要
DPOS	股份授权证明	公有链	51	需要
PBFT	实用拜占庭容错	联盟链	33	不需要

1.2 智能合约

区块链的合约层封装的是一些脚本代码、算法以及由此衍生出的智能合约,是实现区块链系统灵活编程、便于操作的基础。在系统运行中一旦满足触发机制,系统将会自动执行合约内容,避免了人为干扰因素的同时极大促进了工作效率。本文的底层采用Hyperledger Fabric,Hyperledger Fabric智能合约被称为chaincode,当一个区块链外部的一个应用程序需要访问账本时,就会调用chaincode。大多数情况下,chaincode只会访问账本的数据库组件和世界状态(world state),但不会查询交易记录。同时,chaincode可通过多种不同编程语言实现。

1.3 数据加密

数据加密是发送方将一份数据经过密钥加密转换成无意义的密文,接收方通过密钥解密数据还原明文。数据加密的方式可以划分为哈希加密、对称加密和非对称加密,基于区块链技术的公益捐赠系统主要涉及到哈希加密和非对称加密。采用数据加密技术可以保障敏感信息的安全,实现对用户的隐私保护。

1.4 时间戳

时间戳是通过数字签名产生的数据,签名对象包括原始数据信息、时间、参数等,通过时间戳系统能够在对数据签名时添加可信的时间戳,证明原始的数据信息在签名时已经存在,提供可信的时间证明。在区块链公益捐赠系统中添加时间戳,能够确保每一笔交易具有法律认可的时间效力。

2 方案设计

2.1 捐赠模型

基于区块链的公益捐赠系统的参与者包含捐赠者、慈善机构、政府监督部门、受赠者等部分。该平台上节点之间可以直接进行信息共享,将捐赠的资金明细以及流向记录在区块链中,确保捐款到达求助者手中。基于这个模型,可以将捐赠活动的资金流、信息流统一整合起来,加强对慈善项目的整合能力,提高政府监管部门的监管能力。图1为公益捐赠平台模型。

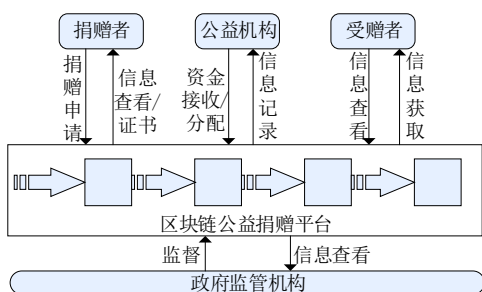


图1 公益捐赠平台模型

(1) 捐赠者是指公益活动中资金或物资的无偿提供方,按捐款主体可分为个人捐赠和集体捐赠,按捐款方式又可以分为公开捐赠和匿名捐赠。个人或集体都可以在公益区块链平台上提交捐赠申请,为了保障捐赠用户的隐私安全,

用户可以选择实名捐赠或匿名捐赠,匿名的捐赠者会以账户地址的方式显示,不会泄露个人隐私。捐赠的信息会在平台上记录下来,并返回给捐赠者捐赠证书。

(2) 公益机构是捐赠过程的中间组织,一般情况下捐赠会通过公益机构将善款送达受赠者手中,起到收集资源与分发资源的作用。慈善捐赠组织可以在慈善捐赠区块链平台上发布公益项目,接受大众捐款,进行资金的接收与分配,并公开透明的记录信息。

(3) 受赠者是接受捐赠的对象,其必须在慈善捐助区块链平台上进行实名验证,并提交求助申请和有关材料,经过审核的对象方可拥有接受捐款的权限。捐赠者处于捐赠链条的末端,是需求节点。对于捐款的数据,捐赠者、政府监管部门、公益机构都有访问权限,可以查询资金的落实情况。

(4) 政府监管部门主要由县级以上人民政府的民政部门进行监督,设置节点的监督与信息查看权限,可以保障监管部门对捐助信息的追溯与查询,快速地定位各条善款的来源去向,实现高效的监管职能。

2.2 整体架构

平台整体架构如图2所示,自下而上可分为基础层、网络层、服务层、应用层、展示层。

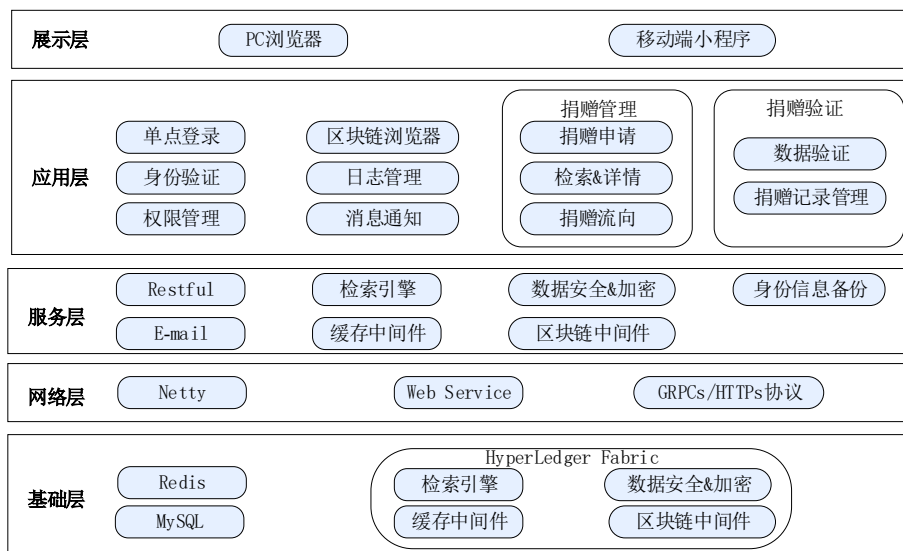


图2 公益捐赠系统总体架构

(1) 基础层包括数据库的存储和底层的区块链技术平台 Hyperledger Fabric。通过 MySQL 数据库, 同时利用 Redis 缓存技术, 能够提升系统性能。区块链组织与账户信息由 Hyperledger Fabric CA 服务颁发, 所有公钥证书等公开信息由本地 MySQL 数据库所备份, 用于应用服务的身份验证。

(2) 网络层是区块链捐赠平台信息传输的必要前提, 通过点对点的组网方式和信息传播机制, 以及传输协议, 促使分布式节点上的参与方都能够记录平台信息。网络层的网络协议和服务如 Netty、Web Service、GRPCS/HTTPs 协议等。

(3) 服务层提供 Restful 接口和 E-mail 网关, 提供检索引擎、数据安全和加密服务、对身份信息备案, 还包括缓存中间件和区块链中间件等。

(4) 应用层是面向用户提供的基本操作功能和具体的应用, 如用户的单点登录、身份验证、权限管理、区块链浏览器、日志管理、消息通知等功能。在具体的捐赠业务处理上, 还包括捐赠行为的管理和捐赠行为的验证服务。如提供捐赠的申请、检索、详情、捐赠流向、捐赠

的数据验证与记录管理。

(5) 展示层是用户与捐赠系统交互的窗口, 可以通过 PC 浏览器或是移动客户端进行相关业务的查看与操作。

2.3 网络架构

各个联盟组织基于 Hyperledger Fabric 区块链技术构建联盟链, 每个组织中分别包含了排序节点对交易进行排序并出块, 以及背书节点对链上交易进行执行、验证、同步。业务系统通过调用捐赠平台接口应用服务(API_SERVER)来上传捐赠数据, 捐赠平台将捐赠的基本信息和摘要信息上传至区块链, 达到数据上链、数据溯源和共享存证的目标。公益机构端的应用服务将开放 Web 网页、小程序供普通用户使用, 开放 Restful API 给普通用户和契约锁调用。图 3 为公益捐赠系统的网络架构。

3 功能设计与实现

3.1 用户及权限设置

区块链公益捐赠平台上的用户角色繁多, 作为该平台的使用和维护方, 公益机构需要对用户进行管理, 并分配用户的权限角色, 不同

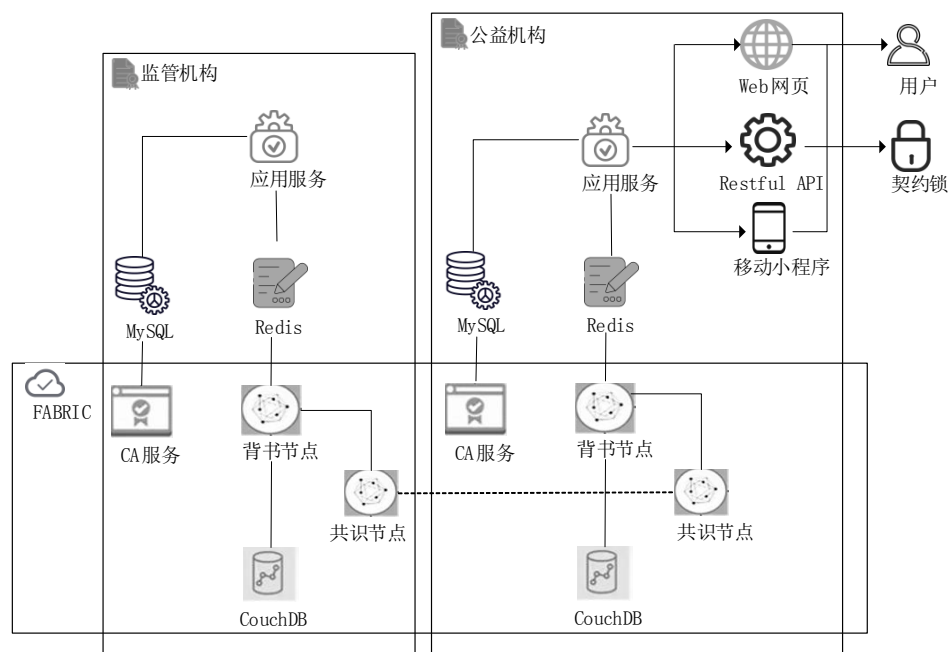


图 3 公益捐赠系统网络架构

的权限所具备的访问能力和操作权限有所区分。对于用户的管理,可以实现用户的增、删、改、查等功能,权限设置可以将角色分配给对应的用户,或是对角色进行数据权限的设计。如图4所示。

3.2 捐赠申请

通过捐赠申请模块,捐赠者可以将捐赠数据存储到公益捐赠平台上,生成唯一的存证哈希值,也可以查看存证的时间及捐赠的交易信息等。同时提供了捐赠信息的统计展示,便于

公益机构了解公益事业的发展情况。如图5所示,为捐赠申请的页面。

3.3 捐赠信息查询核验

捐赠信息的流向,一直是捐赠人以及整个社会关注的问题。对每笔捐赠都生成唯一的捐助资源编号,每一笔善款的打款去向以交易哈希的方式保存在系统中,并且提供在线的文件下载,记录整个项目的基本情况,为善款的使用提供不可篡改的使用证明。如图6所示,为公益捐赠系统的查询核验界面。

用户名称

请输入用户名称

手机号码

请输入手机号码

用户类型

用户类型

创建时间

开始日期 - 结束日期

+ 新增

修改

删除

导出

搜索

重置

☐	用户编号	用户名称	用户昵称	手机号码	用户类型	状态	创建时间	操作
☐	1	admin	超级管理员	13000000000	管理员		2021-07-27 11:12:45	
☐	21	testuser	testuser	13888888888	企业用户		2023-02-15 14:02:19	接口配置 修改 删除 更多
☐	22	admin1	admin1	13333333334	管理员		2023-02-15 14:35:36	接口配置 修改 删除 更多
☐	23	admin2	admin2	13333333335	管理员		2023-02-15 14:42:44	接口配置 修改 删除 更多
☐	24	tp1234	tp1234		普通用户		2023-02-15 15:29:40	接口配置 修改 删除 更多
☐	25	enterprise	enterprise		企业用户		2023-02-15 15:44:58	接口配置 修改 删除 更多
☐	26	sfb	收付宝		企业用户		2023-03-02 15:29:15	接口配置 修改 删除 更多

图 4 用户及权限设置

累计捐赠数

134

当月捐赠数

72

当日捐赠数

6

捐赠申请

存证ID	存证类型	存证哈希	交易哈希	区块链	操作
c724db65-6308-4fc3-8261-61cfa7363384	捐赠申请	9427babb393ed8e650ea5b7047c4bae...	1b95c97ce3585d01630e29925af64c9aa...	129	查看
cdcc6e14-da93-40a9-8594-cf5c2fbc5897	捐赠内容	bd139d686a1850d12033ae5b8bccb74...	9a943850c92a9dc18d5e9bef4fea1c3b8...	128	查看
22696cc9-2de2-4081-83a4-8452e300a728	请输入内容	5a0dfca74f71eac02f1f95b66240dfad76...	29012d1f695822956ac79052f61412ecce...	127	查看
c3eb1f2c-f1cc-44c8-ad42-699bf77cd3a		74fb0c16a95bb3cc8ad212068d4efca0...	e16591a625b0ffc822f563987ae0ab3cb5...	126	查看

图 5 捐赠申请

☐	捐赠资源编号	交易哈希	文件名	文件大小	
☐	26ec91f1-c031-4efe-a2a0-bd59195163d4	c0702de554d8001b8d70fa26aab0bbd24a9bcdabb8b5a813196142c823dd598	A捐赠项目.pdf	1.90 KB	点击下载
☐	aa4fd848-e750-41ec-a3e3-b302294a3c59	c0702de554d8001b8d70fa26aab0bbd24a9bcdabb8b5a813196142c823dd598	B捐赠项目.pdf	1.90 KB	点击下载
☐	2a5aab9-a8f2-4959-ad52-ac7d8bf02385	c0702de554d8001b8d70fa26aab0bbd24a9bcdabb8b5a813196142c823dd598	C捐赠项目.pdf	1.90 KB	点击下载
☐	83e0d38b-dfb8-49f7-86be-1eae50ba5843	c0702de554d8001b8d70fa26aab0bbd24a9bcdabb8b5a813196142c823dd598	D捐赠项目.pdf	1.90 KB	点击下载
☐	40bcfc16-c184-4ee6-bca1-d690cbf8776c	c0702de554d8001b8d70fa26aab0bbd24a9bcdabb8b5a813196142c823dd598	E捐赠项目.pdf	1.90 KB	点击下载
☐	80379dc5-3032-4d3a-81e0-bbb996406a77	c0702de554d8001b8d70fa26aab0bbd24a9bcdabb8b5a813196142c823dd598	F捐赠项目.pdf	1.90 KB	点击下载

图 6 捐赠信息查询核验

4 结语

本文针对公益捐赠所面临的数据不够透明、信息公开不及时、管理效率低下、监管困难等问题,结合区块链技术的特点,提出基于区块链的公益捐赠系统,将捐赠者、慈善机构、监管部门和受赠者有效地连接起来,构建一个信息互通的服务平台。捐赠者的每一笔款项的流通数据保存在区块链网络中,用户可以随时随地查询,保证了信息的可追溯性,极大地提升监管效率。总的来说,区块链技术赋能公益捐赠,能够有效帮助公益机构降低运营成本、提升管理效率,构建可追溯、可信任的公益捐助体系。

参考文献:

- [1] 张卫,张硕.“互联网+慈善”新模式:内在逻辑、多重困境与对策[J].现代经济探讨,2021(11):91-97.
- [2] 徐度康.基于区块链的可信捐助技术的研究[D].南京:南京邮电大学,2022.
- [3] 温直锋.基于区块链的可信慈善系统的设计与实现[D].沈阳:中国科学院大学(中国科学院沈阳计算技术研究所),2022.
- [4] 王芳,李旭东,李强.重大疫情下的医疗物资供应链管理完善与区块链应用研究[J].供应链管理,2022,3(5):85-96.
- [5] 崔军,颜梦洁.区块链赋能慈善捐赠协同治理的框架与应用[J].学术探索,2022(10):117-124.
- [6] 许靖杰.基于区块链的慈善公益平台的设计与实现[D].呼和浩特:内蒙古大学,2022.
- [7] 陈东利,张剑文.区块链技术赋能三次分配:慈善治理公平与效率的现代化表达[J].中国矿业大学学报(社会科学版),2022,24(2):63-74.
- [8] NAKAMOTO S. Bitcoin: a peer-to-peer electronic cash system[EB/OL]. (2008-10-31). [2020-02-25]. <http://www.bitcoin.org/bitcoin.pdf>.
- [9] 袁勇,王飞跃.区块链技术发展现状与展望[J].自动化学报,2016,42(4):481-494.
- [10] 袁勇,倪晓春,曾帅,等.区块链共识算法的发展现状与展望[J].自动化学报,2018,44(11):2011-2022.

Research on public welfare donation system based on blockchain

Li Baodong*, Gao Hongmei

(Software Platform Service Department, Shanghai Development Center of Computer Software Technology, Shanghai 200443, China)

Abstract: Under the current public welfare donation system, donation data is kept inside charities, making it difficult to guarantee the timeliness and authenticity of information disclosure. Blockchain technology is a decentralized distributed database with characteristics such as information traceability, data transparency and non-tampering, which can effectively resolve the crisis of trust in the charity industry. The establishment of a public welfare donation platform based on blockchain technology can effectively connect donors, charities, regulators and recipients, making public welfare donations safe, reliable and well-documented, and contributing to the building of a healthy public welfare ecosystem.

Keywords: blockchain; public welfare donation; donation tracing; information sharing

文章编号: 1007-1423(2023)11-0090-06

DOI: 10.3969/j.issn.1007-1423.2023.11.018

一种低功耗蓝牙 MEMS 振动传感器及其应用

方逸洲^{1*}, 赵中瑞¹, 张青野¹, 张鹏程¹, 艾麦提·布拉丁¹, 曾佑轩², 王 珏²

(1. 上海仪电(集团)有限公司中央研究院, 上海 200233; 2. 上海南洋万邦软件技术有限公司, 上海 200233)

摘要: MEMS 是微机电系统的简称, MEMS 加速度芯片用于测量加速度。处理器芯片通过加速度测量值, 可以计算出运动速度、方向角、振动等信息。蓝牙技术经过 20 余年发展已迭代了 5 代, 从早期单一的外围通讯芯片应用, 拓展到片上系统芯片场景, 尤其从第四代蓝牙技术开始, 其演进了低功耗特性。通过对 MEMS 加速度芯片、蓝牙技术的研究, 基于蓝牙片上系统芯片, 设计一种无需额外处理器芯片的低功耗蓝牙 MEMS 振动传感器, 具有低成本、低功耗、小型化的特点。基于片上系统的处理能力, 结合应用场景需求, 对加速度测量值建立数据处理模型, 并通过蓝牙网关实现数据汇聚、平台展示, 可以形成面向场景的智能化应用。

关键词: 蓝牙; MEMS 加速度; 低功耗; 数据处理

0 引言

MEMS(micro electro mechanical system)是微机电系统的简称, 其主要特点是将传感信号采集、传感信号处理、结果信号输出融合为一体。MEMS 芯片的应用非常广泛, 典型的 MEMS 芯片产品包括手机里的 MEMS 硅麦克风、运动手环中的 MEMS 加速度芯片、工业监测中的 MEMS 压力、MEMS 温湿度芯片等。

本文的研究面向智慧零售场景, 设计了一种低功耗蓝牙 MEMS 振动传感器, 可以分析客户对商超在售物品的拿取等动作行为, 进而给出客户对商超物品的感兴趣指数。同时结合蓝牙无线广播特性和蓝牙网关, 设计了一种移窗过滤均值计算算法, 实现了较为准确的传感器位置判断, 可用于商超物品的防盗管理。本文设计的传感器, 在软硬件层面对产品功耗进行

了深入研究和设计, 采用一颗 CR2032 纽扣电池即可长时间工作。

本文在蓝牙无线通讯领域, 提出了一种基于广播的蓝牙无线数据传输方式。蓝牙网关通过广播扫描可以同时接收多个传感器上传的数据, 并可通过快速周期扫描, 结合本文设计的移窗过滤均值计算算法, 实现对传感器距离的判断。本文设计的传感器是基于 MEMS 加速度芯片的, MEMS 加速度芯片输出 $x/y/z$ 三轴的加速度测量值, 且可以结合滤波、中断等配置, 针对应用场景开发相关数据处理模型^[1]。

本文设计的低功耗蓝牙 MEMS 振动传感器, 结合蓝牙网关及相关处理算法, 实现了智慧零售场景下的防盗管理、客户对商品动作行为和兴趣度分析等功能。重点研究了 MEMS 振动传感器的低功耗设计、MEMS 加速度芯片数据处

收稿日期: 2023-02-27 修稿日期: 2023-03-02

作者简介: *通信作者: 方逸洲(1989—), 男, 硕士研究生, 工程师, 研究方向为人工智能物联网, E-mail: fangyz@rc.inesa.com; 赵中瑞(1993—), 男, 河南周口人, 本科, 研究方向为物联网; 张青野(1991—), 男, 河南平顶山人, 本科, 工程师, 研究方向为人工智能物联网; 张鹏程(1996—), 男, 安徽芜湖人, 硕士研究生, 研究方向为物联网; 艾麦提·布拉丁(1992—), 男, 新疆喀什人, 硕士研究生, 工程师, 研究方向为嵌入式、人工智能; 曾佑轩(1992—), 男, 四川泸州人, 硕士研究生, 工程师, 研究方向为人工智能物联网; 王珏(1989—), 男, 上海人, 本科, 工程师, 研究方向为人工智能物联网

理算法、蓝牙广播无线数据传输、蓝牙接收信号强度的移窗过滤均值计算算法等，最终实现了面向智慧零售场景的应用推广。

本文的研究成果已部署在全国十余个零售门店中，累计部署低功耗蓝牙MEMS振动传感器近1000台。

1 系统方案设计

面向智慧零售场景，本文设计的低功耗蓝牙MEMS振动传感器及其应用系统方案如图1所示。

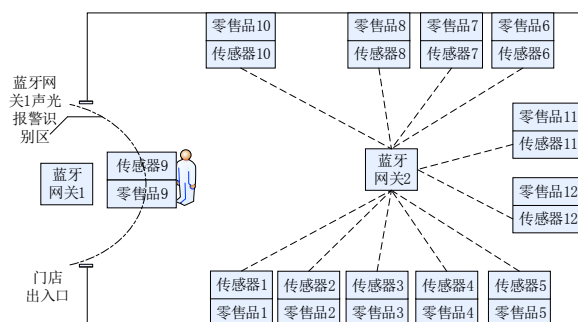


图1 低功耗蓝牙MEMS传感器系统方案

该系统的主要技术实现路径如下：

门店需要监管的每件零售品上，加装一个低功耗蓝牙MEMS振动传感器，传感器采集零售品的加速度信息，并通过蓝牙SoC芯片中设计的数据处理模型，计算出客户对零售品的拿取等动作操作，再通过蓝牙广播将计算分析结果传输到蓝牙网关2。本系统部署时，每个蓝牙MEMS振动传感器具有唯一的序列号，且与对应的零售品绑定，该绑定关系会导入到智慧门店管理平台中，管理平台根据序列号呈现相关零售品的信息。

蓝牙网关2部署在门店的中心位置，其内置功率放大与接收信号增强电路，可以覆盖整个门店。蓝牙网关2通过增强型蓝牙广播扫描，同时接收门店所有传感器的广播数据包，解析出其中的计算分析结果数据，再通过以太网传输到智慧门店管理平台。

智慧门店管理平台接收到蓝牙网关2上传的MQTT数据包后，解析蓝牙MEMS传感器的序列号和动作分析结果，结合数据库中的绑定关系表，分析出零售商品和客户对其的相关操作，并将结果呈现在管理平台界面上。

蓝牙网关1部署在门店的出入口，用于出入口的防盗管理，其内部无需内置功率放大与接收信号增强电路。蓝牙网关1的覆盖范围为1~4 m，部署时可以通过移动端配置软件进行调整。将报警识别距离配置到蓝牙网关1中后，如果有零售品进入到蓝牙网关1的范围内，蓝牙网关1通过接收信号强度的快速周期扫描和移窗过滤均值计算算法，可以较为准确地识别到该零售品，并通过声光报警方式进行警示，实现防盗功能。

2 蓝牙MEMS振动传感器开发

2.1 硬件设计

本文设计的MEMS振动传感器硬件架构如图2所示。

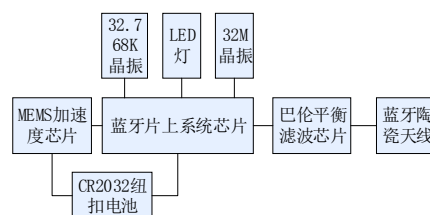


图2 低功耗蓝牙MEMS振动传感器硬件框图

主要模块的技术路径如下：

主控制器：该技术方案无需额外的处理器芯片，选用CC2540F256蓝牙片上系统芯片作为主控制器，在完成数据采集、处理的同时，还可通过蓝牙广播进行数据传输^[2]。

MEMS加速度芯片：选用LIS3DH，该芯片在50 Hz的输出数据速率情况下，低功耗模式下的理论功耗仅6 μ A，适合低功耗传感器的设计。

巴伦平衡滤波电路是射频收发系统中实现平衡转换、阻抗变换等功能的关键模块，本文采用2450BM15A0002芯片，实现该部分功能。

供电：传统的传感器产品，都需要稳压电源芯片。本文设计的传感器无需额外的稳压电源芯片，采用CR2032纽扣电池直接供电的技术路径。CC2540F256的供电电压是2~3.6 V，LIS3DH的供电电压是1.71~3.6 V。以武汉力兴（火炬）电源有限公司生产的CR2032为例，其标称电压为3 V，标称容量是210 mAh，放电曲线如图3所示，可以满足本文传感器的无稳压电源

芯片直接供电的设计方案。

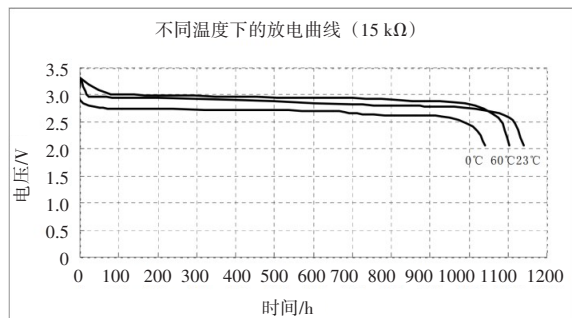


图3 武汉力兴(火炬)电源有限公司生产的CR2032放电曲线

本文设计的传感器实物如图4所示。左边为产品硬件电路正面俯视图，右边为产品组装图。产品尺寸较小，仅为3.7 cm×4.5 cm，可通过挂绳便捷地绑定在零售商品上。



图4 蓝牙低功耗MEMS振动传感器实物图

2.2 低功耗设计

蓝牙低功耗MEMS振动传感器的低功耗设计包括硬件和软件两方面：

2.2.1 硬件低功耗设计

本传感器的硬件低功耗设计主要包括芯片选型、蓝牙芯片引脚电平处理、电源处理三方面。芯片选型方面，选用支持低功耗模式和低功耗中断触发的MEMS加速度芯片LIS3DH，选用支持蓝牙低功耗标准的片上系统芯片CC2540F256。引脚电平方面，由于LIS3DH通过I2C接口与CC2540F256通讯，SDA和SCL两线上有上拉电阻，作为I2C主机的CC2540F256发送完停止信号后，需要避免上拉电阻两边电压差导致的电流消耗。同时，本传感器硬件设计时考虑到兼容性，将LIS3DH的两个中断输出引脚接到了CC2540F256的两个引脚，针对未用到的INT2引脚，需要将CC2540F256对应的引脚配置为输出低电平，防止电流消耗。电源处理方面，如果采用常规的稳压电源芯片设计方

案，会导致漏电流，本产品采用CR2032直接供电方式，减少电流消耗。

2.2.2 软件低功耗设计

软件方面的低功耗主要针对芯片工作模式的配置，CC2540F256需要工作在POWER_SAVING模式下。LIS3DH的工作模式设计较为复杂，包括：配置量程为 ± 2 G；配置50 Hz的数据输出速率；配置为低功耗模式并使能x/y/z三轴输出；配置高通滤波器为正常模式；配置数据输出到寄存器和输入输出缓存器，并使能输入输出缓存器；使能LIS3DH的中断输出引脚1，中断输出信号为高有效；中断阈值为16 mG；触发中断的持续时间为40 ms。通过配置相应的阈值，可以实现动作、非动作的检测^[3]。

软件方面，还对蓝牙广播信号进行了低功耗处理。在没有操作、动作时，关闭蓝牙广播；当LIS3DH中断输出的信号累积到一定程度，判断形成有效的操作动作时，打开蓝牙广播并进行快速周期性广播数据上报，一段时间后再关闭蓝牙广播。

通过上述的软硬件低功耗设计，采用图5所示测试方法对产品静态电流进行测试，为0.008 mA，如图6所示。

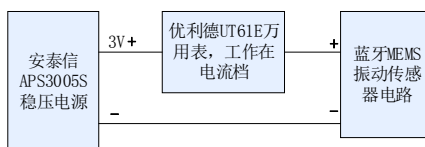


图5 蓝牙MEMS振动传感器静态电流测试方法



图6 蓝牙MEMS振动传感器实测静态电流

2.3 MEMS信号处理算法

针对智慧零售场景下客户对零售品的动作

行为，本文在对 LIS3DH 进行相关设计配置后，还设计了一套 MEMS 信号处理算法，如图 7 所示。

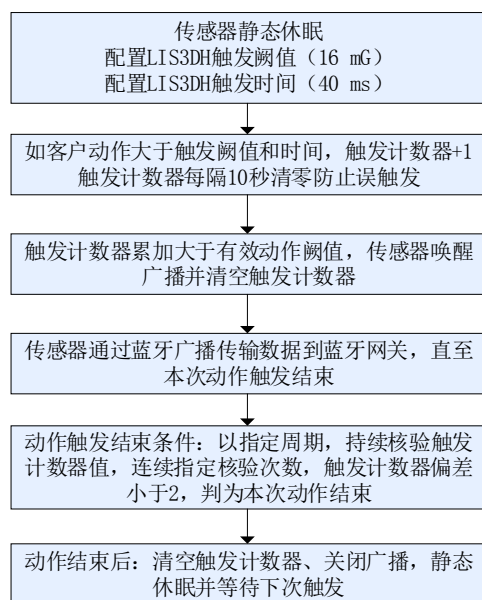


图 7 MEMS 信号处理算法

通过图 7 的 MEMS 信号处理算法，可以使得传感器在无客户动作时，工作在静态休眠模式，降低功耗。同时，有效动作阈值、指定核验周期、指定核验次数均可以通过蓝牙进行再配置调整，便于适配不同场景下的零售品动作。

3 蓝牙无线传输与无线信号处理算法

3.1 基于广播的无线传输

如图 1 所示，本文研究的智慧门店场景下，会有多个传感器。如果采用传统的蓝牙点对多星型网络连接方式进行数据传输，其连接数量受限、传输距离较近，难以满足实际场景要求。本文设计了一套基于广播的点对多蓝牙传输方式，蓝牙传感器的广播包字段格式见表 1。

蓝牙网关快速周期扫描蓝牙传感器，并通过 INESA_SERV_UUID 进行过滤识别，可以有效过滤掉其他非本文设计的蓝牙设备。蓝牙网关设置的广播包单次扫描缓存空间最多支持 40 个蓝牙传感器设备，即支持单一门店的同时 40 名客户对零售商品发生动作的数据上传。经门店部署验证，该技术方案可以满足场景需求。

表 1 蓝牙传感器的广播包字段格式

字段	内容
广播包类型长度标识	0x02
广播包类型标识	GAP_ADTYPE_FLAGS
广播包类型内容	DEFAULT_DISCOVERABLE_MOD E GAP_ADTYPE_FLAGS_ BREDR_NOT_SUPPORTED
数据包内容类型长度标识	根据数据包内容
数据包内容类型标识	0xFF
数据包内容	数据包内容，含客户代码、门店 识别号、传感器序列号、电量、 动作信息、索引值
过滤 UUID 类型长度标识	0x03
过滤 UUID 类型标识	GAP_ADTYPE_16BIT_MORE
过滤 UUID 内容	INESA_SERV_UUID

3.2 基于移窗过滤均值计算的防盗报警算法

零售门店还有防盗的需求，顾客携带未结算的商品离开门店时，行走过程中的振动会唤醒蓝牙 MEMS 振动传感器，使其处于蓝牙广播状态，从而被蓝牙网关 1 识别到。考虑到防盗识别距离较近，蓝牙网关 1 内部不需要功率放大和接收信号灵敏度增强电路。如图 1 所示，在门店出入口部署的蓝牙网关 1，其内部运行本文设计的接收信号灵敏度的移窗过滤均值计算算法，算法流程如图 8 所示。

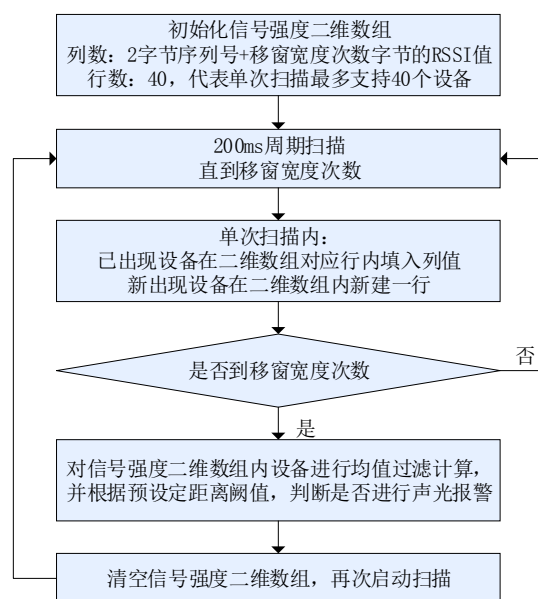


图 8 移窗过滤均值计算算法

基于上述算法,对本文的蓝牙网关1计算出的接收信号强度,以及网关与蓝牙传感器的距离进行了重复性测试,经过测试,形成对应关系表(表2),便于现场部署时进行配置。

表2 本文蓝牙网关1与距离对应关系

蓝牙网关1计算出的 接收信号强度 RSSI	蓝牙网关与传感器的 距离/m
-57	1
-62	2
-66	3
-69	4

4 现场部署与平台展示

蓝牙网关采集到的数据,通过 MQTT 发送

到管理平台,数据包格式采用JSON格式,包括网关序列号、传感器序列号、传感器剩余电量、动作计数累加值、网关部署门店地址等信息,具体格式为:{"Gateway_ID": "网关序列号"; "Command": "Move"; "Tag_ID": "标签序列号"; "Power": 电量值; "Tag_Count": 累加值; "Addr": "shanghai"}。

本文设计的蓝牙 MEMS 振动传感器、蓝牙网关和管理平台系统,部署在上海某大型展会的门店后,可以较为准确地分析出客户对在零售品的动作次数,如图9所示,2合1抓绒毯(黑色款)是该门店中客户最感兴趣的商品。管理平台还对接了商品管理系统,详细展示了商品分类、数量等信息。

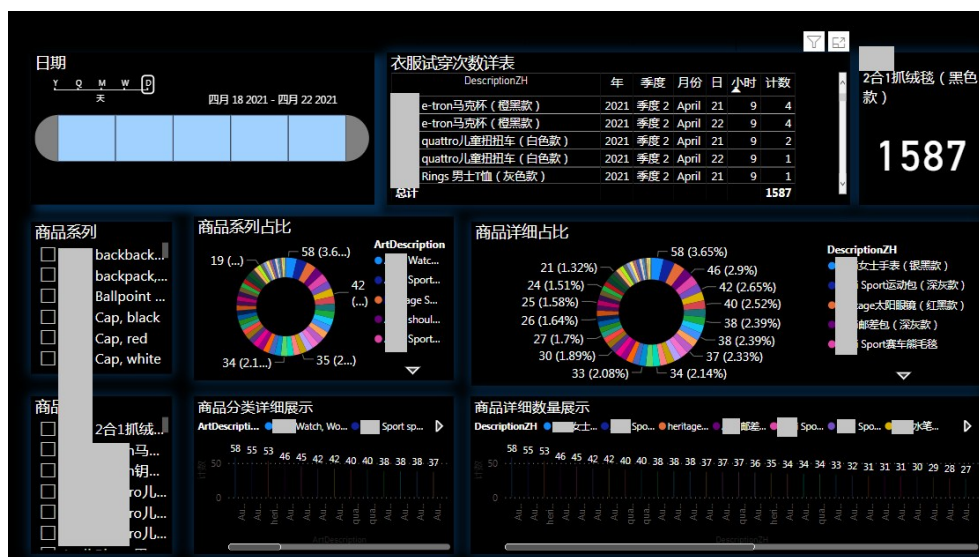


图9 智慧门店管理平台

5 结语

本文设计了一种低功耗蓝牙 MEMS 振动传感器,重点研究了 MEMS 振动传感器的软硬件低功耗设计、MEMS 加速度信号的数据处理算法,并设计了一种基于蓝牙广播的数据通讯方式和接收信号强度的移窗过滤均值计算算法。通过相关技术研究与开发部署,构建了一套包含蓝牙 MEMS 振动传感器、蓝牙网关和管理平台的智慧门店系统,可以采集和统计客户对零售商品的行为动作、对门店零售品的防盗警示等功能。本文的相关研究还可以推广到人员定

位管理^[4]、人员行为分析等更多的场景中。

参考文献:

- [1] 赵虎,王运新,郑佳琳,等.基于MEMS加速度计的物体倾角无线测量系统[J].仪表技术,2013(1): 48-50.
- [2] 陈文辉,罗文广,杨叙.基于蓝牙技术的振动测量仪设计[J].实验技术与管理,2010,27(11):120-123.
- [3] 干娜,王梦轩,余贵萍.基于MEMS加速度的蓝牙手势小车设计[J].数字技术与应用,2016(11):164.
- [4] 陶佳峰,杨晓洪,王剑平,等.基于RSSI的室内测距模型优化技术[J].传感器与微系统,2017,36(1): 24-26,30.

(下转第108页)

文章编号: 1007-1423(2023)11-0095-05

DOI: 10.3969/j.issn.1007-1423.2023.11.019

基于 Android 的防走失 APP 的设计与实现

周 漫¹, 赵嘉颖², 温毅娴^{3*}

(1. 湖南工学院外国语学院, 衡阳 421002; 2. 厦门大学管理学院, 厦门 361005;

3. 湖南工学院商学院, 衡阳 421002)

摘要: 人口有流动就会存在走失的隐患, 人口走失是一个民生痛点。一部智能手机, 一个推送弹窗, 科技重新定义了寻人方式, 让“大海捞针”变成了“精准找寻”^[1]。开发一款功能全面、方便简洁的寻人 APP 变得很有必要。市面现有的相关 APP 大多功能单一, 并不能完整地体现互联网寻人特点, 为此开发基于 Android 的寻人 APP。综合大数据时代特性, 及时发布走失人员信息, 快速检索相关功能信息, 达到高效寻人的效果。APP 使用 Visual Studio Code 进行前端开发, 调用 Google 地图 API 实现精准定位, 利用 MySQL 数据库进行信息存储, 采用 Spring Boot 框架进行后端开发, 建立一个真正高效、便捷的寻人 APP^[2]。

关键词: APP; 人口走失; 寻人

0 引言

在 2021 年 2 月 25 日的“头条寻人”五周年沟通会上, 中民社会救助研究院与“头条寻人”项目组, 联合发布《中国走失人口白皮书(2020)》, 白皮书中的统计数据显示, 2020 年, 在中国, 这个世界上被认为是最安全的国家, 走失人次达到了 100 万^[3]。换算成天数, 意味着我国平均每天处于或经历走失状态的人次为 2739 人。这个数据无疑是令人痛心的, 而每一个走失人员的背后都有着一个苦苦等待、夜夜期盼的家庭。面对此等社会现象, 我国迅速展开行动, 2021 年 1 月起, 部署全国公安机关开展以侦破拐卖儿童积案、查找失踪被拐儿童为主要内容的“团圆”行动, 依托“打拐 DNA”系统整合力量资源, 全面寻找失踪被拐儿童。截止 2021 年 3 月, “团圆系统”的找回率高达 98.1%^[4]。这一消息无疑是令人欣喜的, 不过面对坏人层出不穷的手段, 这些还远远不够。因此, 本文基于 Android 的防走失 APP, 不但针对儿童进行服务, 同时

将老人、重度精神病患者纳入其中。主要囊括信息发布区、社区讨论区、走失大数据地图功能板块等。而该系统的实现帮助家长可以在系统内进行信息交流, 助力走失的孩子、老人、重度精神疾病患者早日找到回家的路。

1 APP 设计

1.1 设计理念

本系统基于通信运营商的覆盖范围和系统搭载使用的衍生产品中储存的数据资料, 形成集精确定位、智能采集信息、实时信息发布及响铃预警、有效信息匹配等于一体的防走失系统, 对接各级公安、医疗、养老等机构平台的接口^[5], 辅助这类机构公安在该指定平台上实时发布信息, 做到事前利用产品降低风险指数; 事中智能穿戴产品响铃并及时传输信息, 家属在数据终端接收及时联系公安, 采取措施, 抢占救援先机; 事后能在该 APP 上传相关受害人信息, 提高数据传播速度, 扩大有效信息来源。

收稿日期: 2023-02-04 修稿日期: 2023-03-06

基金项目: 2021 年湖南省大学生创业训练项目(S202111528010X): 安全铃——专属特殊人群的提醒

作者简介: 周漫(2000—), 女, 湖南衡阳人, 在读本科, 研究方向为商务英语; 赵嘉颖(2001—), 女, 湖南衡阳人, 在读硕士, 研究方向为物流管理与工程; *通信作者: 温毅娴(1986—), 女, 湖南衡阳人, 硕士, 副教授, 研究方向为高等教育, E-mail: 276171532@qq.com

1.2 功能设计

1.2.1 专设走失三大功能

据调查市面上目前尚未出现一款完全针对减少走失的APP可供使用,本项目APP针对走失情况主要提供三大功能:①APP会通过整理用户响铃频率、响铃位置及时间形成大数据个性化预防警示地图,在丢失的高发地点、时间进行提示以此来增强用户警惕性;②同城搜索功能,使用者打开GPS定位进入APP可实时搜索到附近范围内走失事件的相关发布信息或者是新闻,突破传统寻找走失者的途径,更快速、更便捷地完成信息的传播;③拥有走失信息发布集合平台,走失者家属可在APP上发布走失人员信息,其他人员可对信息进行转发或者在评论区提供相关线索,加快走失信息的传播,吸引群众注意力,有效提高寻回机率。

1.2.2 匹配群体操作习惯

市面上的智能手表附带的APP专门服务于已购买其产品的用户,对其他用户设限,但本项目APP可提供两种模式进行登录,开放平台,服务大众。已购买安全铃的用户选择“进入安全铃”登录,可享用APP所有功能,进入APP后可对安全铃的使用模式进行切换,同时设置佩戴者活动范围,超出范围即刻反馈定位信息至监护人终端设备并进行响铃预警,且随时接受并整理佩戴者的健康检测数据以及定位信息。未购买安全铃的用户可选择“游客模式”登陆,可进入“找找”界面,查看APP收集的走失信息、“安全铃”的序列号快捷搜索以及开启临时接送的代接模式。

2 APP实现

从全局角度出发,分为三个板块对防走失软件系统进行总体阐述,包括系统需求、系统总体架构、数据库结构、数据流图、系统各模块分析。

2.1 系统的功能需求

防走失系统系统的建立需要对系统所有的信息进行建模^[6],为了更好地进行系统功能设计与实现,首先需要对系统的用户需求进行分

析。根据消费人群定位,该系统需要具备监测人物佩戴某特定设备时的位置,记录历史定位,并能及时分析出当设备位置出现异常偏移或停留时的警戒范围。除此之外,该系统还兼具电商平台以及信息平台的功能。系统功能层次如图1所示。

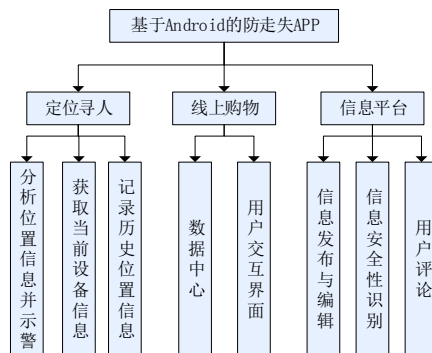


图1 防走失APP的系统功能层次

如图1所示,具体来说,防走失APP的系统中,最重要的模块是定位寻人功能,系统需要具备强大的数据处理能力,首先获取当前佩戴特定设备的人物位置,再通过学习日常人物活动轨迹,当人物位置发生较大偏移时能够及时示警,以便亲友及时发现,避免意外发生;除此之外,还需要系统能够记录历史数据,当位置信息由于设备丢失或信号切断等意外情况无法更新时,能够根据历史数据获取最新情况;系统同时还包含了电商和信息平台的功能,目的是打造多功能一体化的智能信息平台,以便更好地满足客户的多元化需求。

2.2 系统架构设计

在搭建系统平台前,根据系统的功能需求设计系统架构。系统架构如图2所示。本软件系统主要分为前端与后端两个部分,各部分实现不同功能,并存在数据通信,其中前端主要负责获取设备位置信息、连接地图接口、在地图上显示设备实时位置并输入表单;后端实现了数据库连接,读取写入历史位置信息、订单信息、社区信息等以及用户交互操作响应。在前后端进行数据传输的同时存储日志信息到本地存储设备。

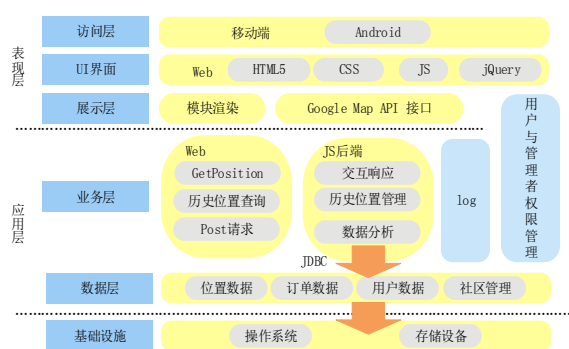


图2 防走失APP的系统架构

防走失 APP 的总体架构采用分层的结构，主要分为表现层、应用层和基础设施层。其中，表现层又分为访问层、UI 界面层和展示层，应用层包含业务层和数据层。由于该 APP 主要面向移动端用户，且为安卓平台软件，所以访问层在 Android 操作系统上运行；UI 界面使用了 Android 控件和 Web 技术，其中使用的技术包括 jQuery、WebView 等；展示层调用了 Google Map API 接口，用于地图的展示，方便用户在移动端查看设备位置。应用层包含了业务层和数据层，业务层描述了系统的业务逻辑，包括获取设备位置信息、历史信息查询、表单的上传、JS 后端上实现的交互响应等；业务层和数据层之间通过使用 JDBC 进行连接，以便数据的录入和查询。最后是软件系统的基础设施，主要包括操作系统和存储设备。

2.3 数据流图和数据库设计

通过对系统架构的描述，可以画出数据流图。表1记录了整个APP中数据的终点/源点、处理过程、数据流、数据存储的信息。

在明确了数据流图的四种成分后,画出数据流图,如图3所示。数据流图描述了数据在系统内部的逻辑流向和逻辑变换的过程,可以帮助我们设计防走失软件系统的功能模型,表达系统的逻辑功能。用户群体分为两类,将使用移动端查看佩戴终端设备者位置的用户定为用户1,将佩戴终端设备的用户定为用户2,将二者的信息包括终端设备信息,用户个人信息以及终端设备历史轨迹等信息存入数据库用户信息中。除此之外,还设置了另外两个数据可用于存储商城信息和社区信息。这些数据都将返

回给管理者以便进行数据筛查和软件维护。

表 1 数据流图的四种成分

终点/源点	处理	数据流	数据存储
用户1(亲友)	产生信息	用户	用户信息
用户2(被保护者)	处理事务	定位设备编号	商城信息
管理员		佩戴者个人信息	社区信息
		历史数据	
		商城	
		订单编号	
		卖家信息	
		买家信息	
		社区	
		帖子	
		发帖人信息	

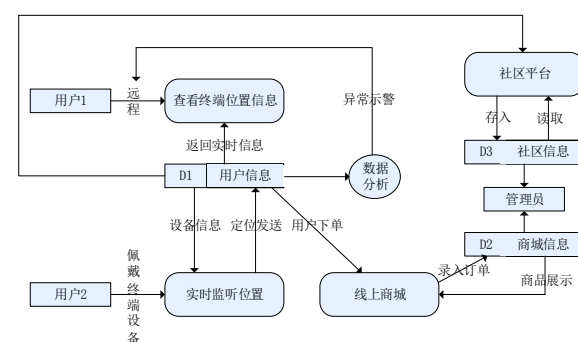


图 3 防走失 APP 的数据流图

在数据流图和系统架构图的基础上，对软件所需的三个数据库进行设计。首先进行用户信息数据库的搭建，在该数据库中包含三个表，如图4所示。在这张数据表中，编号同时作为三张表的主键，便于检索。用户表存放用户的基本信息，健康数据表存放设备监测佩戴终端用户的健康数据，位置数据实时记录佩戴终端用户的位置数据，三个表相互联系，共同构成了用户信息数据库。

其次，作为软件的重要商业功能，还需搭建线上商城的数据库，在这个数据库中同样包含三个表，分别是商品信息表、购物车信息表以及订单表。如图5所示。

最后是社区信息数据库，这个数据库需连接到用户信息数据库来记录发帖人信息，以便管理者管理社区(如图6所示)，这个数据库包含了评论表、发帖信息表，评论表和发帖信息表中都使用了用户编号作为外键来获取用户信息。

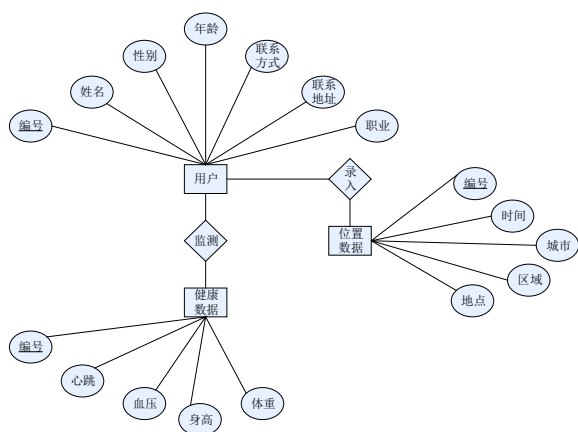


图4 用户信息数据库的E-R图

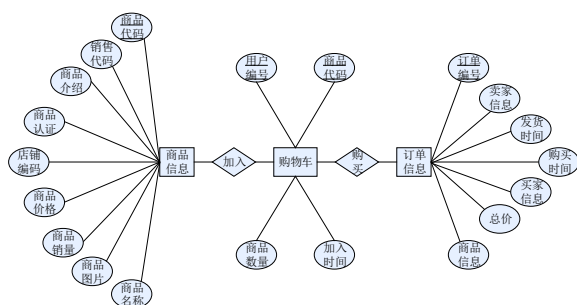


图5 商城信息数据库的E-R图

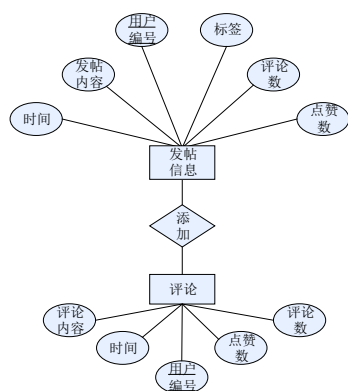


图6 社区信息数据库

2.4 APP实现

APP获取的数据大致为两种,一种是直接请求开发者服务器获得,一种是需要通过开发者服务器转发请求而获得。本平台直接向开发者服务器发送请求,开发者服务器收到后,执行对应的程序,获取到对应的数据,再将数据返回给家属端^[7]。本文主要采取的是第二种获

取数据的方式,将穿戴者与线上家属端相联,线上传输数据,以便线下及时采取措施。

3 结语

基于Android的防走失APP的设计与实现,将搭配智能穿戴设备,有效利用通信运营商的网络追踪定位、数据库资源,及数据匹配、检测、分析,形成穿戴智能设备、监护人数据终端、防走失平台三者合一的智能信息化系统,涵盖多种特殊人群,最大程度利用物联网技术,降低走失风险,提高寻人效率。基于Android开发的APP,能够短时间内升级新功能,并且该系统界面简洁,功能清晰,易于操作,为用户提供一种便捷实用的监控方式^[8]。通过与公安系统对接,以期及时采取救援,抢占救援先机,为降低走失情况的出现提供了强劲的安全保障。

参考文献:

- [1] 木野飞.“头条寻人”万人归:为科技赋予温度,让善意无处不在[N/OL].(2019-07-25). <https://m.gmw.cn/baijia/2019-07/25/33028166.html>.
- [2] 杨仕昌,叶晟澄,陈家乐,等.基于微信小程序的走失老人报备平台设计[J].信息技术与信息化,2021(7):93-95.
- [3] 陈嘉俊,郑伙群,黄雪丽,等.基于百度AI人脸识别的智能救援指挥系统的设计与实现[J].电脑知识与技术,2022,18(11):79-81,83.
- [4] 于晓.公安部:截至2021年3月“团圆系统”失踪儿童找回率达98.1%[N/OL].(2021-03-15) <https://www2.chinanews.com.cn/gn/2021/03-15/9432441.shtml>.
- [5] 田柳,林黄智,鲁磊.基于5G物联网技术的防走失系统设计[J].科技创新与应用,2022,12(25):16-19.
- [6] 种稼惠,杨硕,李晨明,等.儿童防走失胸针设计与实现[J].轻工科技,2022,38(5):48-49,89.
- [7] 杨仕昌,叶晟澄,陈家乐,等.基于微信小程序的走失老人报备平台设计[J].信息技术与信息化,2021(7):93-95.
- [8] 张斌,鄢晓静,邵想,等.基于Android的防走失系统的设计与实现[J].智能计算机与应用,2020,10(1):173-177,182.

(下转第112页)

文章编号: 1007-1423(2023)11-0099-05

DOI: 10.3969/j.issn.1007-1423.2023.11.020

基于 VizQL 的汽车工业数据可视化系统的设计与实现

杨国丹*

(西南交通大学制造业产业链协同与信息化支撑技术四川省重点实验室, 成都 611756)

摘要: 近年来, 伴随着汽车数量的增长, 围绕汽车产业产生了海量数据, 如何利用这些海量的数据进行有效的分析、推理和决策就成为了重要的研究热点。数据可视化对于从海量的数据中发现规律、增强数据表现、体现数据价值、提升交互效率具有重要的作用。对数据可视化进行研究, 提出了一种基于 VizQL 的汽车工业数据可视化系统, 通过良好的交互体验和数据分析能力, 为汽车行业提供更好的生产和服务价值。

关键词: 汽车工业数据; 数据可视化; VizQL

0 引言

随着我国经济的快速发展, 人民生活水平日益提高, 汽车也走进了千家万户。中国汽车保有量继 2020 年底赶超美国成为全球第一之后, 继续保持高速增长, 2021 年中国的汽车保有量将突破 3 亿大关^[1]。随之产生的海量的、不同形式的、不同来源的数据应该如何更好地为汽车产业中的产品设计和研发、制造过程优化、营销和销售、服务和售后赋能是我们亟待解决的重要问题之一。

针对上述问题, 本文设计了基于 VizQL 的汽车行业数据可视化系统, 帮助汽车研发人员更直观地了解市场需求、产品特性和竞争对手情况, 有利于进行产品设计和研发; 帮助企业进行生产计划和流程优化, 提高生产效率和质量; 帮助企业了解消费者需求、市场趋势和竞争对手情况, 从而制定更合适的营销策略和销售计划, 提高销售效率和市场占有率; 帮助汽车企业和经销商提高售后服务质量, 满足消费者需求。

1 相关工作

数据可视化是将数据转化为图形或图表等

视觉元素的过程。通过将数据可视化, 我们可以更直观地看到数据之间的关系、趋势和模式, 进而更好地理解和分析数据。数据可视化可以使得大量的复杂数据更易于理解和沟通, 使得数据分析更加高效、准确。

文献[2]提出了一个数据可视化推荐系统 DeepEye, 通过将一组算子转换数据集, 以一种强制的方式自动可视化给定的数据集。文献[3]提出了一个可视化推荐引擎 SeeDb, 在给定的数据子集中, SeeDb 智能地探索可视化空间, 评估其最有趣的可视化推荐。

一些针对于实际场景的数据可视化分析系统, 包括: 邓智豪^[4]通过数字驾驶舱的方式实现了自定义可视化分析引擎的系统, 用户可以根据不同的业务场景来定制适合业务的可视化面板; 丁超凡^[5]使用 Web 技术和数据可视化技术设计了交通数据可视化分析系统, 通过多角度分析, 呈现交通特性和演变规律。

2 VizQL

2.1 VizQL 概念

VizQL 是由斯坦福大学的 Polaris^[6]系统演化

收稿日期: 2023-02-27 修稿日期: 2023-03-05

作者简介: *通信作者: 杨国丹(1997—), 男, 四川乐山人, 硕士研究生, 研究方向为大数据与云计算, E-mail: Guodan-Yang_nn@163.com

而来,它是数据可视化查询语言,它允许用户通过拖放和点击的方式直接操作可视化元素,从而轻松地探索和分析数据。VizQL将用户的可视化请求转换为SQL查询,然后将查询结果渲染为交互式可视化。通过VizQL,用户可以直接在可视化过程中进行数据操作,例如对图表中的数据点进行筛选、排序、聚合等操作,而无需手动编写代码或执行SQL查询。这种交互式的数据探索方式使得数据分析变得更加直观和灵活,降低了使用者的技术门槛,并帮助用户更快速地发现数据中的关键信息。VizQL的主要优势之一是其高度的可扩展性。它支持多种数据源,包括关系型数据库、非关系型数据库、文本文件、Web数据等,并且可以将多个数据源组合在一起,以便更好地分析和理解数据。此外,VizQL还支持自定义计算字段、多维度分析、动态参数控制等高级功能,以满足各种数据分析需求。

2.2 VizQL的工作原理

2.2.1 表代数

首先需要定义一个正式的机制来指定表配置,为此定义了一个代数。一个完整的表配置由这个表中的三个单独的表达式组成。其中两个表达式定义了表的 x 轴和 y 轴的配置,将表划分为行和列。第三个表达式定义了表格的 z 轴,它将显示划分为多个层次。

这个表代数中的操作数是数据库的维度和度量字段的名称。我们将用 X 、 Y 和 Z 来表示维度场(如数据库表中的企业名称、企业地址等),用 P 、 Q 和 R 来表示度量场(如数据库表中的企业年收入、月收入等)。我们以下述方式为每个字段分配有序集:对维度字段,我们分配字段的维度域的成员;对度量字段,我们分配包含字段名的单个元素集。

$$X = \text{DOMAIN}(X) = \{x_1, \dots, x_n\} \quad (1)$$

$$P = \{P\} \quad (2)$$

将集合分配给符号反映了两种类型字段在表结构中的编码方式的差异。维度字段将表划分为行和列,而度量字段将在空间上编码为窗格中的轴。在代数中,一个有效的表达式是一个由一个或多个符号组成的有序序列,在每对

相邻的符号之间有运算符,并用括号来改变运算符的优先级。代数中的运算符是交集($\times$)、阈值($/$)和关联($+$),按优先顺序列出。每个运算符的精确语义是根据其对操作数集的影响来定义的。

其中关联的运算为

$$\begin{aligned} X + Y &= \{x_1, \dots, x_n\} + \{y_1, \dots, y_m\} \\ &= \{x_1, \dots, x_n, y_1, \dots, y_m\} \end{aligned} \quad (3)$$

$$\begin{aligned} X + P &= \{x_1, \dots, x_n\} + \{P\} \\ &= \{x_1, \dots, x_n, P\} \end{aligned} \quad (4)$$

$$\begin{aligned} P + Q &= \{P\} + \{Q\} \\ &= \{P, Q\} \end{aligned} \quad (5)$$

交集的运算为

$$\begin{aligned} X \times Y &= \{x_1, \dots, x_n\} \times \{y_1, \dots, y_m\} \\ &= \{x_1 y_1, \dots, x_n y_1, x_n y_1, \dots, x_n y_m\} \end{aligned} \quad (6)$$

$$\begin{aligned} X \times P &= \{x_1, \dots, x_n\} \times P \\ &= \{x_1 P, \dots, x_n P\} \end{aligned} \quad (7)$$

阈值运算为

$$X/Y = \{x_i y_j \mid \exists R \text{ set } X(r) = x_i \text{ and } Y(r) = y_j\} \quad (8)$$

其中: R 表示要分析的数据集; r 被定义为一条记录; $X(r)$ 表示记录 r 的字段 X 的值。

对每个运算符使用上述集合语义,代数中的每个表达式可以简化为单个集合,集合中的每个条目是零个或多个维度值与零个或多个度量字段名的有序连接。我们称这个表达式的集求值为规范化集形式。表达式的规范化集合形式决定表的一个轴:表轴被划分为列(或行或层),因此规范化集合中的集合条目和列之间有一一对应关系。

2.2.2 生成数据库查询

通过上述生成可视化规范查询对应于生成数据库查询,该查询选择用于分析的数据子集,然后筛选、排序将结果分组到窗格中,最后将数据传递给图形编码。

首先我们先选择数据源,对于维度字段 X ,用户可以将字段的域的子集指定为有效。如果 $filter(x)$ 是用户选择的子集,那么表示 x 的过滤器的关系是 $X \text{ in } filter(X)$ 。

对于度量字段 P ,用户可以将字段的域的

子集定义为有效。如果 $\min(P)$ 和 $\max(P)$ 是这个子集的用户定义范围,那么表示 P 的过滤器的关系是 $(P \geq \min(P) \& P \leq \max(P))$ 。

所以我们可以将上述关系表示为第一阶段的 SQL 语句: $SELECT * WHERE \{filters\}$ 。

其次将检索到的记录划分为对应于表中每个窗格。正如前文所讨论的,表轴表达式的规范化集合形式决定了表的配置。表被划分为与这些集合中的条目相对应的行、列和层。假设 $Row(i)$ 是表示第 i 行选择标准的代表, $Col(j)$ 是表示第 j 列选择标准的代表, $Layer(k)$ 是表示第 k 层选择标准的代表。如果表的 y 轴是由规范化集定义的: $\{x_1y_1P, x_1y_2P, x_2y_1P, x_2y_2P\}$ 。

那么可以通过一下查询检索到要分区到第 i 行、第 j 列、第 k 层交汇处的窗格中的记录:

$SELECT * WHERE \{Row(i) \text{ and } Col(i) \text{ and } Layer(k)\}$

最后一步就是转换每个窗格内的记录。如果可视化规范包括聚合,则必须为数据库中的

每个度量值分配聚合运算符。如果用户没有指定运算符,那么我们默认地为度量值设定聚合运算符为 SUM。度量值的运算符由两个组成,定义为

$$aggregates = \text{SUM, AVG} \quad (9)$$

所以最终形式的 SQL 语句为

$$\begin{aligned} &SELECT \{dim\}, \{aggregates\} \\ &GROUP BY \{T\} \\ &HAVING \{filters\} \\ &ORDER BY \{S\} \end{aligned}$$

3 系统展示

本文基于 VizQL 开发了汽车工业数据可视化系统,并对该系统的应用进行情况说明。

我们根据指定的数据源,选择不同的场景需求,可以指定以下可视化分析界面,不同类型界面见图1、图2、和图3。

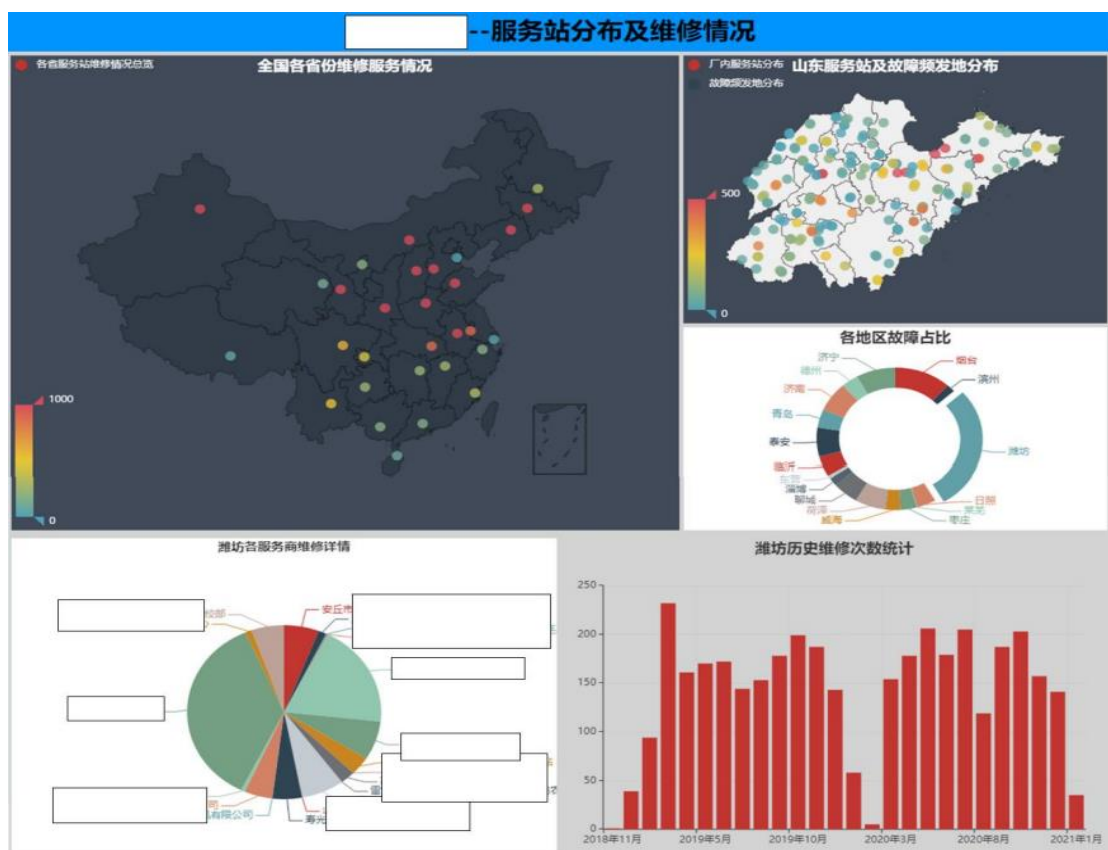


图1 服务站分布可视化分析界面



4 结语

本文提出了一种基于 VizQL 的汽车行业数据可视化系统，通过利用 VizQL 语言实现对汽车行业数据可视化的转换，使用户可以获得更丰富的可视化界面和分析界面。在未来的工作中，交互式响应率和交互历史是一个值得关注的问题。

参考文献：

- [1] 陈秀娟. 全国汽车保有量达 2.81 亿辆[J]. 汽车观察, 2021(1):7.
- [2] QIN X, LUO Y, TANG N, et al. DeepEye: an automatic big data visualization framework [J]. Big Data Mining and Analytics, 2018, 1(1):75-82.
- [3] VARTAK M, RAHMAN S, MADDEN S, et al. S_{EE}DB: efficient data-driven visualization recommendations to support visual analytics [J]. Proceedings of The VLDB Endowment, 2015, 8(13):2182-2193.
- [4] 邓智豪. 基于大数据的自定义可视化分析引擎的设计与实现[D]. 北京: 北京邮电大学, 2021.
- [5] 丁超凡. 交通数据可视化分析及系统设计[D]. 北京: 中国科学院大学, 2021.
- [6] STOLTE C, TANG D, HANRAHAN P. Polaris: a system for query, analysis, and visualization of multidimensional relational databases [J]. IEEE Transactions on Visualization and Computer Graphics, 2002, 8(1): 52-65.

Design and implementation of data visualization system for automobile industry based on VizQL

Yang Guodan*

(Sichuan Key Laboratory of Manufacturing Industry Chain Collaboration and Information Support Technology,
Southwest Jiaotong University, Chengdu 611756, China)

Abstract: In recent years, with the increase in the number of automobiles and the massive data generated by the automotive industry, how to use these massive data for effective analysis, reasoning and decision-making has become an important research hot-spot. Data visualization plays an important role in discovering rules from massive data, enhancing data performance, reflecting data value, and improving interaction efficiency. The research on data visualization found that a VizQL-based automotive industry data visualization system was proposed to provide better production and service value for the automotive industry through good interactive experience and data analysis capabilities.

Keywords: automobile industry data; data visualization; VizQL

文章编号: 1007-1423(2023)11-0104-05

DOI: 10.3969/j.issn.1007-1423.2023.11.021

在线实验报告管理系统的设计与实现

欧阳习彪*, 徐宝林, 蒋 谨

(广东白云学院大数据与计算机学院, 广州 510000)

摘要: 基于当前实验报告不易保存、不易管理、样式排版不统一、批阅太耗时等问题, 对实验教学与管理的实际情况进行研究, 开发了一套符合本学院实际需求, 基于 B/S 架构的实验报告管理系统。系统分为学生端和教师端, 教师端具备创建实验报告任务、在线批阅、批量导出实验报告及成绩等功能; 学生端实现了远程提交、预览、导出实验报告等功能。系统简化了流程, 节约了资源与管理成本, 同时, 加入自动阅卷功能, 极大提高了工作效率和效果。

关键词: 实验报告; 在线批阅; 远程提交; 自动阅卷

0 引言

实践教学在应用型本科院校中尤为重要, 而实验报告通常是实践环节评分的重要组成部分, 能够比较直接有效地反映学生对相关理论知识的理解和掌握程度。目前高校实验教学中, 对学生实验报告提交、收集和管理方式较为繁琐, 主要有邮件收集、U 盘拷贝、FTP 方式、共享文件夹以及纸质实验报告等几种主要方式^[1]。这些方式普遍存在教师需花费大量的时间去批阅每一个学生的实验报告文档, 学生需要花费大量时间去整理格式, 工作效率非常低; 完全靠人工阅读进行考评, 主观性较大等问题, 针对以上弊端, 本文设计研发一种实验报告在线自动生成、提交、批改以及管理的方案^[2]。

1 需求分析

系统分为学生端和教师端, 学生端包括学生基本信息和报告管理两部分, 学生基本信息中学生登录后需要绑定手机号或者邮箱才能使用报告管理功能, 学生可以通过实验报告管理

部分选择对应的实验任务进行编辑提交, 提交后自动生成带格式的实验报告, 学生只需编辑对应的实验内容并不需要关注报告格式; 教师端包括实验参数设置、实验管理、报告管理、学生管理、教师管理五部分, 分为管理员和教师两个角色。先由管理员负责录入学生管理和教师管理的基础数据, 教师进入系统设置实验参数、创建实验任务并发布, 学生通过账号密码登录系统后可看到刚创建的实验任务, 学生编辑任务提交, 在教师端的报告管理列表中生成一条实验报告记录, 如果创建实验室设置的自动阅卷、系统通过定时任务, 每一小时自动批阅一次所有未批阅的实验报告, 未启动自动阅卷功能则需教师手动阅卷打分, 阅卷后学生端便生成一份实验报告档案^[3]。

2 研究的技术路线

该系统采用 B/S 架构, 使用 PHP 语言进行开发, 采用 Nginx 作为反向代理服务器, 选用 MySQL 作为数据库, 系统前后端分离, 前端界面采用 Bootstrap 框架, 使用 Ajax 进行前后端交

收稿日期: 2023-02-28 修稿日期: 2023-03-10

作者简介: *通信作者: 欧阳习彪(1987—), 男, 湖南邵阳人, 硕士, 讲师, 研究方向为软件工程, E-mail: 921632954@qq.com; 徐宝林(1972—), 男, 湖南邵阳人, 学士, 副教授, 研究方向为软件工程、信息系统集成; 蒋谨(1973—), 男, 广州白云人, 硕士, 副教授, 研究方向为人工智能

互，后端基于 THINKPHP6 框架 MVC 架构进行开发，系统中大量使用 redis 作为缓存来提高访问速度^[4-5]，利用 Jieba 第三方中文分词工具实现自动阅卷功能。

3 数据库设计

本系统中主要创建了如下表：

(1) 实验参数表(名为 es_arctype)：主要包括实验名称(typename)、截止时间(deadline)、是否开启自动阅卷(is_auto_decided)、指定发送班级(class_ids)。

(2) 字段表(es_channelfield)：该表的字段在设置实验参数时产生，主要包括字段名称(name)、字段类型(dtype)、默认值(dfvalue)。

(3) 教师表(es_admin)：主要字段有用户名(user_name)、用户密码(password)、手机号(mobile)、邮箱(email)。

(4) 学生表(es_users)：主要字段有用户名(username)、用户密码(password)、手机号(mobile)、邮箱(email)、学生等级(level)。

(5) 学生实验报告记录表(es_archives)：该表不保存具体的实验内容，主要记录学生实验报告分数及教师评语，主要字段有报告作者(author)、实验报告分数(fraction)、教师评语(remark)。数据库中各字段及属性见表 1。

表 1 学生实验报告记录

字段名称	类型	长度	说明
aid	int	10	主键 id
author	varchar	10	作者
fraction	decimal	10	实验报告分数
remark	text	255	评语
add_time	int	10	创建时间

(6) 学生实验报告内容表(es_article_content)：该表与实验报告表关联，记录学生的实验报告的详细信息，该表字段由程序控制进行动态扩建，后台每新增一个实验参数，会在此表生成一个实验参数及其对应的分数字段。数据库中各字段及属性见表 2。

表 2 实验报告内容

字段名称	类型	长度	说明
id	int	10	主键 id
aid	int	10	实验报告记录表 id
group_name	varchar	10	姓名
group_number	varchar	10	学号
principle	longtext		原理内容(新增实验报告任务时自动生成)
principle_score	int	10	原理对应分数(新增实验报告任务时自动生成)
categorize	longtext		分类内容(新增实验报告任务时自动生成)
add_time	int	10	创建时间

4 系统核心功能实现

4.1 实验参数设置

系统管理员在系统中录入教师和学生账号信息后，教师方可进行实验任务的相关操作。实验报告是由一系列参数组构成，教师创建实验任务的第一步是设置实验报告参数。在参数管理中点击新增实验参数，参数中字段类型支持各种输入框类型(各输入框类型已经提前在配置文件中配置好)，参数可以选择是否有分数，比如学生姓名、班级等信息是不需要配置分数，但如实验原理、实验步骤等参数则需要配置分数，界面如图 1 所示。



图 1 实验参数设置

4.2 教师创建实验任务

实验参数设置好之后教师可在实验管理模块中选择增加实验、拟定实验名称、设置实验截止时间和超时扣分数值，设置好相关参数确认提交后会发邮件提醒相关学生有实验任务发布，在截止时间的前两天会自动发送邮件提醒

相对应的学生及时提交,系统支持自动阅卷功能,默认关闭自动阅卷功能,教师必须选择实验参数,实验参数从设置好的参数列表中进行选择,每个参数配置对应的分数、排序值、参考答案(供自动阅卷功能使用,参考答案要求精简,不能有重复的内容以方便计算分数),指定发送班级,提交后学生登录系统自动生成一份实验任务。界面如图2所示。

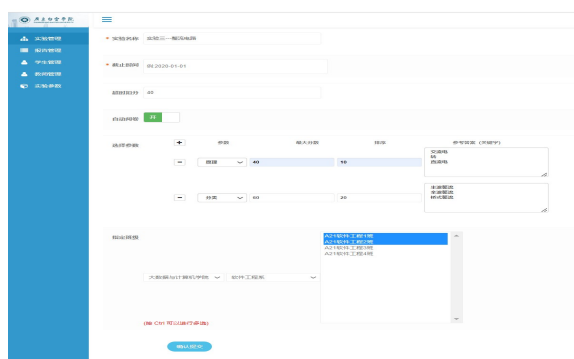


图2 创建实验任务

4.3 学生在线制作实验报告

学生通过账号密码登录实验报告学生端,初次登录需要在学生信息中绑定手机号和邮箱。进入报告管理界面,学生在所属栏目中选择教师布置的任务进行编辑内容提交;在线制作实验报告时学生只需在对应框中填入实验内容,并不需要关注实验报告格式,极大提高了写作效率。在线制作实验报告如图3所示,提交后实验报告效果图如图4所示。

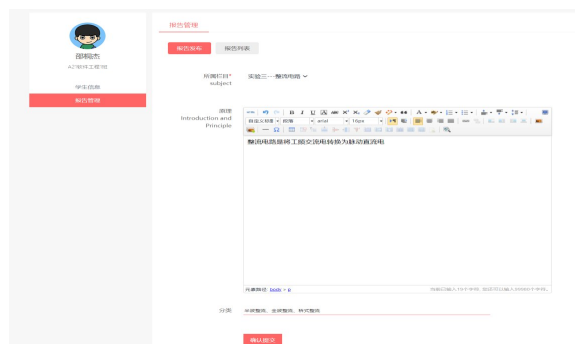


图3 在线制作实验报告



图4 实验报告预览效果

4.4 教师批改实验报告或自动阅卷

教师在报告管理中可按实验任务或其他查询条件,逐篇查看学生提交的实验报告。教师可以依据分类(已批改、未批改或全部)等查询条件筛选相应的实验报告,对尚未批改的实验报告进行批改,实验报告列表如图5所示,批阅实验报告的界面如图6所示。

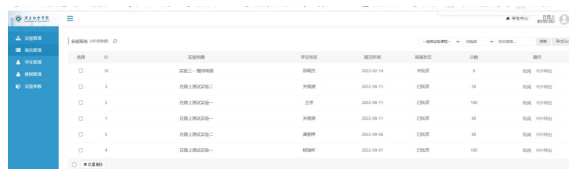


图5 教师查看并批改实验报告



图6 教师批阅实验报告

若教师在创建实验任务时开启自动阅卷功能,利用定时任务每隔一个小时执行一次自动阅卷功能的脚本。自动阅卷功能执行流程如下:

(1) 从数据实验参数表(名为 es_arctype)中取出当天所有未批阅的数据。

(2) 循环处理每一条实验报告记录,调用 Jieba 中文分词 cutForSearch 方法,将实验报告内容进行中文分词,由于分词后可能有重复,需要去重处理。

(3) 从数据库中取出当前循环实验报告任务对应的参考答案,每个参考答案分值平均处理。

(4) 将分词后去重的实验报告内容与参考答

案进行匹配，匹配上就累加分值，最终分值就为实验报告分数。

程序核心代码如下：

```
//引入jieba中文分词
require_once './vendor/fukuball/jieba-php/src/vendor/
multi-array/MultiArray.php';
require_once './vendor/fukuball/jieba-php/src/vendor/
multi-array/Factory/MultiArrayFactory.php';
require_once './vendor/fukuball/jieba-php/src/class/
Jieba.php';
require_once './vendor/fukuball/jieba-php/src/class/
Finalseg.php';
/**
 * 自动阅卷功能
 * Class AutoDecide
 * @package app\command
 */
class AutoDecide extends Command
{
    protected function configure()
    {
        $this->setName('atuodecide')->setDescription
            ('auto decide Lab report');
    }
    protected function execute(Input $input, Output
        $output)
    {
        ini_set('memory_limit', '1024M');
        Jieba::init();
        Finalseg::init();
        //1、获取当天所有需要进行自动阅卷的实验
        $arctypes = Db::name('arctype')
        ...
        $arctypeList = [];
        foreach ($arctypes as $k => $v) {
            $arctypeList[$v['id']][] = $v;
        }
        unset($arctypes);
        //从 archives 表中循环
        foreach ($arctypeList as $K => $v) {
            //取出关键字和实验字段 只处理当天的
            foreach ($v as $Kk => $vv) {
                ....
            }
        }
    }
}
```

```

}
unset($arctypeList);
$updateArr = []; //需要更新的内容
//循环处理每一条实验报告
foreach ($arctypeFormat as $K => $v) {
    foreach ($v as $kk => $vv) {
        unset($vv['aid']);
        $field_name = array_key_first($vv);
        $student_answer = array_values($vv);
        if (!isset($student_answer[0])) break;
        $student_answer = strip_tags(htmlspecialchars_decode($student_answer[0]));
        //参考答案分数平均分配
        $each_score = round($max_scoreArr[$K]
            [$kk] / count($key_wordsArr[$K]
            [$kk]), 1);
        $seg_list = Jieba::cutForSearch($student_
            answer, true);
        //分词后可能重复,需要去重
        $seg_list = array_unique($seg_list);
        $real_score = 0; //实际得分
        foreach ($seg_list as $seg) {
            if (in_array($seg, $key_wordsArr[$K]
                [$kk])) {
                $real_score += $each_score;
            }
        }
        $updateArr[$K][$field_name.'_score'] =
            floor($real_score);
    }
}
...
}
```

5 结语

笔者结合自己所在学校的开放实验教学改革创新思路与开放实验室的定位、功能和目标，以及实验管理的实际需求，充分借鉴、吸收和融合当前在线作业系统的优点，开发了一套具备学校特色的在线实验报告管理系统，本系统功能齐全，支持在线创建实验报告任务、自动阅卷、批量导出实验报告和分数等功能；学生在线制作实验报告并不需要关心报告格式；极大提高了教师和学生的效率，提升了实验教学效果。

参考文献:

- [1] 陈鸿光,唐亚敏,金彪. 编程类课程实验报告自动生成与批改方案研究[J]. 实验室科学,2017,20(3): 50-54.
- [2] 蔡燕敏. 高校开放式实验室管理平台的设计与实现[J]. 实验室科学与技术,2014,12(2):180-181.
- [3] 孟吉冬. 在线作业提交系统的设计与实现[D]. 长春:吉林大学,2014.
- [4] 陈颖,张福洪. 基于J2EE的实验报告管理系统的设计与实现[J]. 现代电子技术,2007,20:69-74.
- [5] 周骏. 基于.Net和Ajax技术的实验教学管理系统的设计和实现[D]. 江苏:苏州大学,2014.

Design and implementation of online experimental report management system

Ouyang Xibiao*, Xu Baolin, Jiang Jing

(School of Big Data and Computer, Guangdong Baiyun University, Guangzhou 510000, China)

Abstract: Based on the current experimental report is hard to save and manage, the typography styles is not uniform, and the review is too time-consuming, this paper starts from the actual situation of experimental teaching and management, develops a set of experimental report management system based on B/S architecture that meets the actual needs of the college. The system is divided into student and teacher model, the teacher model has the functions of creating experimental report tasks, online review, batch export of experimental reports and grades, etc., and the student model realizes remote submission, preview, export of experimental reports and other functions. The system simplifies the process, saves resources and management costs, and adds the automatic marking function, which greatly improves the work efficiency and effect.

Keywords: experimental report; online review; remote submission; automatic marking

(上接第 94 页)

A low-power Bluetooth MEMS vibration sensor and its application

Fang Yizhou^{1*}, Zhao Zhongrui¹, Zhang Qingye¹, Zhang Pengcheng¹, Aimaiti·Bulading¹,
Zeng Youxuan², Wang Jue²

(1. INESA R&D Center, Shanghai 200233, China;

2. Shanghai Nanyang Wanbang Software Technology Co., Ltd., Shanghai 200233, China)

Abstract: MEMS is short for micro-electro-mechanical systems and acceleration measurement is typical usage of MEMS chip. The micro-controller unit (MCU) chip can calculate the motion speed, direction angle, vibration and other information through the MEMS acceleration measurement data. After more than 20 years of development, Bluetooth technology has been iterated for 5 generations, from the early single peripheral communication chip application to the system-on-chip (SoC) usage, especially from the fourth generation of Bluetooth technology, which has evolved the low power consumption characteristics. Based on the research and application of MEMS acceleration chip and bluetooth SoC chip, a low power Bluetooth MEMS vibration sensor without additional MCU is designed, with the characteristics of low cost, low power consumption and miniaturization. Based on the processing capacity of bluetooth SoC chip and the requirements of the application scenarios, a MEMS acceleration data processing model is designed, and data aggregation and platform display are realized through the Bluetooth gateway, which can form a scene-oriented intelligent application.

Keywords: Bluetooth; MEMS acceleration chip; low power; data processing

文章编号: 1007-1423(2023)11-0109-04

DOI: 10.3969/j.issn.1007-1423.2023.11.022

电子校历自动生成系统的设计与实现

刘 哲*, 宋佳豪, 潘梦瑶

(延安大学数学与计算机科学学院, 延安 716000)

摘要: 基于 Python 中的工具包 xlwt+calendar, 结合实践教学管理工作中的需求, 设计并实现了一款电子校历自动生成系统, 能有效提升教学管理人员完成部分教学事务的工作效率。

关键词: 电子校历; 教学管理; Python; calendar

0 引言

高等学校各项工作的良好运转, 离不开广大教学管理人员的辛勤付出。然而面对教学工作过程中的复杂性、多样性, 大多数工作都是通过人工的方式去手动完成的。其中就包括了一部分内容枯燥、重复率高, 实际上又可以通过计算机程序自动完成的工作。教务人员每天都要花费大量的时间和精力去完成这些工作。因此, 如果能够通过计算机程序的方式, 辅助完成这类工作, 可以有效解决教学管理过程中存在的问题, 提升教学管理水平^[1-3]。而校历文件的制定, 是高校教学计划有序执行过程中必不可少的一环。如何能够将其中所蕴含的价值更加充分、高效地利用起来^[4-6], 具有非常重要的研究价值。

1 实验设计

本次实验使用计算机编程语言 Python 3.11 进行系统开发, 使用 Python 标准库 Calendar 获取日历相关数据, 使用 Python 的第三方库 xlwt 将组织好的数据内容写入目标文件中, 并对其进行排版。

1.1 主要组件介绍

Calendar 是 Python 中与日历信息相关的一个标准库, 该模块提供了丰富的对年、月、日等数据进行操作的函数, 使用时无需安装, 可以直接通过 import 语句导入。模块主要包括三大类: Calendar、TextCalendar 和 HTMLCalendar。其中, Calendar 是其它两个类的基类, 可以用来创建一个普通的日历对象。TextCalendar 一般用来创建一个纯文本的日历对象, 使用 HTMLCalendar 生成 HTML 格式的日历对象。

表 1 常用的国内镜像站

镜像源名称	访问地址
清华大学	https://pypi.tuna.tsinghua.edu.cn/simple
中国科技大学	https://mirrors.ustc.edu.cn/pypi/web/simple
豆瓣	https://pypi.douban.com/simple
阿里云	https://mirrors.aliyun.com/pypi/simple

xlwt 是 Python 中常见的用来生成 Excel 类型文件的第三方工具包, 使用时需要通过 pip 命令安装后才能导入, 推荐通过镜像源的方式进行安装^[7], 表 1 列出了一些质量较高的 Python 镜像

收稿日期: 2023-04-17 修稿日期: 2023-04-28

基金项目: 延安大学青年杰出人才项目(QJ001); 延安大学校级大学生创新创业训练计划项目(D2022038)

作者简介: *通信作者: 刘哲(1988—), 男, 陕西延安人, 硕士, 助教, 研究方向为大数据知识工程, E-mail: liu_zhe2023@163.com; 宋佳豪(2003—), 男, 陕西咸阳人, 在读本科; 潘梦瑶(2003—), 女, 陕西咸阳人, 在读本科

源。需要注意的是, xlwt 主要用于将组织好的数据和相关的文本格式信息写入新建的工作表文件中, 不能直接编辑一个已经存在的工作表文件。

1.2 系统设计

系统的基本生成步骤如图1所示。

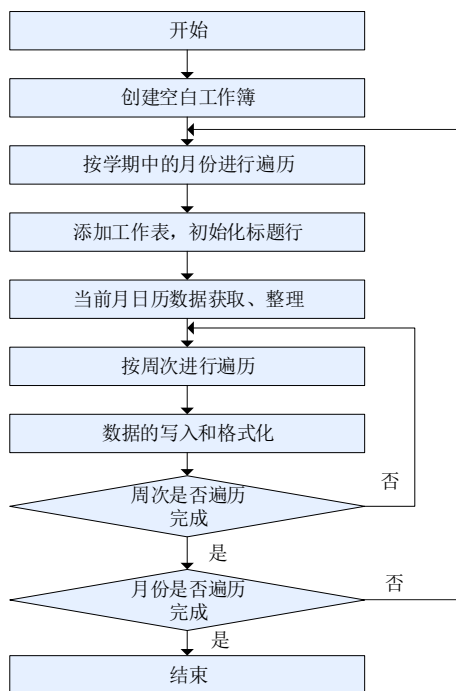


图1 系统流程

(1) 创建一个空白的 Excel 文档, 等待数据写入。

(2) 按学期中的月份进行遍历。逻辑上, 将当前学期的第一个月记为起始月 `start_month`, 将当前学期的最后一个月记为结束月 `end_month`, 使用函数 `range(start_month, end_month+1)` 遍历访问当前学期的月份。

(3) 添加工作表, 初始化标题行。例如, 2022—2023 学年第二学期的起始月是 2 月, 结束月是 7 月。那么, 需要按照从“2 月”到“7 月”的顺序和命名方式, 在工作簿中依次创建每张工作表, 用于存放对应的日历数据。每次迭代只创建当前月的日历工作表, 并初始化工作表的列标题行, 分别是周序号、一、二、……、六、日。其中, “周序号”一列下的数据, 取值范围是教学执行计划中的第一周到最后一周的

序号, 依次递增。

(4) 当前月日历数据的获取和整理。创建 Calendar 日历对象, 通过调用函数 `monthdayscalendar(year, month)`, 获取当前月的日历数据, 整理出相关的中间变量, 等待后续程序的使用。

(5) 按周次进行遍历。例如, 当前迭代月是 2023 年 3 月, 3 月 1 日对应星期三, 3 月 31 日对应星期五。那么, 3 月份在日历中会对应 5 行数据, 按周次遍历应遍历 5 次, 其它月份的遍历依此类推。

(6) 数据的写入和格式化。将当前周次的日历数据, 按照设定好的排版格式, 写入对应的单元格中。例如, 2023 年 3 月 1 日正处于开学后的第一周内。那么, 程序应该在对应列标题下的单元格中, 写入的数据依次是 [‘第 1 周’, ‘’, ‘’, ‘1’, ‘2’, ‘3’, ‘4’, ‘5’]。

(7) 构建剩余月份的日历工作表。重复步骤 (2)~(6)。

2 难点与解决方案

本部分选取了实验过程中遇到的一些难点问题, 给出解决方案, 总结如下。

问题一: 如何让日期出现在正确的列标题下。

如何确定每个月的日期数据, 与列标题的对应关系。例如, 如果当前月的第一天, 对应的列标题是周三。那么, 同一行中对应列标题是周一和周二的单元格, 应该是没有内容的。

通过查询 calendar 模块的官方文档, 以及不断的测试、对比, 最终挖掘出了模块中的函数 `monthdayscalendar(year, month)`。该函数通过传递实参年和月, 能够返回具有如下所示的日期格式数据。

`[[0, 0, 1, 2, 3, 4, 5], [6, ..., 12], [13, ..., 19], [20, ..., 26], [27, 28, 0, 0, 0, 0, 0]]`

示例中的数据是 2023 年 2 月的日历数据, 其中的数字 1~28 代表当前月中实际有的日期, 0 代表非当前月的日期数据。可以看出, 以周为计数单位, 当前月的日历数据(外层的大列表), 共包含 5 个完整的周(内层列表), 即外层的大列表一共有 5 个列表类型的元素。而每个内层列表中的元素, 按顺序依次对应周一、周二、……、

周日。通过这种方式，就可以建立日期数据与列标题之间的映射关系。当将日期数据写入工作表时，如果当前单元格对应的内容是数字0，则跳过。

问题二：教学周序号的确定。

高等院校教学计划的执行、工作安排的制定都是以周为单位进行的。程序自动构建时，如何正确地输出当前周次对应的教学周序号。

正常情况下的逻辑是，设定一个变量 week_index 代表周序号，将该变量的值初始化为1，并关联到开学第一周的日历数据。当每次输出完一周的日历数据时，让变量 week_index 的值加1，然后关联第二周的日历数据，以后的周序号依此类推。如果某个月的最后一天是星期二，对应教学执行计划的第5周。那么下个月的第一天会是星期三。但是，当输出下个月第一周的日历数据时，不能将教学周的序号自动加1。因为当前月的最后一天与下个月的第一天，属于同一个周。

解决办法是，当输出每个月第一周的日历数据时，判断其中是否包含非当前月的日历数据，例如 [0,0,1,2,3,4,5]。这种情况下，就认定教学周序号 week_index 的值不需要加1。

问题三：Excel 文件的行高和列宽的设置问题。

xlwt 工具包的官方帮助文档极少，开发的过程中相关功能的实现和函数的使用，主要来源于网络上的教程。但行高和列宽的设置方法，经过测试后均以失败告终。最终，通过阅读工具包的源代码结合之前的测试，行高和列宽的设置代码如表2所示。两个乘号左边的数字是基本尺寸单位，行高和列宽的设置，主要通过修改乘号右边的数字调试后确定。

表2 行高列宽的设置方法

设置类型	设置方法
行高	work_sheet.row(i).height_mismatch = True work_sheet.row(i).height = 20 * 50
列宽	work_sheet.col(i).width = 256 * 15

3 实验结果与分析

本次实验主要从两个方面对程序自动构建的电子校历文件进行验证：①日期数据是否正确

地显示在对应的列标题下；②教学周的序号与日期数据的关联是否正确。为此，调阅了2020—2021、2021—2022、2022—2023三个学年的校历文件进行对比。

实验结果表明：通过系统自动生成的不同学年、学期的电子校历文件，与相应的纸质校历中的数据完全匹配，并且构建过程不超过2秒。电子校历文件能有效提升教务人员完成特定事务的工作效率，将教务人员从枯燥的行政工作中解放出来，增加幸福指数。其应用场景广泛多样，例如，每个学期开始时都必须尽快制作的学期实验教师值班安排表，就可以使用电子校历文件的方式来快速完成，如图2所示。亦可以将电子校历当作备忘录来使用，记录教学执行计划中相应节点的工作事项。需要时可以方便地查阅以前的工作事项记录，制定新的教学安排。相比于传统纯文本形式的记录，电子校历更加美观大方、一目了然，易于事后翻阅和文件存储。

5 月份实验人员值班表（分组如下）

	一	二	三	四	五	六	日
第10周	1	2	3	4	5	6	7
	①	②	③	①	②		
第11周	8	9	10	11	12	13	14
	③	①	②	③	①		
第12周	15	16	17	18	19	20	21
	②	③	①	②	③		
第13周	22	23	24	25	26	27	28
	①	②	③	①	②		
第14周	29	30	31				
	③	①	②				

图2 电子校历形式的值班安排表

4 结语

电子校历自动生成系统在充分调研的基础上，结合教学管理工作中的实际需求进行研究，具有较强的应用价值。在提升工作效率的同时，将教务人员从枯燥的行政工作中解放出来，减轻工作压力，增加幸福指数。

参考文献：

[1] 赵宇艳. 高职院校培训教学管理中存在的问题与对

- 策[J]. 山西财经大学学报, 2022, 44(S2): 215-217.
- [2] 姜雪. 地方高校研究生教学秘书工作困境及其对策[J]. 继续教育研究, 2016(10): 107-108.
- [3] 刘晶. 高校教学秘书工作规范化探析[J]. 黑龙江高教研究, 2016(2): 75-76.
- [4] 宋文迪. 基于北邮校历的文创设计与研究[D]. 北京: 北京邮电大学, 2020.
- [5] 孙立炜. 一种基于校历的春运铁路客流量预测方法[J]. 中国高新科技, 2021, (14): 119-120.
- [6] 高瞻, 贾向炜, 李雪琴, 等. 基于轻量级整合框架的电子校历系统[J]. 电脑知识与技术, 2015, 11(27): 56-58.
- [7] 刘哲, 马乐荣. 大规模电商平台商品信息采集系统的设计与实现[J]. 延安大学学报(自然科学版), 2020, 39(2): 32-36, 42.

Design and implementation of electronic calendar automatic generation system

Liu Zhe*, Song Jiahao, Pan Mengyao

(School of Mathematics and Computer Science, Yan'an University, Yan'an 716000, China)

Abstract: Based on the toolkit xlwt+calendar in Python, combined with the needs of actual teaching management work, a system for automatically generating electronic school calendars has been designed and implemented, which can effectively improve the efficiency of teaching management personnel in completing some teaching affairs.

Keywords: electronic calendar; teaching management; Python; calendar

(上接第 98 页)

Design and implementation of Android-based anti-wandering APP

Zhou Man¹, Zhao Jiaying², Wen Yixian^{3*}

(1. School of Foreign Languages, Hunan Institute of Technology, Hengyang 421002, China;

2. School of Management, Xiamen University, Xiamen 361005; China;

3. School of Business, Hunan Institute of Technology, Hengyang 421002, China)

Abstract: When there is population movement, there is a hidden danger of people getting lost, which is a pain point for people's livelihood. With a smartphone and a pop-up, technology has redefined the way people are found, turning the "needle in a haystack" into an "accurate search". It has become necessary to develop a comprehensive, convenient and simple crowd finder APP. The existing apps have a single function and do not fully reflect the characteristics of Internet Finder Crowds, so it is important to develop an Android-based crowd finder APP. It uses Visual Studio Code for the front-end of the platform, calls Google Map API to achieve accurate positioning, and uses MySQL database for information storage, and adopts Spring Boot framework for back-end development to build a truly efficient and convenient crowd finder APP.

Keywords: APP; lost people; crowd finder

文章编号: 1007-1423(2023)11-0113-04

DOI: 10.3969/j.issn.1007-1423.2023.11.023

面向汽车旧件再制造组合赋权模糊综合评价模型

王华政*

(西南交通大学制造业产业链协同与信息化支撑技术四川省重点实验室, 成都 611756)

摘要: 围绕汽车产业价值链维修服务过程中产生的汽车旧件是否可再制造的模糊性决策问题, 经过对再制造决策问题的分析与主客观赋权评价算法的研究, 给出基于主客观组合权重的模糊综合评价方法, 对旧件相关指标计算, 评价旧件是否可再制造, 辅助供应商进行回收再制造决策, 并实际应用验证该方法的有效性。

关键词: 汽车旧件; 组合赋权; 模糊综合评价

0 引言

2021 年国务院印发了《关于印发汽车产品生产责任延伸试点实验方案的通知》^[1], 其中明确提出了对汽车产品实施生产责任延伸制度, 鼓励生产企业建立逆向回收利用体系, 支持回收报废汽车, 推广再制造产品。汽车售后服务行业是汽车产业价值链中的关键环节, 在售后服务中存在如此现象: 大量的汽车旧件(因维修而被更换的零部件)成为三包期内索赔的对象被零件供应商回收, 未被再制造利用, 甚至放弃旧件回收再制造, 导致大多数旧件未能发挥其循环利用的潜在价值。然而汽车旧件是汽车零部件再制造的重要原材料, 有效地判断出旧件是否可再制造, 决定着零件供应商是否回收该旧件、资金分配及原材料采购决策。针对此类问题, 研究旧件再制造决策评价模型, 辅助供应商进行旧件回收再制造决策具有一定的现实意义。

1 相关研究

目前国内大部分相关论文对可再制造回收确立的评价指标主要包括可再制造技术可行性、

经济可行性、环境可行性、资源利用率、能量损失率、质量等体系指标。文献[2]关于评价模型文献针对报废汽车拆解前的成本评估问题, 建立了一种基于拆解难度和模糊聚类的一般性综合成本评价模型。文献[3]提出了一种基于两级模糊计算方法的决策支持系统, 该方法结合了产品的定性和定量属性, 考虑技术、经济、资源利用和环境四个指标来确定产品的可再制造性。文献[4]建立一种基于实例推理的废旧零/部件可再制造性评价模型。

2 模型构建

本文选取主客观赋权平衡权重融合的方法进行指标赋权。由于各评价指标数据的计量单位不一致, 存在定性与定量的数据, 故在评价赋权时需进行数据归一化处理, 消除数据的差异性。本文对于评价对象及数据的归一化公式如下:

用 x 表示评价对象, $x = \{x_1, x_2, \dots, x_n\}$, 其中 x_i 表示第 i 个对象, $i \in \{1, \dots, n\}$, 这里 n 是评价对象的个数。 x_i 的第 j 个二级评价指标表示为 x_{ij} , 数据同向归一化公式如下:

收稿日期: 2023-02-18 修稿日期: 2023-03-04

作者简介: *通信作者: 王华政(1997—), 男, 四川广安人, 硕士研究生, 研究方向为计算机应用技术, E-mail: wanghuazheng97@163.com

$$\begin{cases} x_{ij} = \frac{x_{ij} - \min\{x_{1j}, x_{2j}, \dots, x_{nj}\}}{\max\{x_{1j}, x_{2j}, \dots, x_{nj}\} - \min\{x_{1j}, x_{2j}, \dots, x_{nj}\}}, \text{正向指标} \\ x_{ij} = \frac{\max\{x_{1j}, x_{2j}, \dots, x_{nj}\} - x_{ij}}{\max\{x_{1j}, x_{2j}, \dots, x_{nj}\} - \min\{x_{1j}, x_{2j}, \dots, x_{nj}\}}, \text{负向指标} \end{cases} \quad (1)$$

旧件的评价结果集定义为 $T = \{T_1, T_2, T_3\}$, T_1 表示无用旧件, T_2 表示可再利用旧件, T_3 表示可修复旧件。无用即无利用价值的旧件、可再利用即还有使用价值, 可拆卸使用作为原材料; 可修复指经过加工后, 可恢复原有的性能。

2.1 指标选取

评价模型的可行性、可靠性及评价结果的有效性很大程度上依赖评价指标体系的建立。评价指标的选取依赖于评价对象的刻画目的, 本文参考其他学者在旧件可再制造评价上确定的旧件指标, 并依据第三方云平台长期以来服务于汽车产业价值链集聚的旧件相关数据, 从技术可行性、经济可行性、环境可行性等方面去刻画旧件是否可再制造, 将旧件分为可修复、可再利用、无用等三类。各项指标维度刻画如表1所示。

表1 汽车旧件再制造评价指标

一级指标	二级指标
经济可行性	1.再制造成本; 2.再制造价格; 3.剩余价值
技术可行性	1.可拆卸装配性; 2.可再加工性; 3.损坏程度
环境可行性	1.污染指数

2.2 指标赋权

对于赋权方法的选择, 本文选取主观赋权AHP层次分析法与客观赋权熵权法。用 W 表示权重矩阵 $W = [w_1 \dots w_n]$ w_i 表示二级指标的权重值, n 是二级指标的个数。

AHP^[5]是一种定性和定量相结合的、系统化的、层次化的分析方法。将一个复杂的目标决策问题作为一个系统, 将目标分解为多个目标或准则, 由专家通过指标相对的重要程度给出判断矩阵, 计算指标权重, 用一致性指标和一致性比率衡量判断矩阵的有效性, 由AHP得到的权重为 W_1 。

熵权法^[6]是根据指标信息熵的大小对指标

客观赋值的一种方法, 对数据进行同向化归一化处理, 再由信息熵公式进行计算, 信息熵越小, 代表该指标的离散程度越大, 包含的信息就越多, 所赋予的权重就越大。由熵权法得到的权重为 W_2 。

2.3 指标平衡权重组合

关于权重组合, 本文采用权重差异度^[7]的思想来平衡组合权重。

第1步: 计算主客观总体权重得分, 取主观权重为 W_1 , 客观权重为 W_2 , 取主观权重的平衡因子为 θ_1 , 客观权重的平衡因子为 θ_2 , 平衡权重为 W , W 计算如式如下, 其中 $\theta_1 \geq 0, \theta_2 \geq 0, \theta_1 + \theta_2 = 1$

$$W = \theta_1 * W_1 + \theta_2 * W_2 \quad (2)$$

第2步: 权重差异度为 $\sum_{i=1}^m (\theta_1 * W_1 \times x_i - \theta_2 * W_2 \times x_i)^2$, 其中 x_i 是评价数据的第 i 个数据, 为了降低主客观权重对指标赋权的差异性, 建立以下目标约束问题

$$\begin{aligned} \min & \sum_{i=1}^m (\theta_1 x_i W_1^T - \theta_2 x_i W_2^T)^2 \\ \text{s.t.} & \begin{cases} \theta_1 \geq 0, \theta_2 \geq 0; \\ \theta_1 + \theta_2 = 1; \end{cases} \end{aligned} \quad (3)$$

第3步: 代入优化目标 $\theta_1 = 1 - \theta_2$, 求 θ_1 导数:

$$\begin{aligned} \dot{\theta}_1 = & 2 \sum_{i=1}^m (\theta_1 x_i \times W_1^T - (1 - \theta_1) x_i \times W_2^T) * (x_i \times W_1^T + x_i \times W_2^T) \end{aligned} \quad (4)$$

令 $\dot{\theta}_1 = 0$ 在 $0 \leq \theta_1 \leq 1$ 内求解 θ_1 , $\theta_2 = 1 - \theta_1$ 即可。

2.4 模糊评价

第1步: 根据上面的值求解, 评价数据个体一级指标维度的值及一级指标权重值分别为式(5), 式(6), 其中 m 是各一级指标下的二级指标个数。

$$S_i^{\text{economy}}, S_i^{\text{technology}}, S_i^{\text{environment}} = \sum_{j=1}^m (W_j * x_{ij}) / \sum_{j=1}^m W_j \quad (5)$$

$$W^{\text{economy}}, W^{\text{technology}}, W^{\text{environment}} = \sum_{j=1}^m W_j \quad (6)$$

第2步: 根据式(5)得到一级指标赋权得分,

根据专家的经验判断,给出一级指标对评价结果集的隶属矩阵 R ,如 $R^{\text{economy}} = [a_1 \ a_2 \ a_3]$ 表示第一级指标经济可行性对评价结果集 O 的隶属度。一级指标对评价结果集的隶属度梯度函数如下:

$$\begin{cases} R_1^i = \begin{cases} 1, x \leq a_1 \\ \frac{a_2 - x}{a_2 - a_1}, a_1 < x < a_2; \\ 0, x \geq a_2 \end{cases} \\ R_2^i = \begin{cases} \frac{x - a_1}{a_2 - a_1}, a_1 < x \leq a_2 \\ \frac{a_3 - x}{a_3 - a_2}, a_2 < x < a_3; \\ 0, x \geq a_3 \vee x \leq a_1 \end{cases} \\ R_3^i = \begin{cases} 0, x \leq a_2 \\ \frac{x - a_3}{a_3 - a_2}, a_2 < x < a_3 \\ 1, x \geq a_3 \end{cases} \end{cases} \quad (7)$$

第3步:通过隶属函数计算评价对象对于评价结果集的隶属矩阵为 $\hat{R} = [R_1^i \ R_2^i \ R_3^i]$,这里 i 表示一级指标。评价对象的隶属度得分为 T ,如式(7),选择 T 中最大的值对应的评价即是该评价对象的评价结果。

$$T = [W^{\text{economy}}, W^{\text{technology}}, W^{\text{environment}}] \hat{R} \quad (8)$$

3 实验分析

3.1 实验数据

本文算法实验数据源于第三方云平台数据空间及相关网站(环境方面是The Eco-indicator 99及其生态数据库^[8]获取旧件在加工制造运输等过程中对环境影响的相对负荷数据),进行数据预处理及按式(1)同向化处理,其中再制造成本、损坏程度与污染指数为负向指标,其他均为正向。

3.2 实验结果分析

通过上述方法计算了AHP、EWM及AHP-EWM组合的权重,得到的权重对比图如图1所示,AHP-EEM组合权重结合了主观经验判断和数据客观的熵权信息。

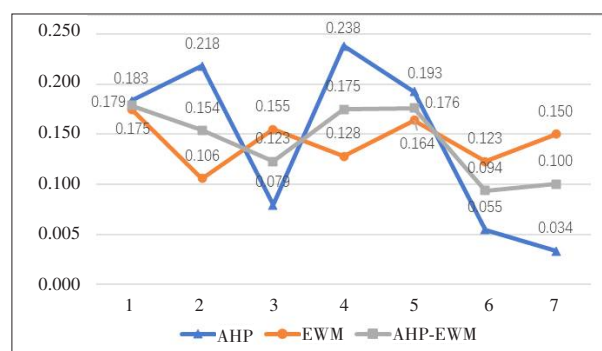
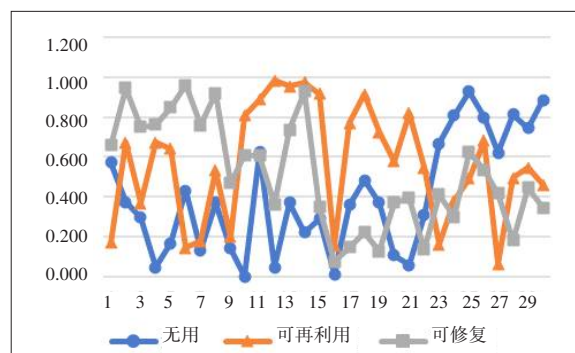
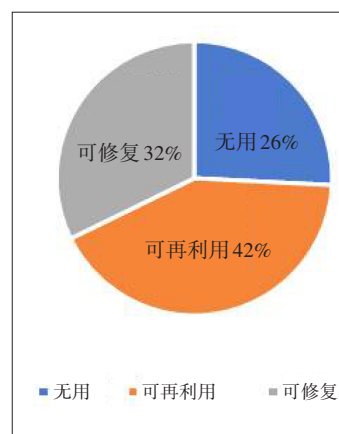


图1 权重对比

根据模糊评价的方法计算AHP-EWM组合赋权得到模糊评价结果,评价对象对于三种评价结果得分与整体的评价结果如图2所示。在实际的旧件再制造决策通过客观情况分析与实际厂商的实际应用沟通,组合赋权得到的评价结果在辅助决策中效果更佳,综合考虑了数据的客观关系情况与厂商的主观意愿。



(a)



(b)

图2 评价结果

4 结语

本文以汽车产业价值链旧件管控系统回收再制造辅助决策的需求为基础,分析其所需的旧件再制造决策评价模型与国内外相关的研究情况,设计了基于AHP-EWM平衡组合赋权的模糊综合评价方法,用于供应商回收再制造辅助决策。实验结果证明,该模型较单一赋权模型的评价效果更佳,为汽车旧件再制造决策评价提供了有效可行的解决方案。

参考文献:

- [1] 工业和信息化部,科技部,财政部,等. 关于印发汽车产品生产责任延伸试点实验方案的通知:工信部联节函[2021]129号[A/OL]. (2021-05-26). http://www.gov.cn/zhengce/zhengceku/2021-06/10/content_5616601.htm
- [2] 张春亮,陈铭. 基于拆解难度和模糊聚类的泛化报废汽车拆解成本预测[J]. 中国机械工程,2019,30(7):840-845.
- [3] OMWANDO T A, OTIENO W A, FARAHANI S, et al. A bi-level fuzzy analytical decision support tool for assessing product remanufacturability [J]. Journal of Cleaner Production, 2018(174): 1534-1549.
- [4] 蒋小利,江志刚,张华,等. 基于实例推理的废旧零部件可再制造性评价模型及应用[J]. 现代制造工程,2013(12):6-9.
- [5] 吕康,叶军,李天淳. 基于分辨矩阵与层次分析法的组合属性权重决策方法[J]. 河北师范大学学报(自然科学版),2022,46(2):126-138.
- [6] 程宏,江志刚,向鹏,等. 基于熵权和TOPSIS法的再制造逆向物流模式决策研究[J]. 制造技术与机床,2016(8):71-76.
- [7] 任春华,孙林夫. 面向协作企业群的三阶段主客观公平权重动态评价[J]. 计算机集成制造系统,2019,25(11):2874-2891.
- [8] 谷振宇,刘飞,任凡. 基于生态指示器99的制造过程环境影响分析模型[J]. 中国机械工程,2010,21(8):931-935.

Weighting evaluation model for remanufacturing combination of used automobile parts

Wang Huazheng*

(Sichuan Key Laboratory of Manufacturing Industry Chain Collaboration and Information Support Technology, Southwest Jiaotong University, Chengdu 611756, China)

Abstract: Based on the research of subjective and objective weighting algorithm and evaluation algorithm, a fuzzy comprehensive evaluation method based on subjective and objective weight combination is proposed to solve the problem of fuzzy decision on remanufacturability of used auto parts generated in the process of maintenance service of auto value chain. By calculating relevant indexes of old auto parts, the decision on remanufacturability of old auto parts is evaluated. The effectiveness of this method is verified by practical application.

Keywords: automobile used parts; combinatorial empowerment; fuzzy comprehensive evaluation

文章编号: 1007-1423(2023)11-0117-04

DOI: 10.3969/j.issn.1007-1423.2023.11.024

一种基于 YOLO 的口罩佩戴检测终端的设计

张 良*, 陈周欢, 林耿名, 叶绮雯, 何镇轩

(韶关学院信息工程学院, 韶关 512005)

摘要: 对新冠疫情现状的调查, 依托 RT-thread 平台设计一款用于公共场所进出口对口罩佩戴的检测和提醒(未佩戴口罩时)终端。通过植入推理模型, 使终端在脱机状态下正常工作。硬件平台由主板 ART-PI 以及多媒体拓展板和 GC0328 摄像头模块构成。选用 YOLOv5s 和 YOLO-Fastest 网络模型分别应用于联机状态和脱机状态。经过训练的 YOLO 目标检测模型, 对同一组 100 张图片测试, 在联机状态下准确率达到了 92.3%, 在脱机状态下达到了 74.9%。

关键词: RT-Thread; YOLO 目标检测; 口罩识别

0 引言

自 2019 年末, 新冠病毒(COVID-19)已席卷世界各个角落。发展至今, 随着病毒毒性减弱, 社会面全面放开, 随之而来的便是大面积的感染。根据实际情况和相关研究, 个体仍有二次感染病毒的风险。因此群众对个人的防护措施依然不可松懈, 公众场合、私密场合佩戴口罩依然是最有效的防护措施之一。因此设计一款轻量的口罩佩戴检测终端, 既可满足有联机条件的工作环境, 也能满足脱机使用条件的可行性设计方案。

1 整体方案

系统由图像信息采集模块、图像显示模块、数据处理模块、数据传输模块、报警模块构成。本次设计使用 ART-PI 开发板作为硬件平台, 为数据处理模块以及数据传输模块, 其具有 32 M 静态 RAM, 能更好地适应推理模型的植入。图像信息采集模块使用 GC0328 摄像头模块来捕获视频图像信息, 用于进一步的口罩识别处理。多媒体拓展版上的 3.5 寸 LCD 显示屏用于显示被

摄像头捕捉的图像经推理模型识别后的结果。每当捕捉到行人图像, 经模型推理后, 对佩戴口罩的对象使用绿色框框出, 对未佩戴口罩的对象使用红色框框出, 并通过多媒体板的音频输出提醒行人佩戴口罩。此外, 在联机条件下, 将由上位 PC 担任数据处理模块, 来完成对图像中行人佩戴口罩的识别, 通过串口与 ART-PI 主板进行数据传输。

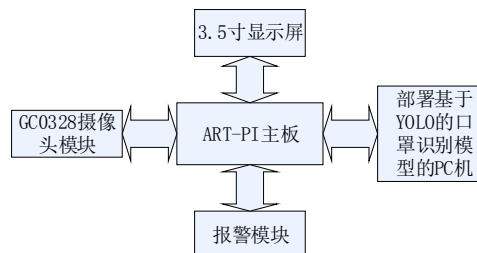


图 1 方案设计框图

2 硬件设计

2.1 主控平台 ART-PI

RT-Thread 作为一款国产嵌入式 IoT OS, 与

收稿日期: 2023-02-23 修稿日期: 2023-03-13

基金项目: 省级大学生创新创业训练计划项目资助(S202210576059)

作者简介: *通信作者: 张良(2001—), 男, 浙江绍兴人, 在读本科, 研究方向为嵌入式系统, E-mail: 1772249421@qq.com; 陈周欢(2002—), 男, 浙江衢州人, 在读本科, 研究方向计算机软件及计算机应用; 林耿名(2001—), 男, 广东湛江人, 在读本科, 研究方向为嵌入式开发; 叶绮雯(2001—), 女, 广东河源人, 在读本科, 研究方向为嵌入式系统; 何镇轩(2002—), 男, 广东广州人, 在读本科, 研究方向为人工智能图像识别、神经网络预测

其他 RTOS (FreeRTOS、uC/OS 等) 的主要区别是: RT-Thread 除了嵌入式系统的内核, 还包含网络、文件系统、GUI 界面等中间层组件, 具有极强的扩展性^[1]。ART-PI 为 RT-Thread 团队设计的一款硬件平台。在 RT-Thread 系统下拥有更好的贴合度和使用性。搭载 stm32H7 系列高性能芯片, 同时板载 Wi-Fi、蓝牙、TF 卡槽、预留摄像头、多媒体模块等各类外设接口。因此在依托 RT-Thread 平台下选择 ART-PI 作为硬件平台。

2.2 ART-PI 多媒体拓展板

多媒体拓展板为 ART-PI 官方配套拓展板, 其搭配一块 3.5 寸电容 LCD 触摸屏, WM8988 音频输入输出接口, 预留 GC0328/OV2640 摄像头排线接口。整体硬件与 ART-PI 主板组成结构简约、美观。

2.3 RT-AK

RT-AK 为 RT-Thread 团队对实时操作系统开发的 AI 套件, 其具有将 AI 模型部署进工程的能力, 提供 AI 模型推理 API, 在目标硬件平台下获得对 AI 模型极致性能推理的能力。RT-AK 简单架构如图 2 所示。

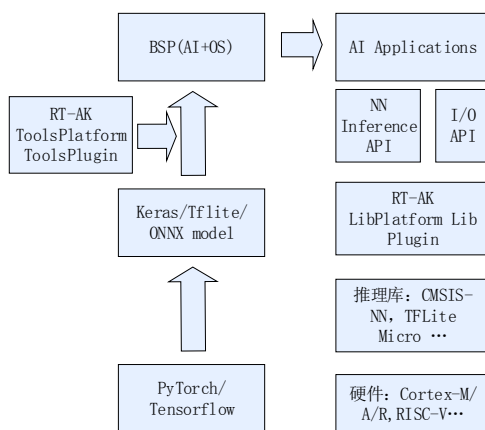


图 2 RT-AK 架构

首先在上位机通过 Tensorflow 等框架训练得到 AI 模型文件, 将其转化为 Keras、TfLite 模型格式, 再通过 RT-AK Tools 针对不同硬件平台进行适配得到集成了 AI 的 BSP。最后通过 RT-AK Lib 在嵌入式端使用 AI 模型, 完成 AI 模型的推

理运行。在嵌入式端对 RT-AK Lib 的使用流程包括: 查找模型句柄、初始化模型、数据输入, 得到推理结果如图 3 所示。

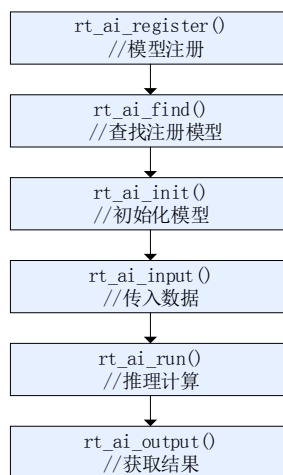


图 3 嵌入式端推理流程

3 软件设计

3.1 YOLO 目标检测

YOLO 网络模型是目标检测算法中应用广泛的模型之一, 相比于两阶段检测算法, YOLO 所使用的单阶段检测算法运行速率更快, 但会损失一定精度, 对于当前硬件算力有限的条件下, 选择 YOLO 目标检测算法更为合适。目前其具有八个官方版本, 同时拥有若干个经二次修改的网络用于特殊的针对性场景。本次设计针对联机和脱机情况, 对不同平台选用不同的版本。对联机情况, 使用算力更强的 PC 机选用官方第五版本中的 YOLOv5s^[2]; 对脱机状态, 硬件平台选用 YOLO-Fastest^[3]模型, 为当前结构最小、算力依赖最低的 YOLO 网络模型。图像目标检测的重要评价指标主要有平均精度 (average precision, AP) 和平均精度均值 (mean average precision, mAP)^[4]。选用网络模型概况见表 1。YOLO-Fastest 模型相比较 YOLOv5s 来说, mAP 减小近 56%, 但模型大小为 YOLOv5s 的 5%, 浮点计算量也仅为 YOLOv5s 的 1.5%。边缘硬件的算力显然不如上位 PC 机的 CPU, 甚至是 GPU 算力, 损失一定精度来换取更小的运算量是合适的。

表1 模型概况

Model	Input_size	Flops /G	Params /M	Size /M	*mAP0.5 /%
YOLOv5s	640×640	16.5	7.23	14.0	56.8
YOLO-Fastest	320×320	0.25	0.35	1.4	24.4

*mAP on COCO 2017val

3.2 模型训练

训练数据集来自 kaggle 的开源数据集，进行格式的转化后得到 xml 格式和 txt 格式。针对两种模型的训练，应用到了 Tensorflow^[5] 和 Darknet^[6] 两种框架进行训练和量化。此外 RT-AK Tools 对于模型文件的类型也有一定的规定，在部署 AI 模型前需要将 Darknet 框架训练得到的 x.weight 量化为 x.tflite 模型类型。表2为训练结构。这里在 Darknet 框架下的 max_batches 不同于 tensorflow 框架下的 epoch。batch 和 epoch 都为超参数。一个 epoch 是对整个数据集做一次迭代，一次迭代中包含若干次训练，而一次训练的样本大小则为 batch 的值，max_batches 则为最大训练容量，也称之为最大迭代次数，在某种程度上 max_batches 和 epoch 是可以互相换算的。计算方式为：

$$\max_batches = \text{数据集样本数} \times \text{epoch} / \text{batch}$$

在 Darknet 框架中，cfg 配置文件仅使用到 max_batches，这里仍做保留。

表2 训练结构

目标平台	训练框架	Classes	Epoch /max_batches	Input-size
PC	Tensorflow	With-mask、 Without-mask	300	640×640
ART-PI	Darknet	With-mask、 Without-mask	100000	160×160

3.3 系统流程设计

系统开机后，判断是否处在联机状态，联机状态下将图片推理部分交由 PC 来处理，若在脱机状态下将由设备自己完成图片的检测。YOLO 目标检测算法，粗略来说是对若干个 bounding box^[7] 进行预测，每个 box 的预测结果将由 x 、 y 、 w 、 h 、 c 、 $class_1$ 、 $class_2$ 、 $class_3$ …… 构

成，对于口罩佩戴识别可以认为二分类检测（佩戴口罩和未佩戴口罩），每条输出结果包含七个属性。该 box 所属类的置信度 $X=c*class_1(class_2)$ ，当某 X 大于预先设置的阈值时，认为该 box 内容属于该类。图4为系统流程。

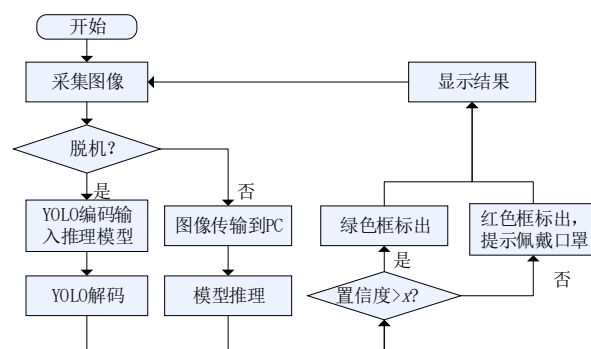


图4 系统流程

4 系统测试

4.1 模型测试

针对模型训练过程中的训练集和测试集，在模型训练完成后，使用该推理模型对数据集进行抽测。PC 和 ART-PI 两个不同目标平台下的模型测试效果分别如图5、图6所示。

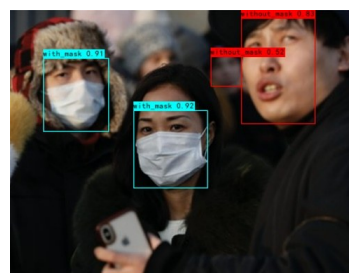


图5 YOLOv5s测试



图6 YOLO-Fastest测试

经对 100 张图片，包含约 350 个对象进行测试后，结果见表3。

表3 测试结果

测试框架	GPU	Models	准确率/%	用时/s
Tensorflow	RTX2060	YOLOv5s	92.3	7.5
Darknet	RTX2060	YOLO-Fastest	74.9	2.4

4.2 脱机测试

本项测试针对实际情况进行实际的脱机测试,在脱机状态下,手持硬件摄像头对准需要识别的图片进行测试,摄像头捕捉现实图像后,在硬件侧进行模型推理,最后在LCD屏上显示推理结果。整体延迟较低,效果同PC机测试YOLO-Fastest模型时基本一致。这里仍以模型测试中展示图片进行展示比对,如图7所示。



图7 终端脱机测试

5 结语

本设计依托RT-Thread平台设计的口罩佩戴检测终端,实现设备在联机和脱机两种状态下的正常工作。经实际测试,联机状态下的识别率极高,在人流较大的情况下依然可以精准检

测。脱机状态下,由于模型较小,相对来说存在一定误差,相对于PC的算力也相对较弱,运算速度较慢,能满足一般需求。另一方面,如何对YOLO-Fastest网络模型进一步剪枝网络体量,优化网络结构,来提高在边缘硬件侧的推理效率和准确度是下一步的研究内容。本系统具有实际实施成本较低、使用方便、应用场景广等特点。可以应用于商场出入口,小区出入口等公共场所,提醒往来人群及时佩戴口罩,做好个人防护。

参考文献:

- [1] 张晓芳. 基于RT-Thread的室内环境监测系统的设计[D]. 新乡:河南师范大学,2021.
- [2] 葛云飞,祁云嵩,孟祥宇. YOLOv5改进的轻量级口罩人脸检测[J/OL]. 计算机系统应用: 1-7 [2023-02-13]. DOI:10.15888/j.cnki.csa.009021.
- [3] 蒋超,张豪,章恩泽,等. 基于改进YOLOv5s的行人车辆目标检测算法[J]. 扬州大学学报(自然科学版),2022,25(6):45-49.
- [4] 沈立峰. 基于边缘计算和深度学习的铁路防入侵关键技术研究[D]. 长春:长春理工大学,2022.
- [5] POINTER I. TensorFlow, PyTorch, and JAX: choosing a deep learning framework [EB/OL]. (2022- 08-29). <https://www.infoworld.com/article/3670114/tensorflow-pytorch-and-jax-choosing-a-deep-learning-framework.html>.
- [6] 刘瑶. 改进Darknet框架的多目标检测与识别方法研究[D]. 西安:西安工程大学,2019.
- [7] 周伦钢,孙怡峰,王坤,等. 目标多种多值属性的终端快速识别网络[J]. 计算机工程与应用,2021,57(9):182-190.

Design of a mask wearing detection terminal based on YOLO

Zhang Liang*, Chen Zhouhuan, Lin Gengmin, Ye Qiwen, He zhenxuan

(School of Information Engineering, Shaoguan University, Shaoguan 512005, China)

Abstract: Investigate the current situation of the COVID-19 epidemic, relying on the RT-thread platform to design a terminal for detecting and reminding the wearing of masks at the entrance and exit of public places (when not wearing masks). By embedding the inference model, the terminal can work normally in offline state. The hardware platform is composed of main board ART-PI, multimedia expansion board and GC0328 camera module. YOLOv5s and YOLO-Fastest network models are applied to online and offline status respectively. The trained YOLO target detection model tested 100 images of the same group, and the accuracy rate reached 92.3% in online state and 74.9% in offline state.

Keywords: RT-Thread; YOLO object detection; mask identification

