# 软件计划文档

## 1.1系统概述

小云鹰平台作为网页式服务平台，通过遥感农业大数据对农村客户的生产活动和行为，进行面向自然灾害风险的金融化定价，部分替换和提升风险定价过程中的人力勘察、调研、核算过程，减少金融机构发放贷款时的人力成本，同时在贷后自动监测与预警方面实现自动化。平台业务逻辑如下图1-1。

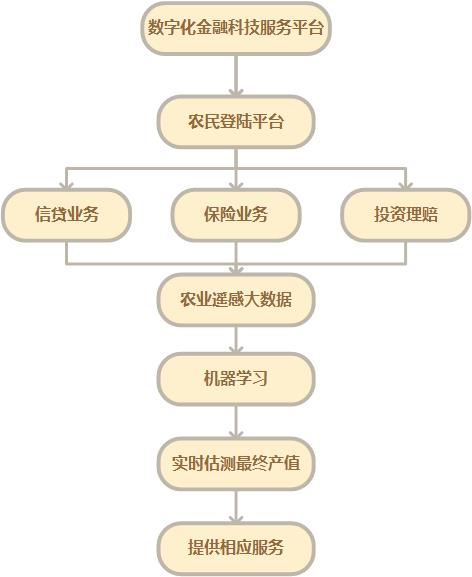


图1-1 平台业务逻辑图

小云鹰平台将利用遥感农业大数据监测审批，对实时收集到的气温、降水等可以观测到的指标进行整理、分析、核算，更快速地完成灾害受损状况监测与辅助定损，进一步提高验标、查勘定损的真实性与准确性，持续跟踪以完善贷前审批、贷后监测工作，从而拓宽金融业务范围，同时响应国家号召，实现多赢。

小云鹰数字化农业金融科技服务平台系统架构图如下图1-2。



图1-2 平台系统架构图

有如下创新点：

1.对开源遥感卫星图片和气象进行清洗和筛选，计算用户农田面积，对种植作物进行产量预估。通过大数据等数字科技手段收集农民种植地区各种遥感信息，结合深度学习人工神经网络模型对土地产量和农民收入进行精准预测。

2.结合往年和当期期货市场农产品价格，通过机器学习联合风控模型辅助各类金融机构为家庭农场客户基于作物收益完成生态农业生产环节的风险定价，实现信贷额度、利率和还款周期的合理匹配，提高金融机构对农业转向支持贷款的扶持力度。

### 1.1.1发展战略和经营目标

社会责任目标：以科技化方式替代人工实现高准确率的风险定价的金融服务，为农民贷款、灾害保险提供保障。

经营战略措施：吸引农民大量入驻平台，形成规模优势，同时将人数规模优势转化为生态农业生产资源的规模化数据，并通过平台的新算法、新商业/服务模型的应用，实现数据的“资产化”，助力各类金融机构在内循环经济领域发挥支柱性的核心作用。

### 1.1.2 拓展业务

小云鹰运营部门通过微信公众号、抖音、微博等自媒体渠道进行宣传，积极与政府开展合作积累流量，宣传当地的旅游业、特色农产品等，带动乡村地区经济发展。

发展助农商城电商业务，鼓励农户售卖绿色有机农产品，为解决农村农业存在的售出难、收购难、风险大等问题提供解决方案，拉进原产地与消费者的距离，助力新零售。

## 1.2 软件开发过程模型

|  |  |  |  |
| --- | --- | --- | --- |
| **软件生命周期** | **是否首先定义好所有的需求** | **是否有多个开发周期** | **是否有中间软件** |
| 瀑布型 | 是 | 没有 | 没有 |

## 1.3 项目估算

### 1.3.1规模估算

经过对系统的需求分析的了解，我们对系统的功能模块进行了划分，下面是本系统所涉及的功能模块图。

a.用户：信贷申请、保险申请、理赔申请、申请状态查看、申请管理、灾情视图。（如图2-5）

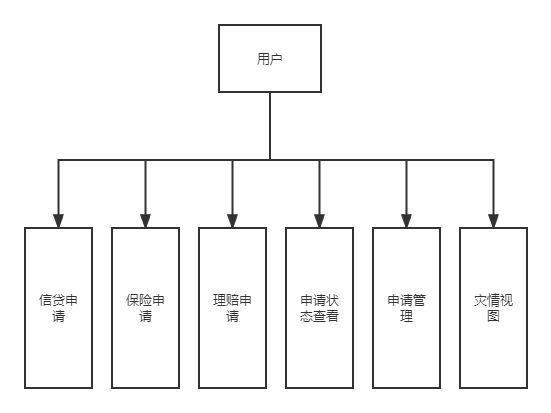


图3-4 用户功能模块

b.后台管理人员：用户申请管理、用户管理。（如图2-6）



图3-5 管理人员功能模块

c.系统业务流程分析（如图2-7）

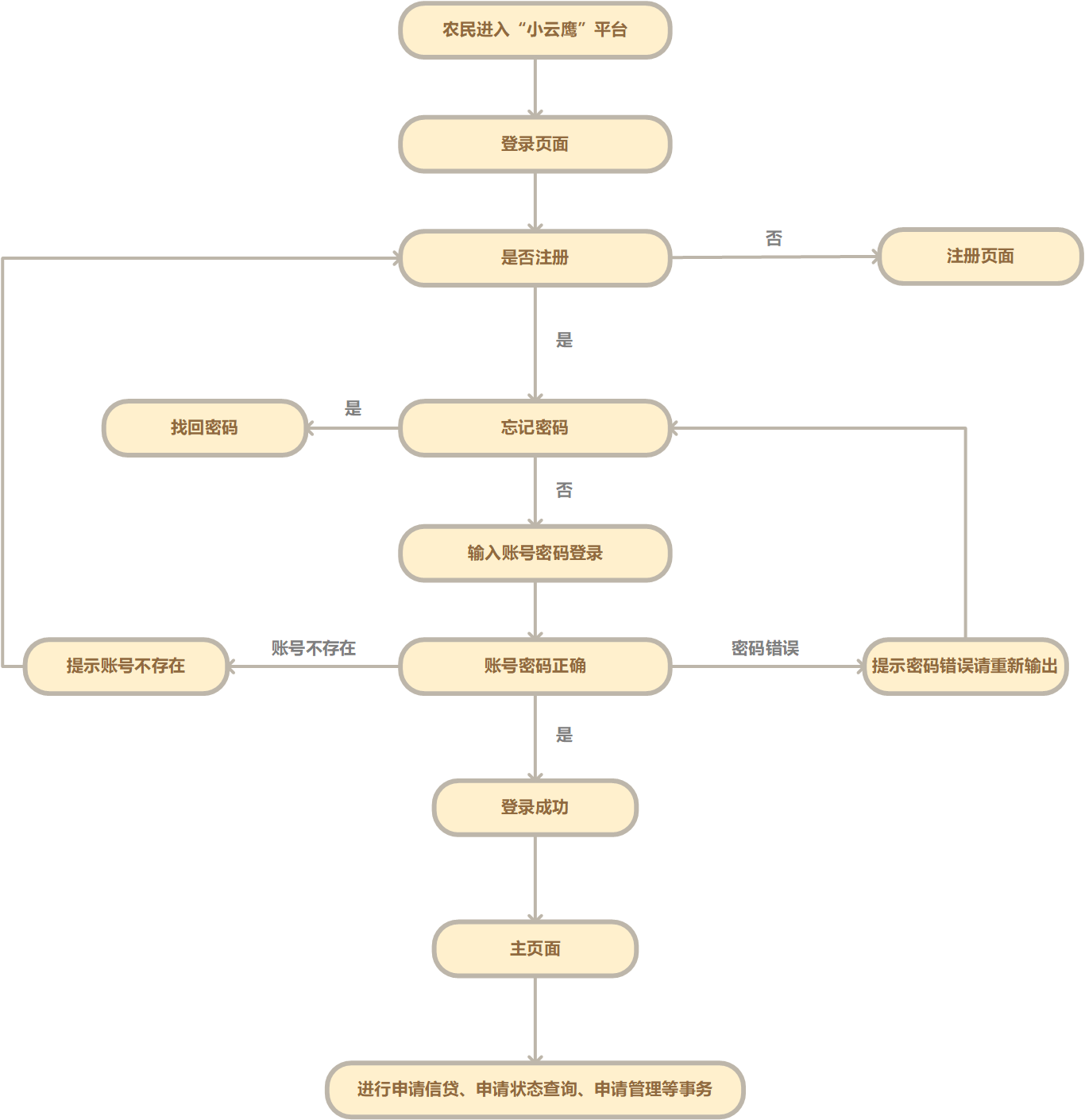


图3-6 用户登录及服务使用流程

d.平台业务流程图（2-8）

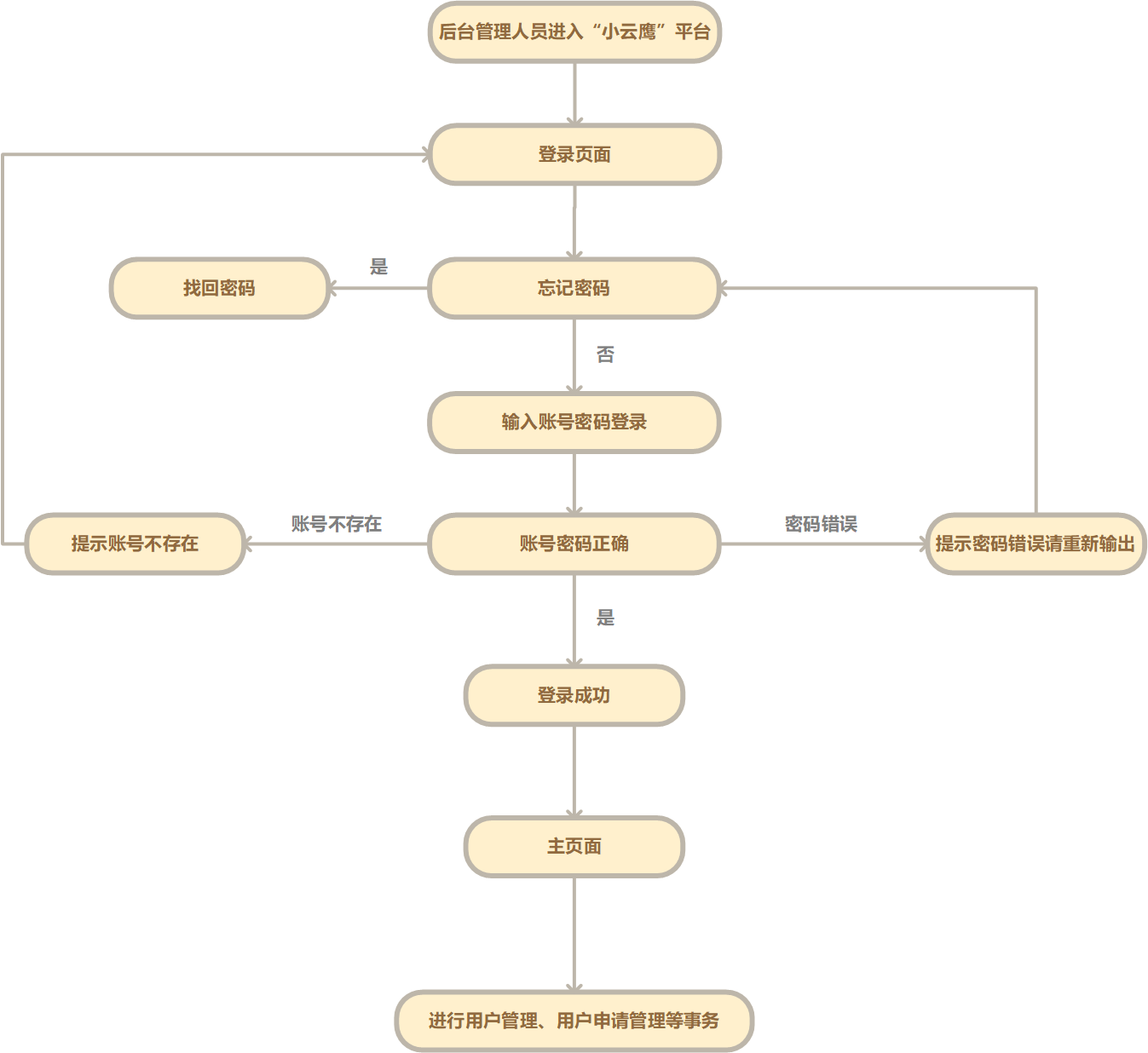


图3-7 后台管理人员登录及系统操作流程

### 1.3.2进度估算

|  |  |  |  |
| --- | --- | --- | --- |
| **阶段** | **开始日期** | **结束日期** | **工作日** |
| 软件计划 | 2月26日 | 3月11日 | 13 |
| 需求调研 | 3月12日 | 4月19日 | 7 |
| 需求分析 | 3月20日 | 4月2日 | 13 |
| 概要设计 | 4月3日 | 4月15日 | 12 |
| 详细设计 | 4月16日 | 4月30日 | 14 |
| 编码 | 4月20日 | 5月15日 | 25 |
| 测试 | 5月18日 | 5月24日 | 6 |
| 部署及验收 | 5月26日 | 6月1日 | 7 |
| **总计** | 3个月 | 3个月 |  |

### 1.3.3工作量估算

例如：

|  |  |  |
| --- | --- | --- |
| **阶段** | **阶段工作量占总工作量百分比** | **工作量（人.天）** |
| 软件计划 | 2周 |  |
| 需求调研 | 2周 |  |
| 需求分析 | 2周 |  |
| 概要设计 | 10天 |  |
| 详细设计 | 15天 |  |
| 编码 | 20天 |  |
| 测试 | 1周 |  |
| 部署及验收 | 5天 |  |
| **总计** | 3个月 |  |

### 1.3.4成本估算

“小云鹰”基于遥感农业大数据的涉农普惠金融平台初期运营每年需要约72万元的资金，主要用于人力成本、购买数据库产品以及后期运营维护工作。整体开发工作周期约为6个月，运营时长预测取为4年

具体各项支出见下表：

小云鹰项目支出统计表

|  |  |  |  |
| --- | --- | --- | --- |
| **项目类别** | **项目** | **概述** | **预算** |
| 人力成本 | 后端开发 | 1. 负责设计和开发系统中后台的工程架构 2. 承担AI、BI中台的工程能力落地以及相关架构的优化设计工作 3. 配合数据和算法工程师以及业务部门，针对业务需求，搭建搞笑的驱动算法迭代和数据流转的底层架构 4. 负责开发搞可靠性、高性能的分布式计算架构，为产品的高效迭代提供有力的保障 | 4000/人/月，共2人 |
| 前端开发 | 1. 负责依据产品需求文档，与UI设计师沟通，进而实现用户界面构建与展现 2. 负责与产品其他相关负责人员沟通，进行项目可行性与项目预算、时间评估 3. 负责对产品前端功能实现进行排期 4. 负责与后端开发人员沟通进行联调 5. 负责形成相关功能文档，并与测试人员沟通，对项目中可能出现的问题进行及时反馈与更正 | 4000/人/月，共1人 |
| 用户界面设计和维护 | 1. 负责参与具体的产品图形界面美术设计，根据产品需求完成图形界面的原型设计图和效果图设计 2. 负责优化图形界面的用户体验，适配用户习惯，根据产品实现进行人机交互的图形界面设计 3. 负责为指定产品设计并制作产品海报、LOGO以及其他电子美术资源。 4. 负责参与并指导前端工程师进行产品图形界面的开发，并负责部分的测试工作 5. 负责与产品测试及时沟通，对用户界面进行维护、升级 | 4500/人/月，共1人 |
| 保险产品人员 | 1. 负责进行保险产品开发工作，参与其具体设计及模型建构 2. 负责对保险市场进行调研，从事调研、设计、定价等产品全生命周期的开发、调研以及维护工作 3. 负责关注市场动态，紧跟市场发展方向。跟踪最新监管政策并对已有产品进行实时重新评估与修改。 4. 负责监控同业情况、政策导向等方面，对产品的设计、开发和维护提出建议 5. 负责进行保险产品管理以及产品分析，对上市后的产品进行维护支持，并负责后续的追踪分析以及统计工作。 | 4000/人/月，共2人 |
| 运营推广人员 | 1. 负责执行信用风险政策，包括投向政策、结构优化调整政策等 2. 依据国家金融法律法规、产品和货币政策以及信贷经营方针、政策和规章制度，对各类授信业务进行初审和复审，做出审查意见报告 3. 根据审查情况，分析宏观经济发展态势与热点行业系统性风险，并做出报告 4. 参与拟订信贷审查、审批各项规章制度，统一并明确各类授信业务的审查标准与审查要点 5. 负责对规定范围内的客户信用评级认定、客户准入的审核、放款核准等工作 | 4000/人/月，共2人 |
| 数据类 | 天气数据 | 1. 获取产品适应地区的历史天气数据，作为模型拟合、训练使用 2. 数据包含降水（毫米）、近地面短波辐射（瓦特每平方米）、近地面长波辐射（瓦特每平方米）、空气湿度（质量比）、风向风速（米每秒）、气温（开尔文）、近地面大气压（帕斯卡） 3. 数据按省平均值获取，时间颗粒度为年 | 0.01元/条  项目开发中需要约1000条，运营过程中实时更新。 |
| 历史收成数据 | 1. 获取产品适应地区的历史作物收成数据，作为模型拟合、训练使用 2. 数据包含小麦、水稻。未来可以扩充为多种植物。 3. 数据按省亩产平均值获取以反映各地不同的土地状况。时间颗粒度为年 |  |
| 遥感农业数据 | 1. 获取往年农田的遥感卫星影像，训练语义分割模型 2. 通过遥感农业大数据数据实时监控农田生长情况并进行预测 | 商业授权，5万/年 |
| 公司运营 | 场地租金及固定成本投入 | 1. 办公室场所租金 | 5.4万/年 |
| 1. 服务器购买 | 0.01万/件，购入10件 |
| 1. 固定电话及网络 | 0.15万/年 |
| 1. 计算机 | 0.5万/件，购入10件 |
| 1. 计算机桌椅 | 0.06万/件，购入十件 |
| 1. 打印机 | 0.2万/件 |
| 其他运营费用 | 1. 水电费用 | 1.2万/年 |
| 营销投入 | 线上推广 | 1. 创立微信公众号、微博等自媒体账号，专门设置人员进行运行等一系列活动 | 计入人力成本 |
| 1. 主要为与各大搜索网站合作，投放相应广告，如百度、谷歌、必应等几种大型的搜索网站 | 共计6万/年 |
| 线下运营 | 1. 公司将派专业人员与相关政府部门与金融机构商谈，以期达成合作 | 计入人力成本 |
| 合计 |  |  | 72.251万/年 |

## 1.4 人力资源安排

按照项目开发阶段或者模块进行人力资源的分配。例如

|  |  |
| --- | --- |
| **阶段** | **成员** |
| 软件计划 | 刘翼宇、潘怡珺 |
| 需求调研 | 刘翼宇 |
| 需求分析 | 刘翼宇 |
| 概要设计 | 潘怡珺、林洁 |
| 详细设计 | 罗梓丹、顾志远 |
| 编码 | 潘怡珺、林洁、顾志远 |
| 测试 | 罗梓丹 |
| 部署及验收 | 刘翼宇 |

## 1.5 开发环境

本作品分为四个部分：Python后端、SSM后端、前端、Jupyter 分析文件。其中 Jupyter 分析文件是进行数据处理与数据分析的文件，不需要部署，其他三项都需部署。

Python 后端需在 Python3.7及以上环境运行。通过安装最新版 Anaconda，并进入到项目目录下，使用pip工具安装requirements.txt中的依赖包。随后，使用Python分别运行Forrest.py以及FCN.py，从而启动图像识别以及预测服务。由于使用了深度学习技术进行高精确的识别和预测，对机器的内存、CPU 要求较高，因此部署 Python 服务时务必确保机器中至少有4GB可用内存。

SSM后端采用IDEA2021个人版进行开发。安配置好NodeJS和tomcat后，进入项目目录，执行npm install 命令安装依赖，随后执行命令 node bin/www 启动后端服务。

前端采用VUE技术打包，需要安装NodeJS。安装后，进入项目目录，执行npm install命令安装依赖，随后执行命令npm start启动前端服务器，使用浏览器访问本地8080端口浏览页面。后端依赖Redis、MySQL等数据库，数据的安装与部署不再赘述。

## 1.6 测试环境

由于本项目基于遥感卫星形象分析识别估产的特殊性，只能在实验室中进行初步的数据集交叉验证。

## 1.7 风险估计和控制

|  |  |
| --- | --- |
| **风险描述** | **降低风险策略** |
| 前后端分离式设计面临跨域风险 | 将主系统在本地服务器上运行 |
| 数据集获取难度大，拟合精确度难以保证 | 扩大数据集，使用多种数据和训练模型交叉验证取最优结果 |
| 预测产量的算法准确度有待商榷 | 使用ndvi叶面积指数方法时序预测，提高置信度 |

## 1.8 测试计划

| **测试** | **测试类型** | **负责人及成员** | **时间** |
| --- | --- | --- | --- |
| 1 | 单元 | 刘翼宇、林洁、顾志远 | 2-4月 |
| 2 | 集成 | 潘怡珺、罗梓丹 | 4月下旬到5月上旬 |
| 3 | 系统 | 潘怡珺 | 5月中旬 |
| 4 | 验收 | 刘翼宇 | 6月 |

# 需求分析

## 2.1 系统概述

小云鹰平台的核心技术是基于计算机视觉、机器学习的遥感农业大数据新应用，通过FCN网络模型处理遥感图像识别农作物，运用遥感领域的归一化植被指数处理方法来评判农作物长势优劣，使用改良的蝙蝠算法为农作物长势划分等级，与多种地理、气象数据共同作为变量，输入LGB模型建立最优预测产量模型。并运用深度学习的VGG、残差网络模型对比验证精确度。

## 2.2 需求概述

### 2.2.1 问题来源

近年来，国家在一号文件多次强调应当强化农村金融体系的构建，鼓励金融下乡，但在实际推进过程中却始终困难重重。一方面是由于农民主要资产大多为非标的资源性资产，不可以作为抵押品，因此无法实现抵押贷款，而信用贷款也只能停留在小额阶段；另一方面，农村地区客户分布较散，金融机构因其服务成本较高、风险较难控制等因素无法有效拓展并服务农村客群。两个方面的问题共同导致了我国的金融下乡工作迫切需要能突破生产端数据壁垒、调整家庭农场借贷方式的创新型金融工具。

然而国内目前能实时获得这类遥感农业数据的平台多用于科研领域，在与金融领域相结合方面几乎是一片空白。这就催生了需精准对接需求的服务于金融机构下乡的金融工具平台的创新与发展。

传统信贷放贷后需要人工周期性实地探查农作物长势、是否出现自然灾害等情况，小云鹰平台依靠遥感农业大数据能够定期勘察农田情况，帮助金融机构对农民的生产经营状况进行识别，在节省人力成本提升服务质量方面起到重要作用。

由于农业生产的长周期性，因此规模化家庭农场对金融服务的需求极高，而中国70万以上规模较大的家庭农场存在分布较为分散且作物、气候、环境多样化的问题，造成了人工贷款审核、后续监测困难的局面。且农业保险赔付的主要核算依据是人工审查，需要人工计算各家农户所承包托管的农田面积以及相应地理位置，由于农户状况的复杂性，数据的缺失及误差会给保险公司带来赔付力度的困难。

### 2.2.2 现有解决方案

竞品分析部分，主要和网商银行旗下的大山雀卫星风控系统进行对比分析：

表2-1 竞品分析

|  |  |  |
| --- | --- | --- |
| 产品名称 | 网商银行“大山雀”系统 | 小云鹰数字化农业金融科技服务平台 |
| 产品定位 | 卫星风控系统 | 金融科技服务平台 |
| 目标用户 | 有贷款需求的农民 | 有对农民发放贷款、保险等金融服务意愿的金融机构 |
| 主要功能 | 农民圈出土地，银行根据卫星遥感技术测算的土地面积、预测收成等状况为农民发放贷款 | 金融机构在小云鹰系统上传目标客户的相关数据，后台通过农业遥感大数据+机器学习联合风控模型给出风险评级 |
| 特色功能对比 | 大山雀系统通过调用支付宝的地图功能，农民在支付宝中圈出自己农田面积，卫星遥感技术远程分析获取相关数据 | 小云鹰平台为合作金融机构提供百度地图接口，智慧调取卫星影像并提供圈地服务。通过与农民个人信息交叉验证确保数据可信度 |
| 结合气候、行业景气度等情况，通过几十个风控模型，就可以预估产量和价值，从而向农户提供额度与合理的还款周期。 | 通过农业遥感大数据与计算机视觉、机器学习技术结合，通过联合风控模型给出目标客户参考风险评级和参考借贷额度与还款周期 |
| 无保险相关业务功能 | 提供保险评估业务，通过遥感和气候信息联合预测智能生成参考保费，实时监测和显示目标客户受灾情况，提供灾情导致的减产分析，实现智能理赔 |
| 结论 | “大山雀”作为网商银行内嵌式风控系统，为农民申请信贷提供了卫星遥感技术支持，简化农村信贷申请流程 | 小云鹰平台本质上是一款面向金融机构的专家系统，利用合作金融机构提供的有关数据+平台自身遥感农业大数据技术，简化金融机构面向农村弱信用群体的信用评估流程，给出可靠风险评估方案 |

### 2.2.3 本作品要解决的痛点问题

表2-2 用户痛点及解决方案

|  |  |
| --- | --- |
| 用户痛点 | 解决方案 |
| 农民资产多为非标的资源性资产，且农村土地流转状况复杂、农作物生长有不确定性，金融机构采用传统人力放贷成本过高 | 小云鹰平台通过基于计算机视觉和机器学习的遥感农业大数据新应用，通过向银行提供生产数据以及预估产量解决信息不对称的问题，开辟新型借贷模式，打破家庭农场因资源性资产无法进行抵押贷款从而融资困难的僵化局面。 |
| 由于农村土地和作物生长状况复杂性，保险机构在农业保险上的市场难以打开，缺少高效手段获得农业生产端数据 | 通过遥感农业大数据和以预估产量投保相结合的金融模式创新，旨在让农村客户的土地、农作物等不易估值的资产，发挥其最大价值，为农村客户的信用评价增添助力，同时极大提升农村理财产品的保险理赔效率，减少人力成本，实现金融机构和农业经营个体的共赢。 |
| 农作物生产与天气挂钩，天气等因素具有随机性，传统金融机构缺少手段对农村生产状况进行持续监测，难以掌握贷后信息和保险理赔信息，导致金融服务无法穿透农业生产端的数据壁垒 | 通过遥感农业大数据协助金融机构采集生产端数据，构建农业生产端数据库，简化实地调研过程。同时能够实时监测和更新农田数据信息，为金融机构把握投保农田的基本信息以及跟进农田经营状况提供有力保障。以预估作物产量从而给出参考保额的方式为金融机构进行投保提供有效参考，助力金融机构在因技术和数据壁垒尚处于空白阶段却蕴含广阔金融市场的农业领域抢得市场。 |

## 2.3运行环境

|  |  |  |
| --- | --- | --- |
|  | **名称** | **版本** |
| 后端 | Python | Python3.7及以上 |
| SSM | IDEA2021 |
| 数据库平台 | Redis、Mysql |  |
| 前端 | VUE | NodeJS |
| 客户端软件 | 无 |  |

## 2.4需求规定

以下展示核心信用评估功能实现界面：

### 2.4.1面向信贷申请的风险评估业务

金融机构获取有贷款需求的农民的相关信息，通过小云鹰平台提供的接口提交，小云鹰平台通过农民所圈土地面积与相关证明交叉验证，结合农业遥感大数据、期货市场等数据，利用后台算法进行信用评估，生成风险评级报告。



图1-3 农民信贷申请信息视图



图1-4 作物地区选择视图

### 2.4.2面向保险申购的风险评估业务

金融机构获取有购买农业保险需求农民的相关信息，通过小云鹰平台提供的接口提交，小云鹰平台通过分析遥感卫星图像对农民土地面积、作物生长态势等信息进行分析，结合农业遥感大数据、期货市场等数据，利用后台算法进行保费计算。

![C:\Users\mzy\AppData\Roaming\Tencent\Users\1277951424\QQ\WinTemp\RichOle\0OXCAP0IU](VKICN2D[1IPN.png](data:image/png;base64,)

图1-5 保险评估方案视图

# 三、概要设计

## 3.1系统概述

小云鹰的核心技术是基于计算机视觉和机器学习的遥感农业大数据新应用。我们用FCN网络模型来分析可见光波段的开源卫星遥感数据，从中提取出主要农作物的RGB值，之后输入FCN模型，可以识别出农作物的种类。之后我们获取哨兵二号卫星红外波段和近红外波段的数据，得到四个光谱波段的卫星遥感图片，运用遥感领域的NDVI处理方法提取农作物归一化植被指数的特征参数数据，结合改良的蝙蝠算法为农作物长势划分等级，从而给出一个农作物长势优劣的评判。

紧接着，我们从国家青藏高原科学数据中心得到实时的气象数据如气温、降水、湿度等，从农业知识服务系统搜索得到化肥使用量和往年亩产量的数据，与我们得到的农作物长势共同作为变量输入LGB模型建立最优预测产量模型，并运用深度学习的VGG、残差网络模型对比验证精确度。

图3-1 技术路线思维框架图

## 3.2 技术方案背景

我国是传统的农业大国，而且农业生产是当前我国社会经济的重要组成部分。农业生产好与坏不仅关系到一县、一市的农村经济问题，而且往往会关系到整个省区甚至全国。

以往金融机构都是采取人工的方法对农作物的面积和种类进行统计，农作物的产量也都是到秋后才能统计出来。这种情况下，我们只能以往年的农业收成数据（如图3-2所示）来对今年的收成进行预测。但由于天气、气温、环境和种植面积等的变化，每年的收成显然存在很大的时间性、空间性和地域性差异。速度慢、精准性差、人为因素过多，导致往年农作物产量数据不能真正的反映当年农业生产的状况。

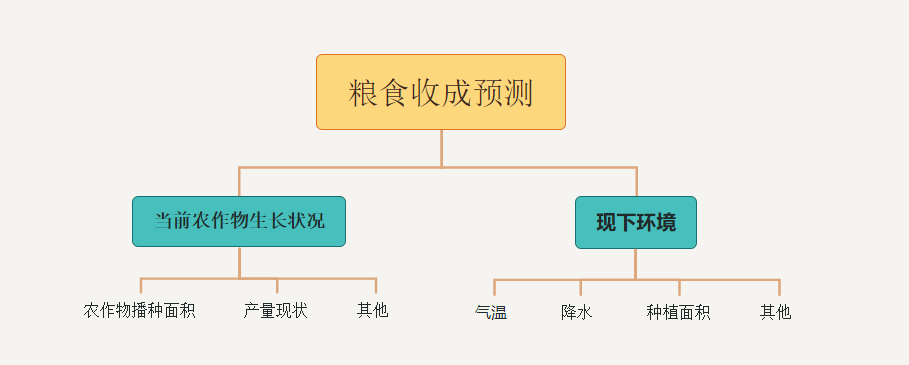


图3-2 收成预测影响因素

与各种传统的统计方法相比，遥感农业大数据具有多种类、多平台、多时段、多波段的特色和信息丰富、信息周期短、现时性和宏观动态性强等优势，无疑是最快速、有效的掌握农作物的生长状况和产量的信息之一。利用遥感农业大数据监测作物的长势包括作物的苗情、生长状况及其变化等，可及时识别农作物种类、提取农作物面积和预测农作物产量。

表3-1 传统统计方法与遥感农业大数据衍生技术对比

|  |  |  |  |
| --- | --- | --- | --- |
|  | 应用领域 | 特点 | 劣势 |
| 传统统计方法 | 传统金融行业信贷评估 | 采用人工统计的方式搜集农业生产端数据 | 速度慢、精准性差、人为因素过多、不能真正的反映农业生产的现状 |
| 遥感农业大数据衍生技术 | 精准农业、利用多时相影像发现土地利用变化和农作物估产 | 多种类、多平台、多时段、多波段的特色和信息丰富、信息周期短、现时性和宏观动态性强，技术先进 | 1. 多用于科研领域，数据保密性强，获取难度高 2. 精度问题 3. 遥感图像处理主流手段仍是目视解译。采用各种机器学习的分类算法实际应用目前还比较少   （4）一些极端天气、冰雪覆盖会对遥感影像产生影响 |

早在1979年，中国就开始关注农作物遥感农业大数据估产的意义，利用气象卫星监测作物生长状况始于20世纪80年代中期。从“六五”计划开始开展了农作物遥感农业大数据估产研究，并在区域尺度上开展产量估算试验。1984年开始，国家气象局组织北方省市开展冬小麦气象卫星遥感综合测产技术研究，组建了全国冬小麦遥感综合测产地面监测系统，开展了气象卫星监测冬小麦长势的研究，建立了不同类型的气象卫星遥感面积测算与估产方法。“八五”期间，遥感估产成为国家技术攻关内容，开展小麦、玉米和水稻大面积遥感估产试验研究。1998年中国科学院初步建立了国家级的农情监测系统，开展全国尺度的农作物长势监测。

表3-2 遥感估产在我国的发展历程

|  |  |
| --- | --- |
| 年份 | 发展状况 |
| 1979年 | 我国开始关注农作物遥感估产的意义 |
| 20世纪80年代中期 | 开始利用气象卫星监测作物生长状况，开展了农作物遥感估产研究，并在区域尺度上开展产量估算试验 |
| 1984年 | 国家气象局组织北方省市开展冬小麦气象卫星遥感综合测产技术研究 |
| 1998年 | 中国科学院初步建立了国家级的农情监测系统，开展全国尺度的农作物长势监测。 |

用遥感农业大数据监测农作物是一个时空变化的过程，即同一时相的作物长势在空间地域上存在差异，并且同一空间地域的作物在不同的时相上也存在差异。地表的自然状况决定着遥感农业有关图像的质量、色调，要想使遥感农业有关图像与现实地理环境相吻合，就要充分了解被调查区域的自然概况，去除干扰图像质量和不正常色调的因素。图3-3描述了遥感农业大数据估产中农作物收成预测流程。

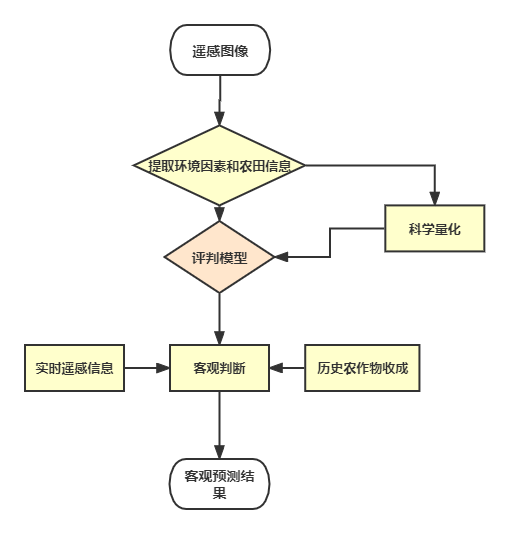


图3-3 农作物收成预测流程图

1. 软件架构及说明（软件架构图）

## 3.3后端实现

### 3.3.1系统分析

**1.需求分析**

**系统目标：**

本系统的目标是要完成一个基于B/S架构的分析遥感卫星图像和气象数据的涉农平台。该平台可以实现判读农作物种类，计算农作物面积，根据实时的天气和气候数据来预测当年的产量的功能，并以此减少金融机构在考察农民土地实际情况时的人力成本。同时，平台还可以实现自动化贷后监测与贷款风险预警。我们通过用户管理、事务管理、记录管理等功能模块，实现对本平台管理的自动化、系统化、规范化。本系统前端开发使用Vue框架，后端采用SSM框架，结合MySQL数据库开实现对数据的查询、添加、删除、修改等功能。

**用户需求和功能模块：**

用户：信贷申请、保险申请、理赔申请、申请状态查看、申请管理、灾情视图。

后台管理人员：用户申请管理、用户管理。

### 3.3.2系统概要设计

**1.系统功能模块图**

经过对系统的需求分析的了解，我们对系统的功能模块进行了划分，下面是本系统所涉及的功能模块图。

a.用户：信贷申请、保险申请、理赔申请、申请状态查看、申请管理、灾情视图。（如图2-5）

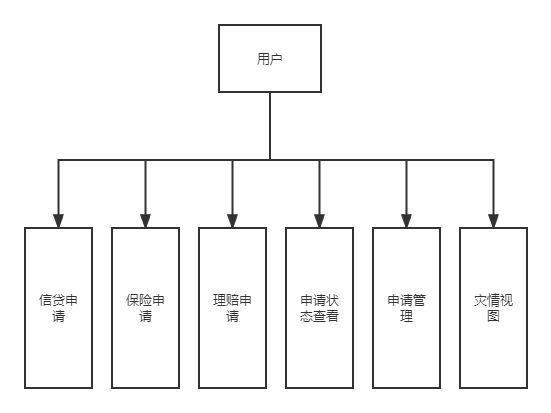


图3-4 用户功能模块

b.后台管理人员：用户申请管理、用户管理。（如图2-6）



图3-5 管理人员功能模块

c.系统业务流程分析（如图2-7）

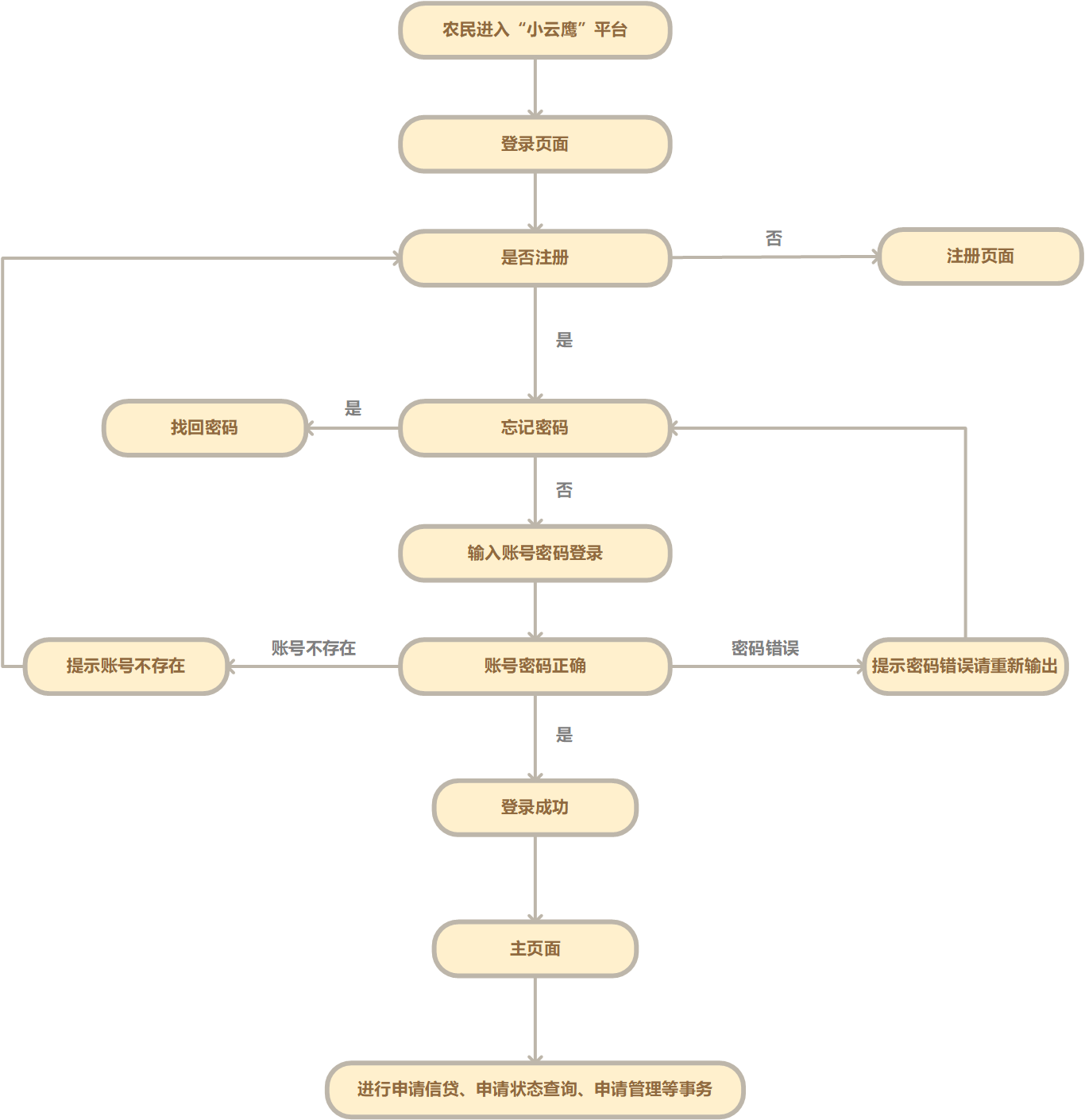


图3-6 用户登录及服务使用流程

d.平台业务流程图（2-8）

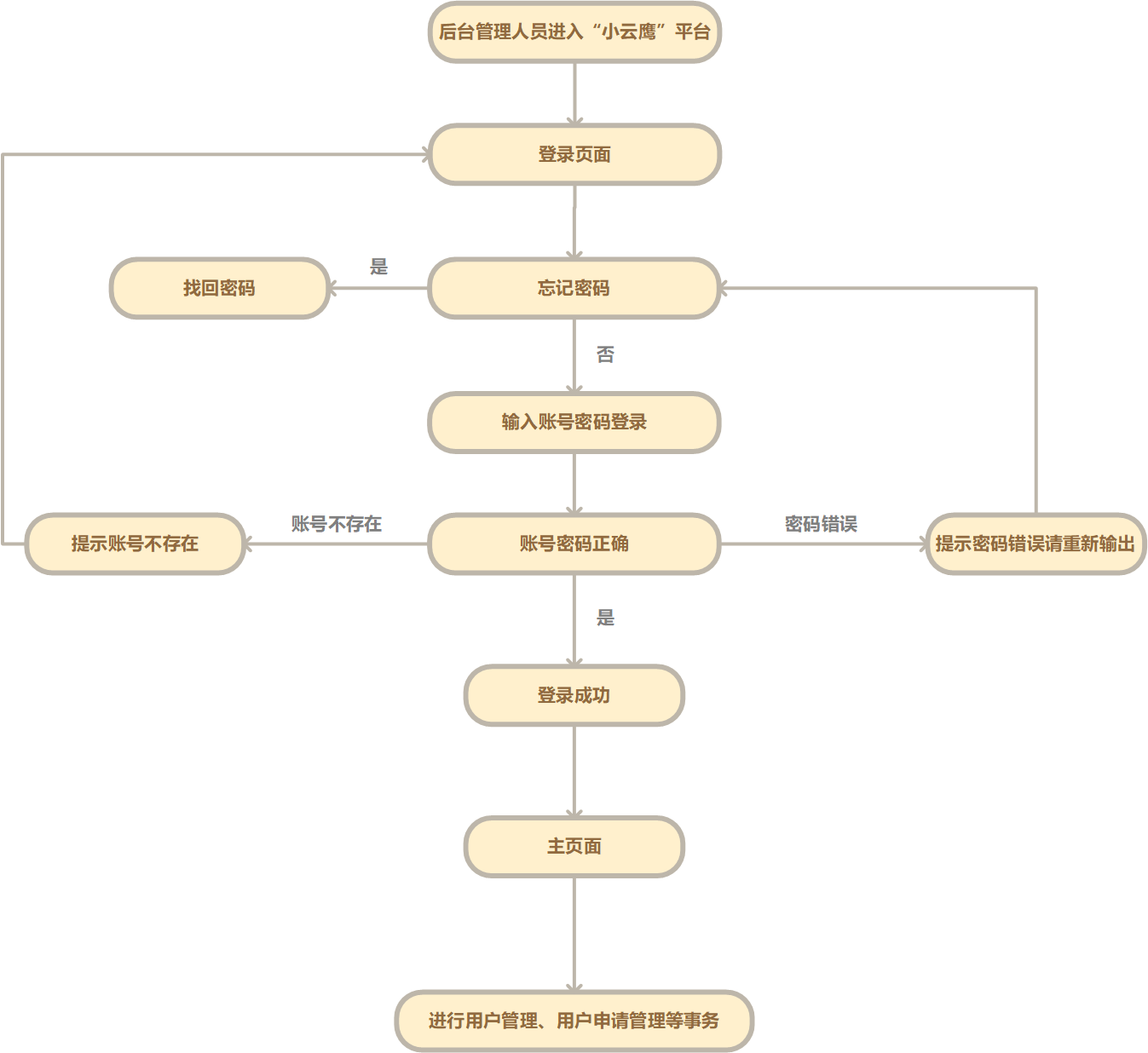


图3-7 后台管理人员登录及系统操作流程

### 3.3.3平台交互数据库设计

根据以上对数据库的需求分析，我们建立如下图2-9的E-R模型图。

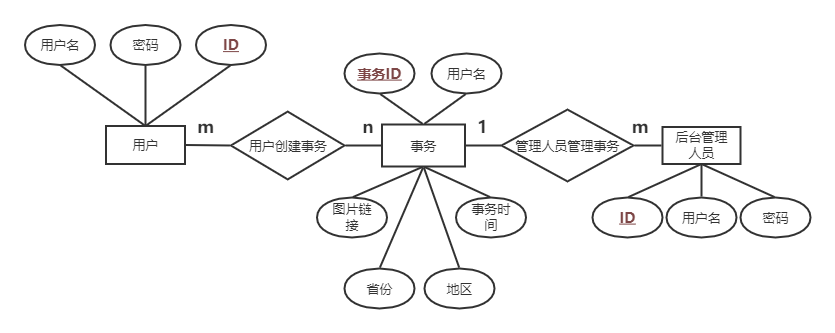


图3-8 E-R图设计

E-R图转换为关系模型如下：

管理员信息（ID，管理员帐号，管理员密码）

用户信息(ID，用户名，密码)

事务信息(事务ID，用户名，图片链接，省份，地区，事务时间)

数据表结构如下：

根据相关部分数据流程，我们建立了以下数据库中的表。

表2-3 transaction表

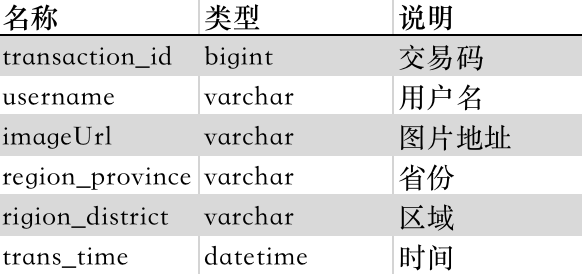


表3-4 coordinate表



## 3.4关键技术与算法

### 3.4.1 基于可见光波段遥感卫星影像数据的判读

用于农作物面积和种类的提取的可见光波段遥感卫星图片分辨率最好为10m左右，因为当分辨率太高时容易造成数据处理时金钱成本和时间成本的指数级增加，当分辨率太低时又容易造成较大的误差。我们选用百度卫星高清图片作为工作的基础技术资料，一方面保证准确度，另一方面保证用于训练的数据集的误差尽可能小。该图像色彩鲜明、层次丰富，其细微地表特征清晰可见，几何精度也能够符合要求，完全可以满足对图像中农作物RGB值的提取和农作物种类识别的工作。

由于全国各地的种植环境与种植情况，作物种类等的不同，我们选出十个代表性的地区数据作为我们研究的范例，分别为：

**1.种植作物代表：小麦**

表3-5 麦田图像收集表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 编号 | 地点 | 作物 | 观测时间 | 图片张数 |
| 1 | 郑州黄河北岸的原阳县河南省农科院小麦育种基地 | 普通小麦 | 六月上旬 | 30 |
| 2 | 潍坊寒亭区高里街道一空桥村 | 普通小麦 | 六月上旬 | 30 |
| 3 | 安徽省濉溪县 | 普通小麦 | 六月上旬 | 30 |
| 4 | 内蒙古巴彦淖尔市临河区 | 普通小麦 | 六月下旬 | 30 |
| 5 | 山东省德州市宁津县 | 普通小麦 | 六月上旬 | 30 |

**2.种植作物代表：水稻**

表3-6 稻田图像收集表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 编号 | 地点 | 作物 | 观测时间 | 图片张数 |
| 1 | 江苏淮安盱眙县 | 籼稻 | 9月上旬 | 30 |
| 2 | 安徽省广德市邱村镇 | 虾稻 | 10月上旬收割 | 30 |
| 3 | 黑龙江省五常市 | 五常水稻 | 4－9月，9月下旬收割 | 30 |
| 4 | 辽宁盘锦市兴隆台区国营兴隆农场 | 盘锦大米 | 10月上旬收割 | 30 |
| 5 | 黑龙江省友谊农场（国营） | 水稻 | 9月下旬成熟 | 30 |

由于选取的图像都为绿色系，RGB值差别不大，因此我们需要对可见光波段遥感卫星图像进行数据加强，用像素值的均值和标准偏差对像素值进行标准化，并将分类变量由多维数组降为一维，处理结果如下图3-9所示。我们采用opencv处理图像，将图片转化为灰度图。此时，我们就可以读取图片中的像素值，并对图片的衍生数据如torch.tensor等进行处理。之后我们选择用dataloader对数据进行四分类分批用四个不同的进程处理，并在每次训练前将数据打乱以降低模型的耦合度。

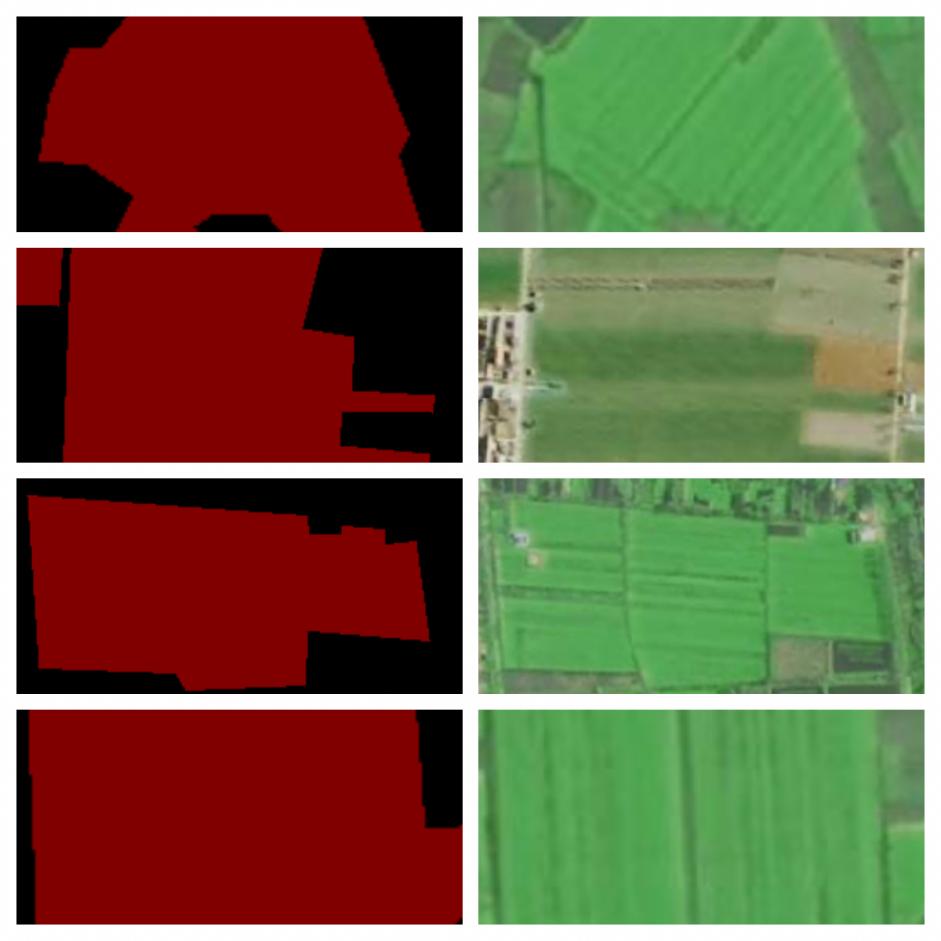


图3-9 遥感图像数据加强对比效果

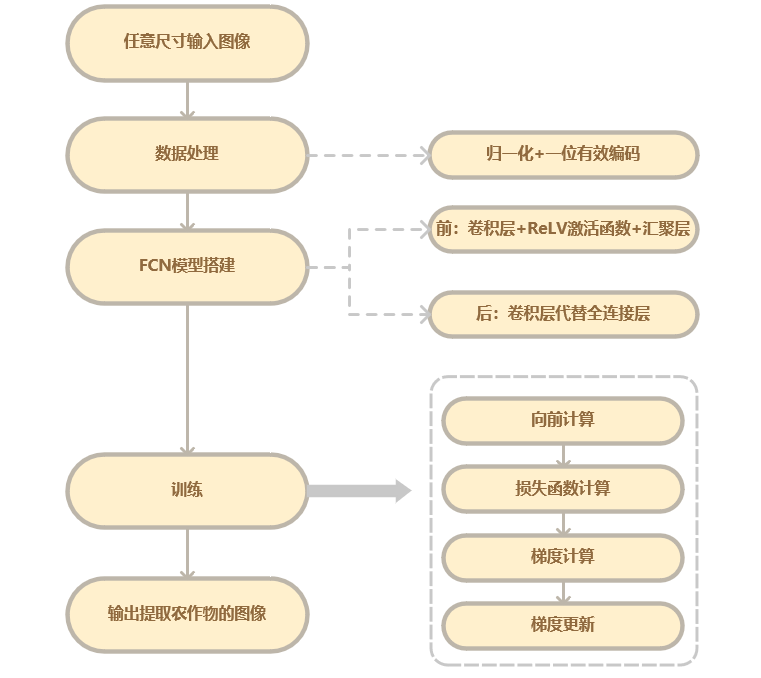


图3-10 语义分割提取农作物流程图

### 3.4.2 基于哨兵二号多光谱波段数据的农作物长势评价

**1.NDVI分析**

遥感农业的发展表明可用叶面积指数LAI来反映农作物的生长状况，而且其可以作为主要指标协助进行农作物产量的估计。通过遥感影像的红波段和近红外波段中提取出的遥感信息计算得出的植被指数与作物的叶面积指数、太阳光合有效辐射、生物量成正相关，可以用来评估农作物的长势情况。其中归一化植被指数NDVI与LAI具有很好的相关关系，因此我们选择NDVI值来反映作物的长势情况。通过多年遥感多光谱数据资料累积，我们可以计算出常年同一时段的平均植被指数，然后由当年该时段的植被指数与常年值的差异程度作为衡量指标，来判断当年作物长势优劣。具体流程如下图3-11所示。

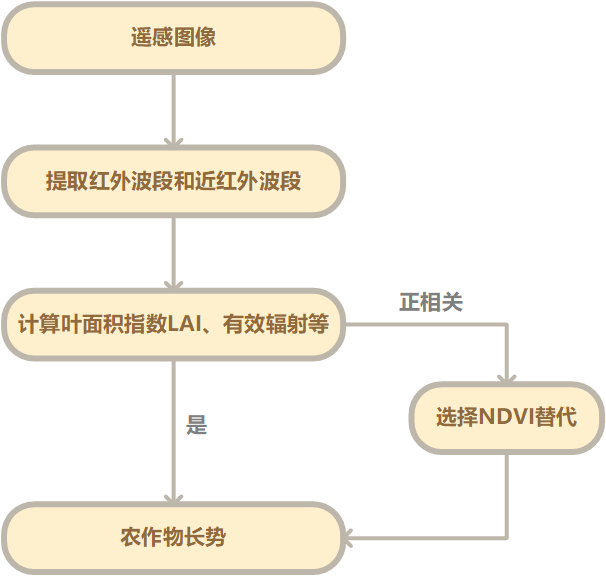


图3-11 NDVI分析流程图

对于多光谱遥感影像中蕴含的多种信息和数据，NDVI与作物生长周期关系密切，而NDVI又与近红外波段和红波段的线性组合呈现正相关关系，因此近红外波段和红波段的线性组合可以很好的反映作物的生长过程特征。具体表现为：作物生长初期，随着作物生长，叶子结构中叶孔的增加，叶子表面散热能力增强，近红外波段值逐渐增加，叶绿素吸收能力增强，红波段的值逐渐减少，NDVI 值逐渐增加; 而在作物生长末期，由于枝干由绿色变为黄色，叶绿素吸收能力减小，相应的红波段的反射值将会增加，叶面的叶孔相对收缩，散发的热量降低，近红外波段的值将会减小，NDVI有明显的下降。

图3-12 农作物生长阶段与光谱以及NDVI关系

如果将作物的 NDVI 值以时间为横坐标排列起来, 就可以形成作物生长的NDVI动态迹线，该动态迹线能够以最直观的形式反映作物从播种、出苗、抽穗到成熟收割NDVI的变化过程。以湖北省后湖水稻为例，用NDVI曲线模拟的水稻长势完全符合水稻的干物质积累，过程如下图3-13所示。

图3-13 水稻干物质积累与NDVI关系

因此，通过对农作物时序NDVI曲线的分析, 不但可以了解实时作物的生长状况，而且还能够反映作物生长的趋势, 作物NDVI曲线的提取与分析是作物长势监测的基础。其中，与NDVI关系密切的对应光谱波段号如图3-14所示：

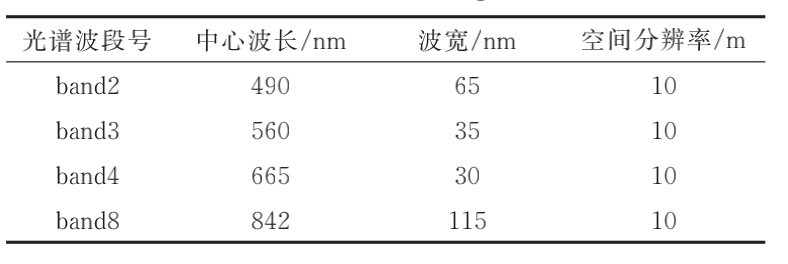


图3-14 光谱波段号对应数据

我们可以利用曲线形态变化与作物苗情变化的响应关系，提取NDVI曲线的特征参数，推测作物的生长发育状况，监测作物长势。作物特定生育期内，NDVI 的累计值与其最终生物量有较好的相关关系，可以依据此值进行年际间对比，从而定量的进行生长情况分析，为作物产量的计算提供依据。

不同地区影响作物长势因素不同，曲线特征参数变化具有区域性，需结合遥感农业气象大数据和地域性的特点来具体分析。因此需要提取时间过程曲线的特征参数，通过不同年份的特征参数与产量的相关分析，从中筛选出敏感的参数因子，来进行作物长势的定量监测。

**2.研究方法**

在 SNAP 处理下，选择含云量少，地物特征明显在返青期与抽穗期的影像，云处理代码如下图3-15所示，处理结果如3-16所示。

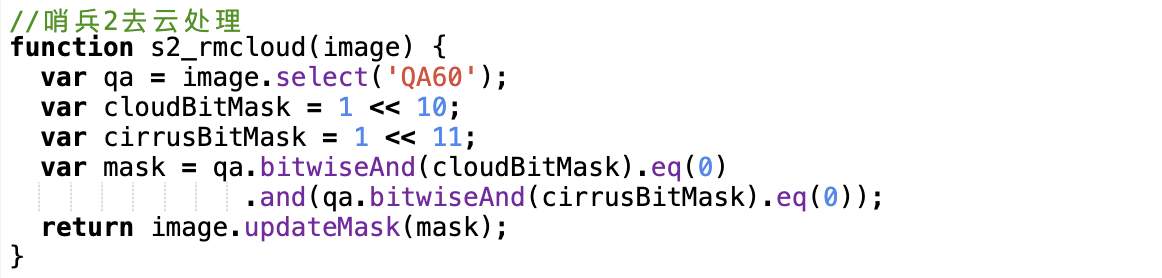


图3-15卫星光谱图去云处理核心代码



图3-16卫星光谱图去云处理结果

**3.Theil-Sen Median 趋势分析和 Mann-Kendall 显著性检验**

Theil-Sen Median 趋势分析法是一种稳健的非参数斜率估计方法，抗噪性强，受异常值影响小，可用于长时间序列数据变化趋势的研究中，能够科学直观地反映时间序列数据在一段时间内的变化趋势，评估其变化趋势。计算公式如下：

,∀j>i

式中，和分别为第i和第j年的农作物NDVI值，其中 i，j=1，2，3…，n；median为取中值函数。当Slope>0时，表示农作物NDVI整体呈上升趋势；当Slope=0时，表示农作物NDVI整体基本保持不变；Slope<0时，表示农作物NDVI整体呈下降趋势。我们运用该方法对农作物NDVI时间序列变化趋势进行显著性检验，对于给定的置信水平α，当|Z|> −时，可以认为在α水平上NDVI 时间序列有显著变化趋势，反之则为轻微变化。我们定义变化趋势在α=0.05下时，为显著变化；在α=0.01下时，为极显著变化。

**4.多元线性回归分析和残差分析**

考虑到农作物生长对气候变化具有一定的滞后效应，首先，采用相关分析法，计算不同的农作物NDVI与前0-12月气温、降水、相对湿度和日照时数的相关系数；然后，采用一阶偏导法，得到不同的农作物NDVI与气温、降水、相对湿度和日照时数的最大相关系数及其对应的滞后期；接下来，通过T检验法判断其相关性是否显著，若最大相关系数通过P<0.1显著性检验，则认为该气候因子对农作物NDVI影响显著。在此基础上，以农作物NDVI观测值和通过显著性检验的最大相关系数对应滞后期的气温、降水、相对湿度和日照时数为自变量，分区建立多元回归分析模型生成农作物NDVI预测值()，农作物 看作气候变化影响下的农作物NDVI。忽略其他非主要影响因素的条件下，建立残差分析模型，即可算出农作物NDVI残差值()，从而剥离气候变化的影响，得到人类活动作用下的农作物NDVI值。

式中，为系数；为最大相关系数对应滞后期的各气候因子；i 为气候因子类别，包括气温、降水、相对湿度和日照时数；ε为常数。当 >0时，说明人类活动对农作物生长具有促进作用，当 =0，说明人类活动对农作物生长的作用力微弱；当 <0 时，说明人类活动对农作物生长具有抑制作用。

**5.改良的蝙蝠算法**

目前鉴于结合群体智能算法在农业特征分类与评价领域的研究还非常有限，我们大胆突破，采用最新的群体智能算法——蝙蝠算法，其具有较好的智能型和鲁棒性，适应性比较广泛，在寻优过程中具有比较大的潜力，因此我们用其对归一化植被指数数据进行分析和处理，提取挖掘规则来构建模型。

每一条规则的挖掘需要蝙蝠在多维解空间中确定各个作物生长特征所对应的最优区间：即，、分别表示最优区间的下限和上限，即找到每一种评价指标所对应的最优分割点。对于有n特征指标的作物数据来说，蝙幅即在一个2n维的空间内进行最优解搜索。

* 评价规则构造过程：

初始化规则集R，此时规则集为空集。对蝙蝠种群进行初始化，记第i只蝙蝠的初始位置为。

其中为第i只蝙幅在第j个评价指标上的下限和上限。

初始速度,其中分别为上限和下限方向上的最大速度。

评价模型的输入是我们之前用哨兵二号多光谱波段数据经过处理、分析所得的NDVI特征参数，而其输出则是作物的生长态势等级。我们基于蝙蝠算法构造出一种新的分类器B-E，具体规则如下图3-17所示，利用蝙蝠群体的相互协作寻找各作物评价特征的最优上下限，来构造分类评判规则，并利用这些规则构建综合评价模型，对作物生长态势进行评判。评价模型的建立可以分为两个部分：首先建立基于蝙蝠算法的规则分类挖掘算法，将原始数据集划分为训练集和检验集，并利用训练集进行规则挖掘，完成基于规则的分类评价模型的设计；然后利用之前划分的检验集对设计好的评价模型的性能进行检验，测试其分类评价效果。利用挖掘算法建立的综合评价模型就能够将输入的模型的新样本，依据模型中的最优规则，给出与其所对应的生长态势等级。

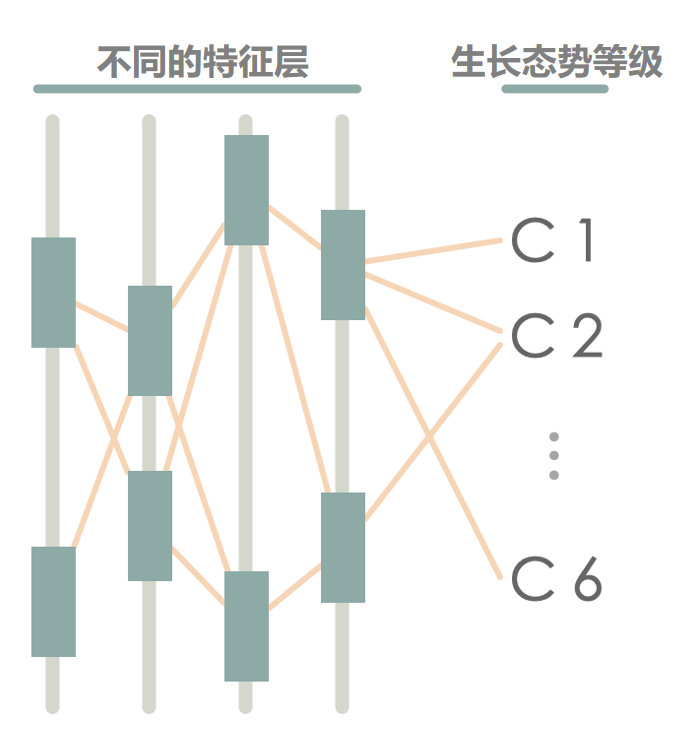


图3-17 B-E规则挖掘原理

* 具体过程：

运行蝙幅算法后计算各蝙幅的适应度值，比较每只蝙蝠当前时刻与迭代前的适应度值，若优于迭代前的最优位置则继续进行更新，否则保持当前最优位置不变。计算整个蝙蝠种群中所有蝙蝠的个体最优值，适应度最优的个体即为全局最优值。

随后按照公式更新蝙蝠的速度和位置。当全局最优值的适应度达到设定的阈值或者迭代次数超过最大迭代次数时终止这一次的规则挖掘，此时得到一条最优分类规则。

将得到的蝙幅最优位置置于规则集中R，并在训练数据中移除这一条规则所覆盖的数据，即作物评价指标与生长态势等级均与此规则相匹配的样本数据。训练集中剩下的数据继续用来进行训练，若某一类别的数据个数小于阈值时，视为该类别数据过小，不再进行这一类别数据的规则挖掘，转而进行下一类别的规则挖掘。直到所有类别的规则都挖掘完毕，则完成规则集的建立。

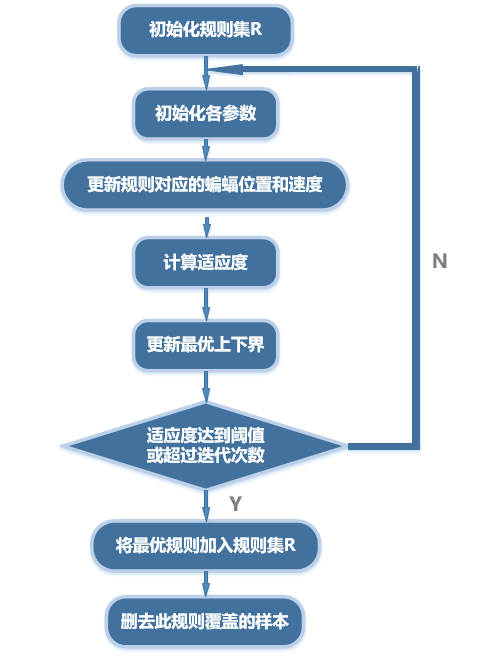


图3-18 蝙蝠算法

* 规则适应度评价：

的值为当前规则的质量。式中TP,FP,FN,TN和均为训练集中满足条件的样本个数。TP表示满足规则前件，且与规则的预测类别相同的样本数；FN表示不满足规则前件，但与规则的预测类别相同的样本数；FP表示满足规则前件，但与规则的预测类别不相同的样本数；TN表示不满足规则前且与规则的预测类别不相同的样本数。

表3-7 农作物影像经过一系列图像处理、分析所得的最终结果



### 3.4.3 基于实时气象数据和往年亩产量数据的农作物收成预测模型

我们对一系列气象数据，化肥数据等与农作物产量有关的数据进行了处理，可以得到13个特征值，1个目标值，此时总数据量为1872，是一个1872行，14列的二维数组。

我们使用Robust方法对所有的数据进行归一化，此方法可以去除噪声和离群点的影响，增加系统适应数据变化的容忍度，以此提高鲁棒性。之后用箱式图（图3-19）检测数据的异常值，可以发现V1,V3有很多数据偏离正常值，处于四分位点以下，故这些数据存在许多较大的异常值，可以移除。为了进一步规范数据，我们还绘制了Q-Q图（3-20）查看数据是否近似于正态分布，结果表明数据基本随对角线分布，不需要后续继续使用数据变换。

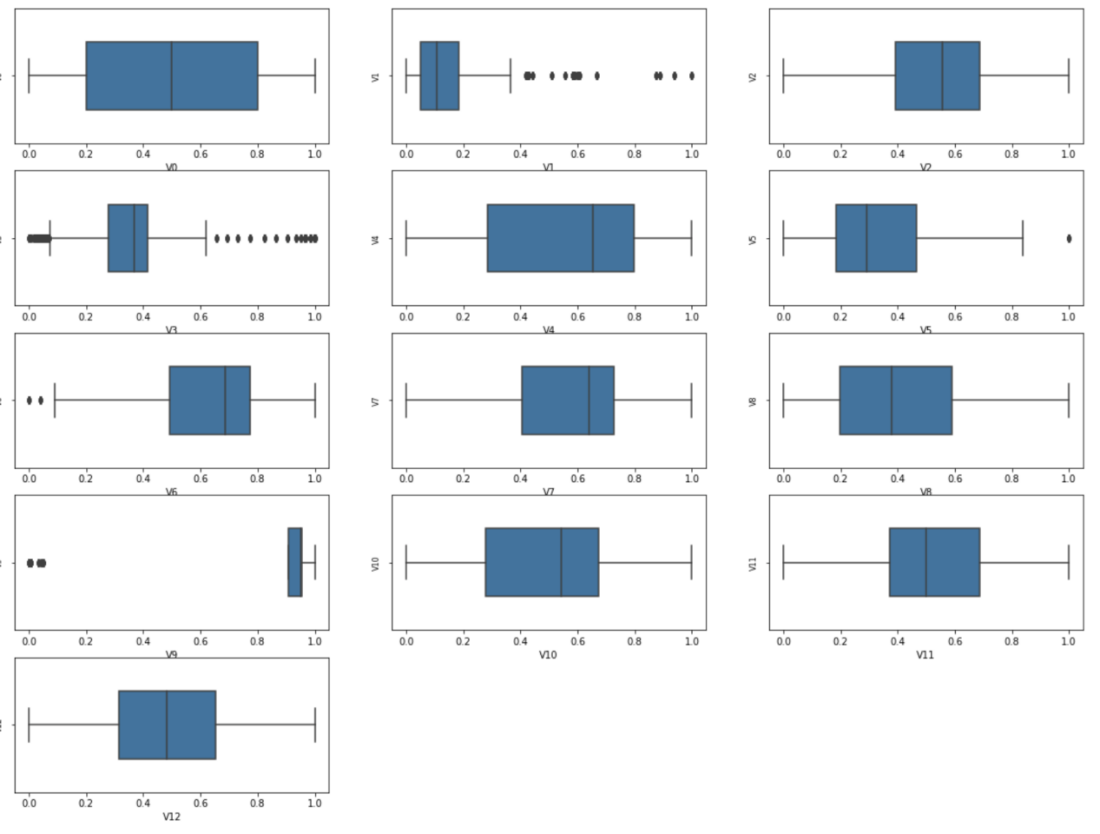


图3-19特征箱式图

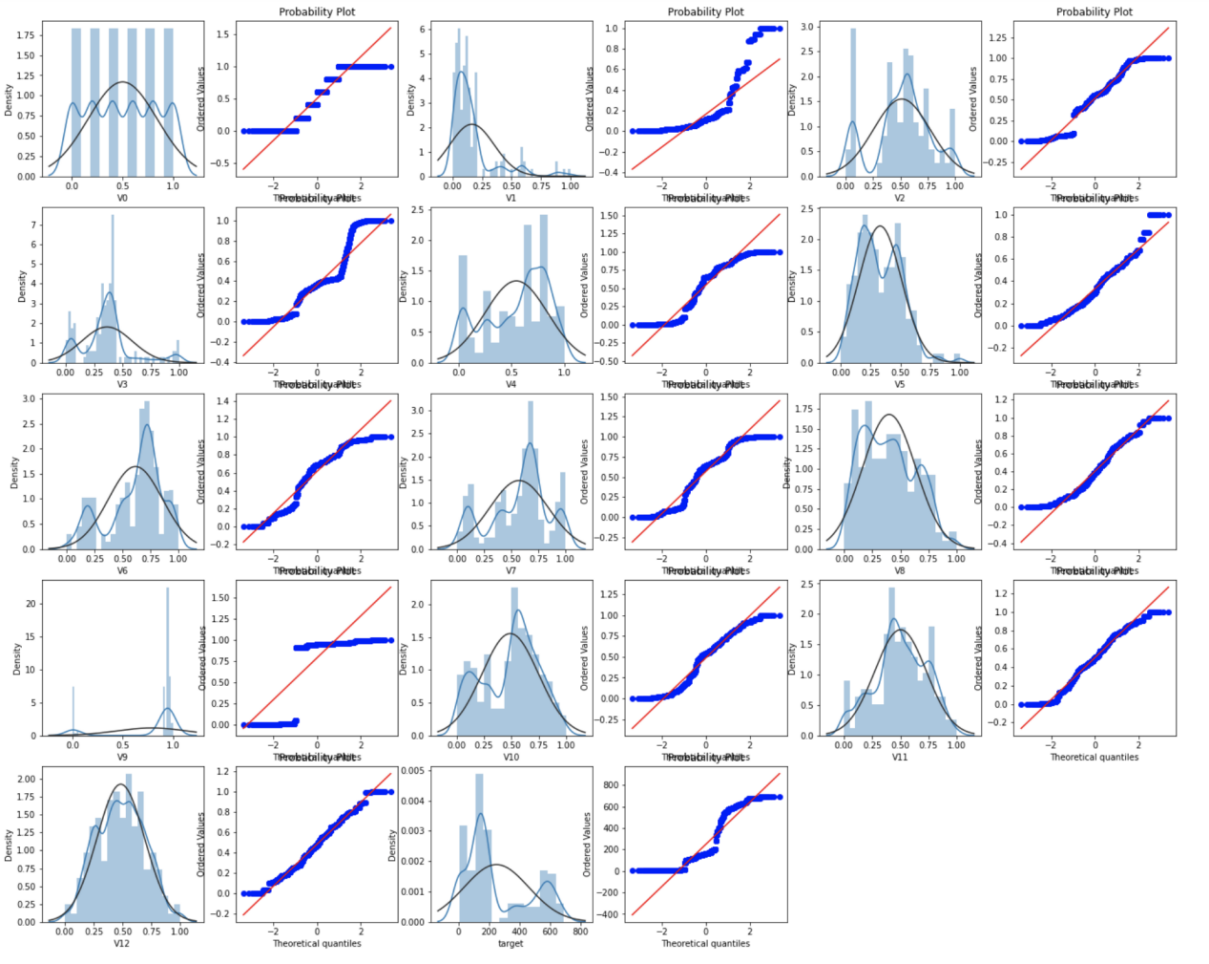


图3-20 数据分布直方图和Q-Q图

在对数据的相关性分析中，画出梯形热力图（图3-21），结果为所有特征变量和target变量之间的相关性系数，与对角线数据保持一致。

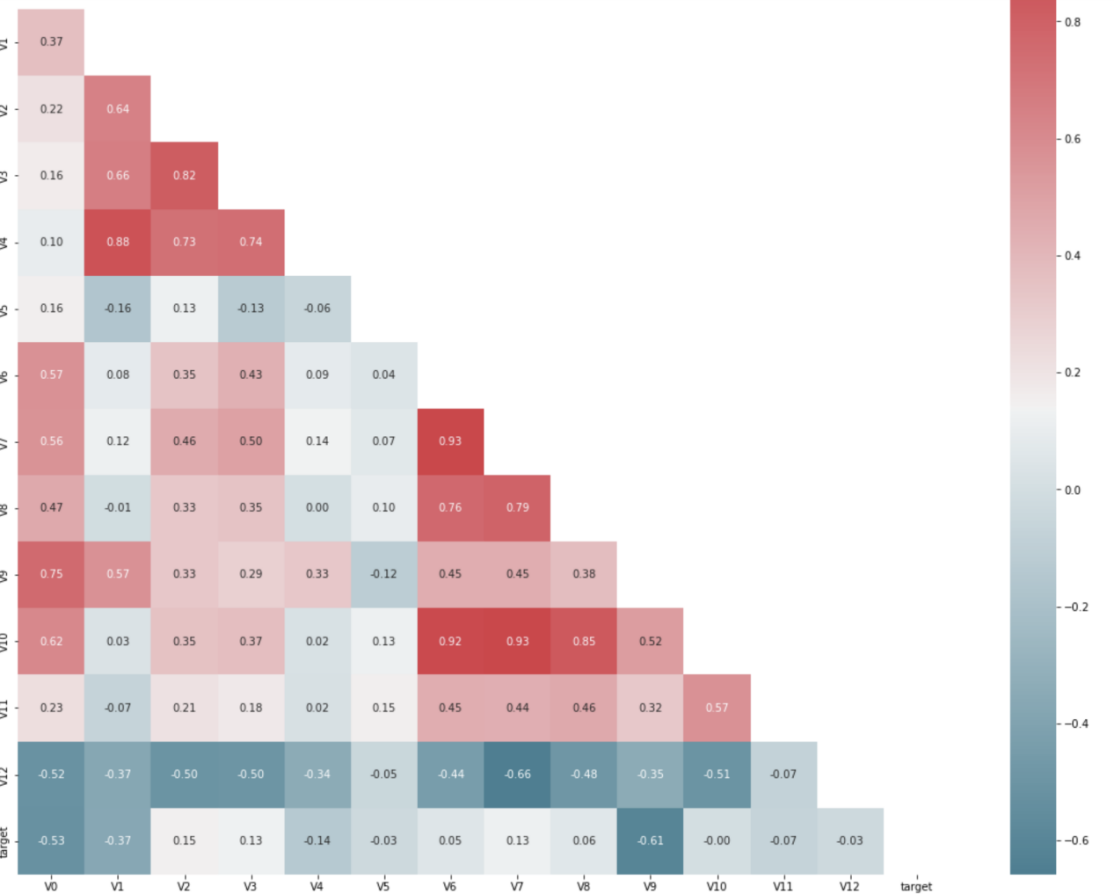


图3-21 热力梯度图

我们筛选出相关性较大的7个数据并利用岭回归（图3-22）对train 数据进行建模过滤异常值。

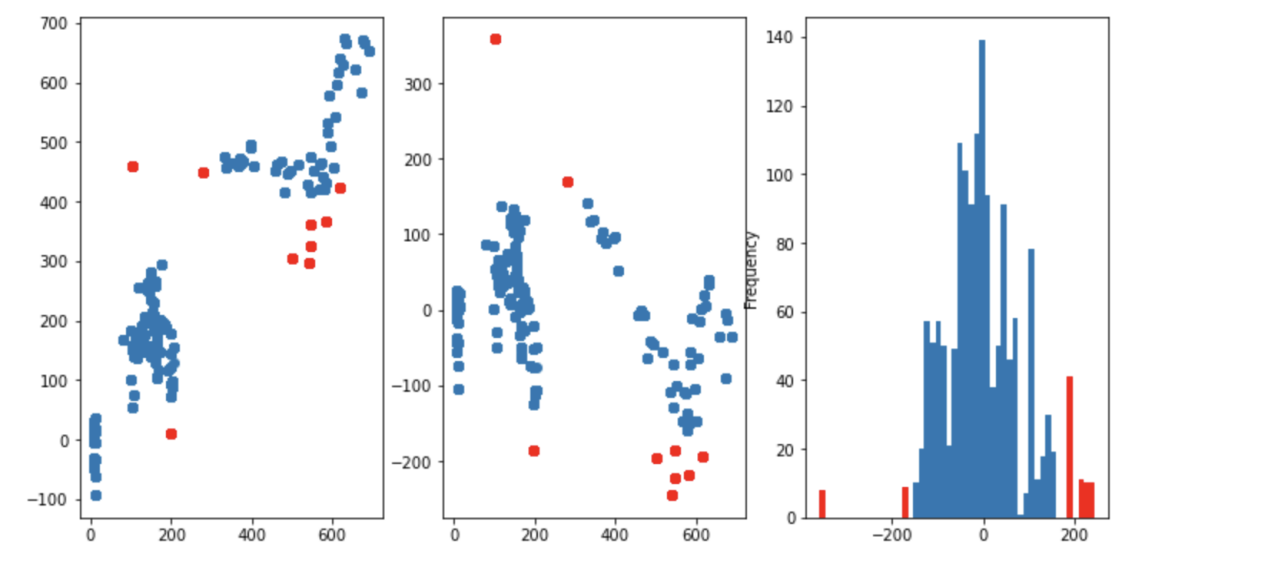
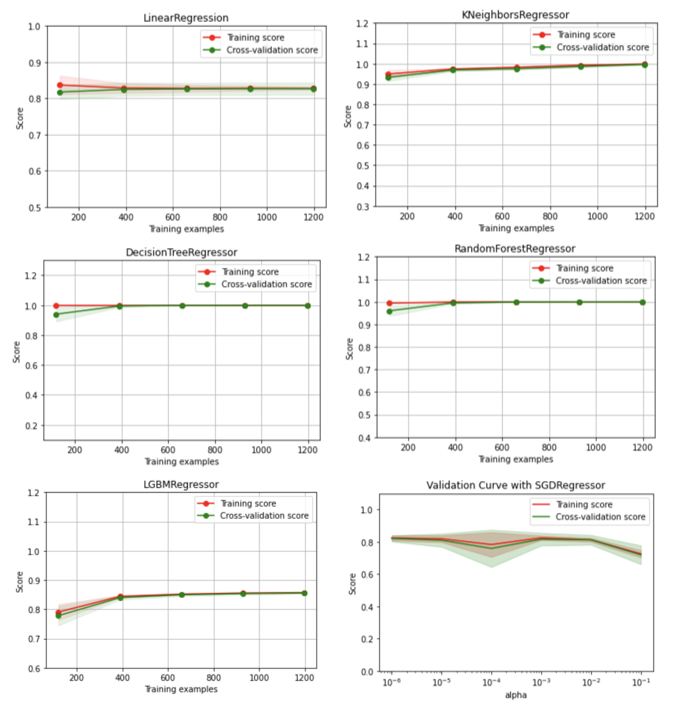


图3-22 岭回归样本异常值可视图

我们使用交叉验证法来训练优化选择模型，分别采用了线性回归、k邻近回归、决策树回归、随机森林回归、LightGbm模型、SGD回归的方法（图3-23）。

图3-23 不同模型性能比较及验证图

综上，我们发现LGB 模型均方损失误差最小，故它在预测产量上的效果最好。LGB是实现GBDT (Gradient Boosting Decision Tree)算法的框架，其主要思想是利用弱分类器迭代训练以得到最优模型，该模型具有训练效果好、不易过拟合等优点。它支持高效率的并行训练，并且具有更快的训练速度、更低的内存消耗、更好的准确率、支持分布式处理数据，因此可以减少我们模型预测时间，之后我们采用深度学习方式，用ResNet和VGG模型的预测结果与上述模型结果进行对比，进一步进行准确率分析，防止由于个别数据的偏差引起模型的失真。具体训练过程如图3-24所示。

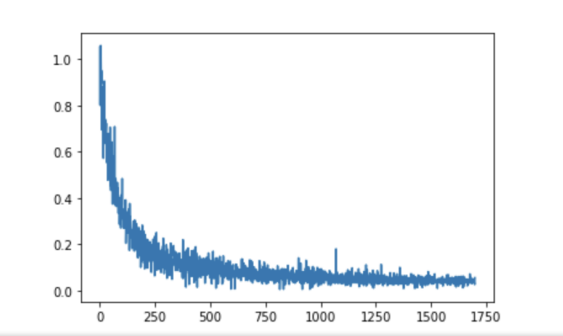


图3-24 模型训练过程



图3-25 产量预测流程图

## 3.5 系统功能设计

### 3.5.1 主界面

小云鹰涉农平台首页界面如下图所示。



图3-26 小云鹰涉农平台主界面UI

### 3.5.2 业务功能界面

**1.气候信息实时视图**

小云鹰涉农平台气候信息视图如下图所示，以统计图的形式展现近10小时地区气温/降水量情况，并实时更新。



图3-27 环境视图功能实现

**2.全国灾情实时视图**

小云鹰涉农平台全国灾情视图如下图所示。



图3-28 全国实时受灾情况视图

点击具体的省区，可以展示对应市级单位的实时受灾情况。



图3-29 陕西省（例）实时受灾情况视图

**3.信贷申请功能视图**

小云鹰涉农平台信贷申请视图如下图所示。



图3-30 信贷申请视图

提交信贷申请单后，可以在当地的卫星地图中选择具体作物地区。



图3-31 作物地区选择视图

**4.保险申购功能视图**

小云鹰涉农平台信贷申请视图如下图所示。

![C:\Users\mzy\AppData\Roaming\Tencent\Users\1277951424\QQ\WinTemp\RichOle\0OXCAP0IU](VKICN2D[1IPN.png](data:image/png;base64,)

图3-32 购买保险视图

1. **理赔申请功能视图**

小云鹰涉农平台理赔申请视图如下图所示。



图3-33 理赔申请视图

## 3.6 数据库设计

后端依赖Redis、MySQL等数据库，主要数据分为图像数据集和产量预测数据集，图像数据集分为遥感卫星图像数据集和无人机航拍数据集，产量预测数据集分为气象数据集和农作物期货价格及涉农公司股价数据集。

数据来源：

**（1）图像数据集：**由遥感数据由具有较高的空间分辨率和时间分辨率的哨兵二号卫星得到，已经过气象处理，再进行云层处理得到分辨率为10m的农作物图像。

**（2）预测数据集：**通过民用卫星哨兵二号获取多光谱图像(GEE)，选用其中的第8波段为红外波段，第4波段为红光波段，进行两者之差比两者之和，提取农作物四个生长时期的NDVI，进行生长态势分析，进行生长态势的评估作为预测的特征向量

**（3）气象：**国家青藏高原科学数据中心

**（4）化肥：**农业知识服务系统搜索

**（5）亩产量：**农业知识服务系统搜索

数据规模：

预测数据量为1872，是一个1872行，14列的二维数组。

图像数据为200个农作物图像及其标签，以及为扩大数据集后续对图像进行的增广。

表3-8



# 四、测试计划

## 4.1 后端测试报告

表5-1 测试报告

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **测试项目** | **测试内容** | **测试要求** | **测试结果** | **备注** |
| 软件结构测试 | 是否使用B/S架构 | 1.判断客户端输入有效  2.判断服务器可以对客户端的输入做出正确的响应  3.关闭服务后客户端输入无效 | 使用了 B/S 架构对客户端属于做出正确响应 |  |
| 网络结构测试 | 验证网络数据流向 | 使用 TCPDUMP 验证节点间数据流向 | 网络数据流向正常 |  |
| 先进性测试 | 根据系统总体结构和采用的技术对系统的先进性作出评价 | 由相关组织，部门，机构或个人在了解系统构成的基础上对系统的先进性作出合理评判 | 系统设计模式为MVC 设计模式 |  |
| 开放性测试 | 是否采用标准协议，接口规范 | 由相关组织，部门，机构或个人在了解系统构成的基础上对系统的开放性作出合理评判 | 符合标准协议，接口规范 |  |

## 4.2 数据处理与测试

**（1）图像数据集增广**

图像增广在对训练图像进行一系列的随机变化之后，生成相似但不同的训练样本，从而扩大了训练集的规模。 此外，应用图像增广的原因是，随机改变训练样本可以减少模型对某些属性的依赖，从而提高模型的泛化能力。 例如，我们可以以不同的方式裁剪图像，使感兴趣的对象出现在不同的位置，减少模型对于对象出现位置的依赖。 我们还可以调整亮度、颜色等因素来降低模型对颜色的敏感度。具体处理方式如下图：

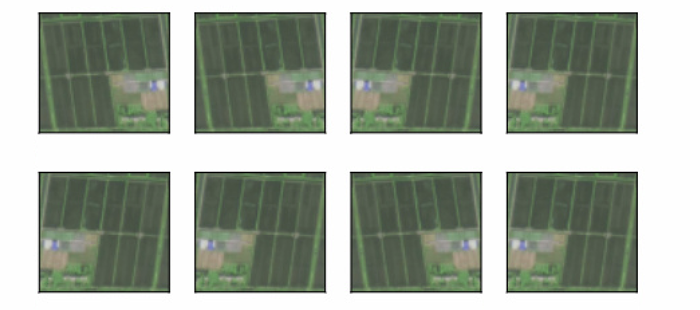


图5-1 不同方位增广

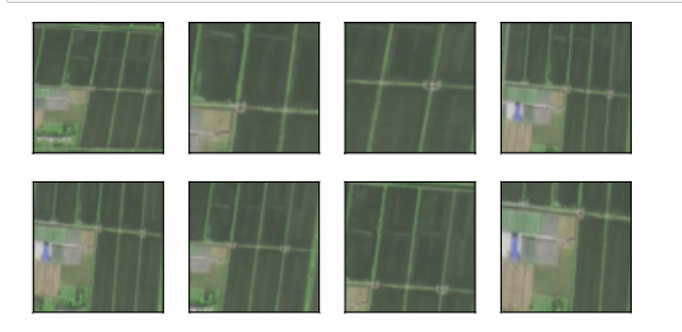


图5-2 进行缩放增广

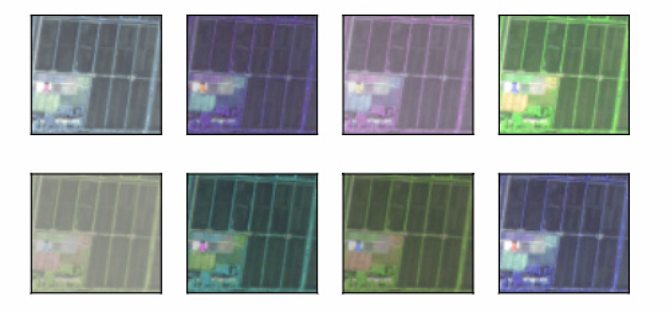


图5-3 颜色与亮度调节进行增广

1. **相关数据处理**

首先是导入数据，进行查看。如图5-4所示，可以看到有13个特征值，1个目标值，数据量为1872，是一个1872行，14列的二维数组。

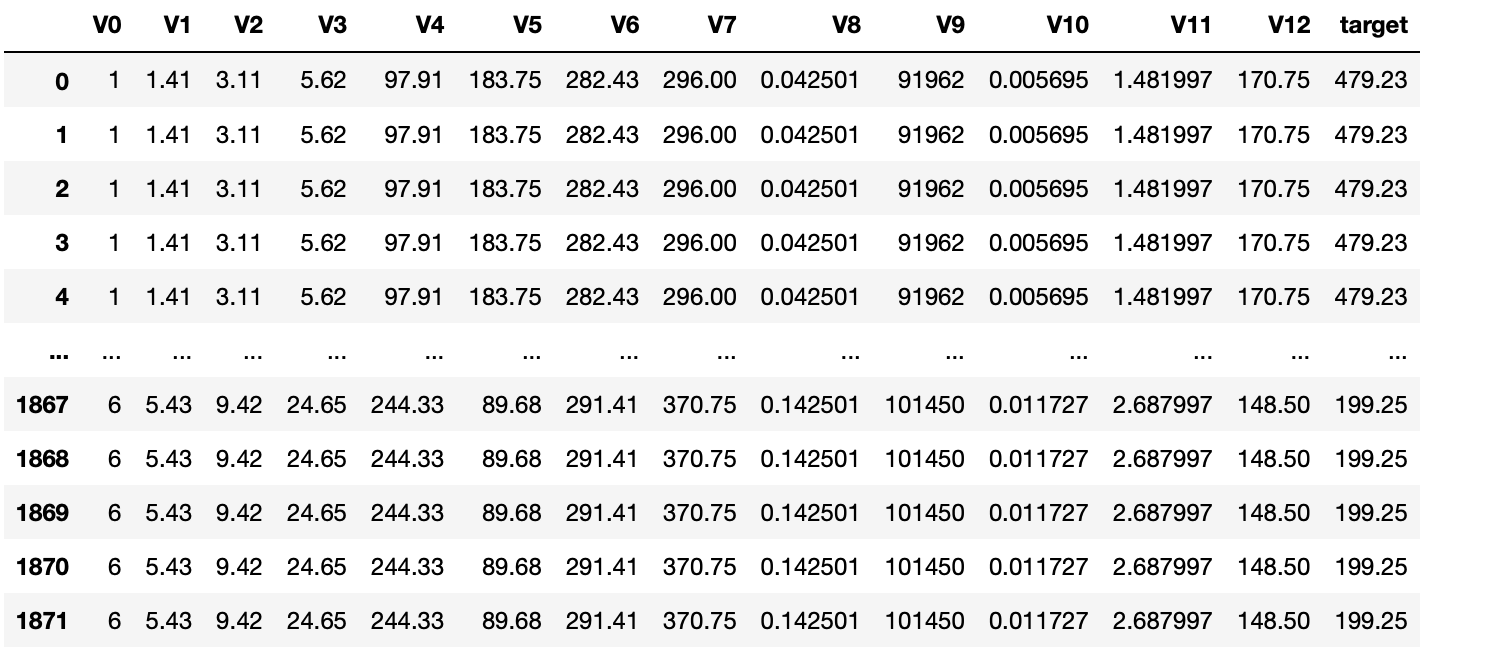


图5-4 数据集导入

这里使用Robust方法对数据集进行归一化处理，可以使是模型具有较高的精度或有效性，这也是对于机器学习中所有学习模型的基本要求。去除噪声和离群点的影响，因为用鲁棒性理解述算法对数据变化的容忍度有多高。鲁棒性并不同于稳定性，稳定性通常意味着特性随时间不变化的能力，鲁棒性则常被用来描述可以面对复杂适应系统的能力，需要更全面的对系统进行考虑。处理结果如图5-5所示。

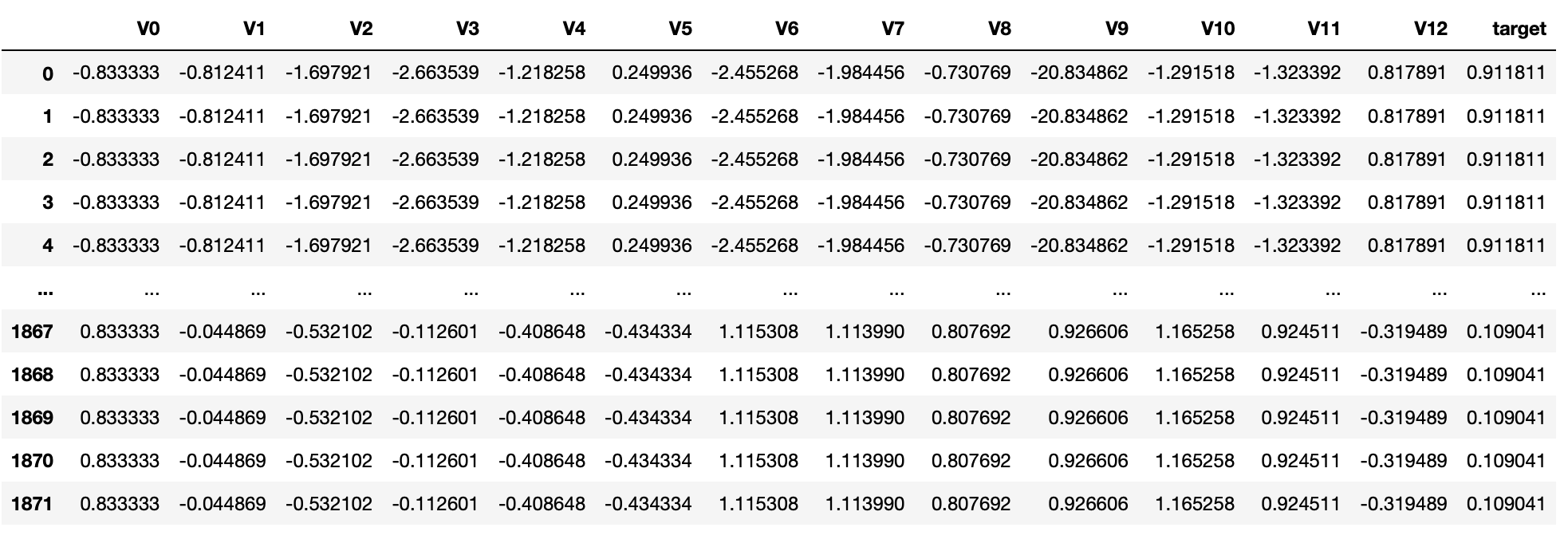


图5-5 归一化处理

接下来查看箱式图，可用于异常值检测

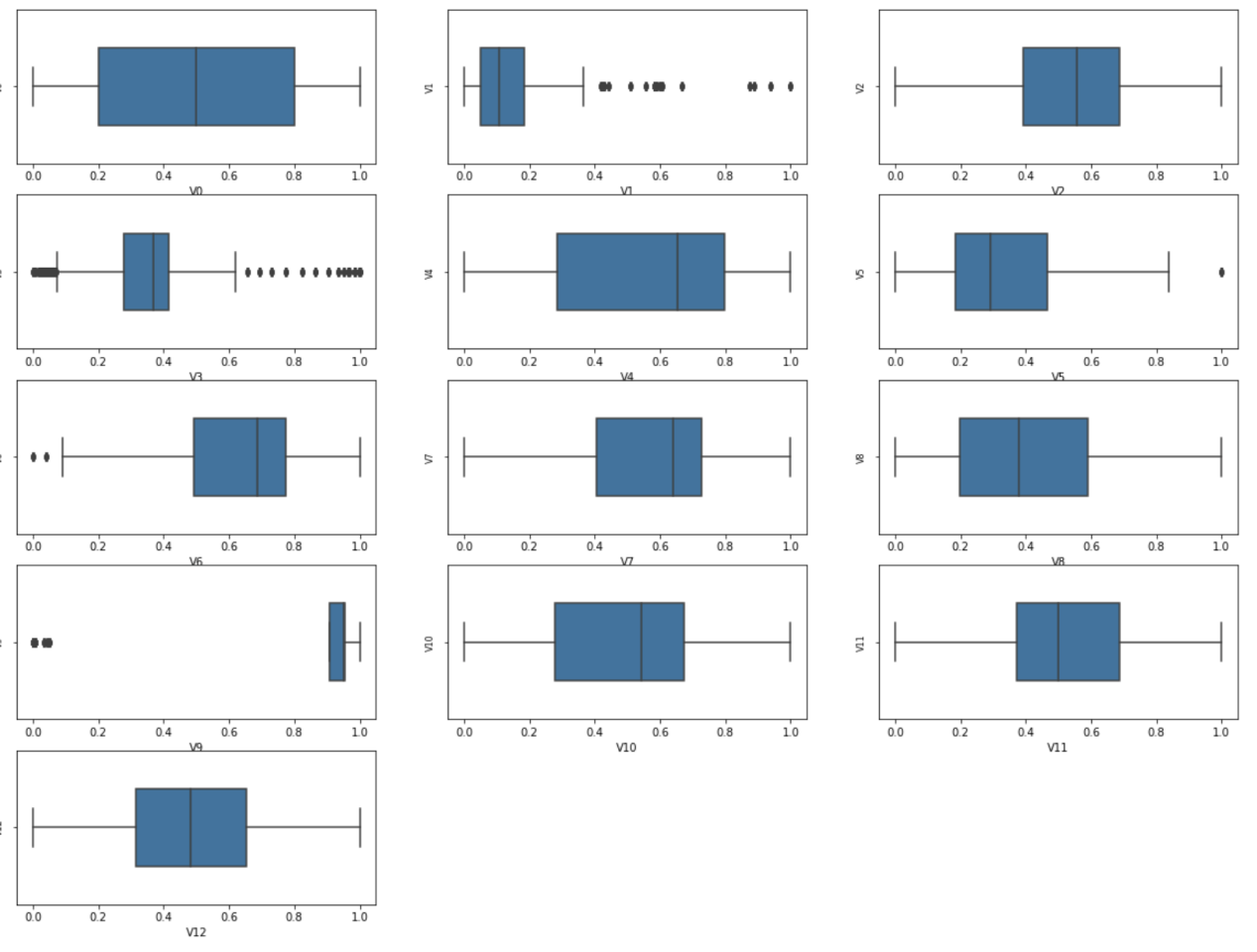


图5-6 箱式图

观察图5-6可以发现V1,V3有很多数据偏离正常值，处于四分位点以下，故这些数据存在许多较大的异常值，可以移除。

查看特征变量的数据分布直方图，并绘制Q-Q图查看数据是否近似于正态分布。

这是指数据的分位数和对比参照的图，如果数据符合正态分布则所有的点都会落在直线上。

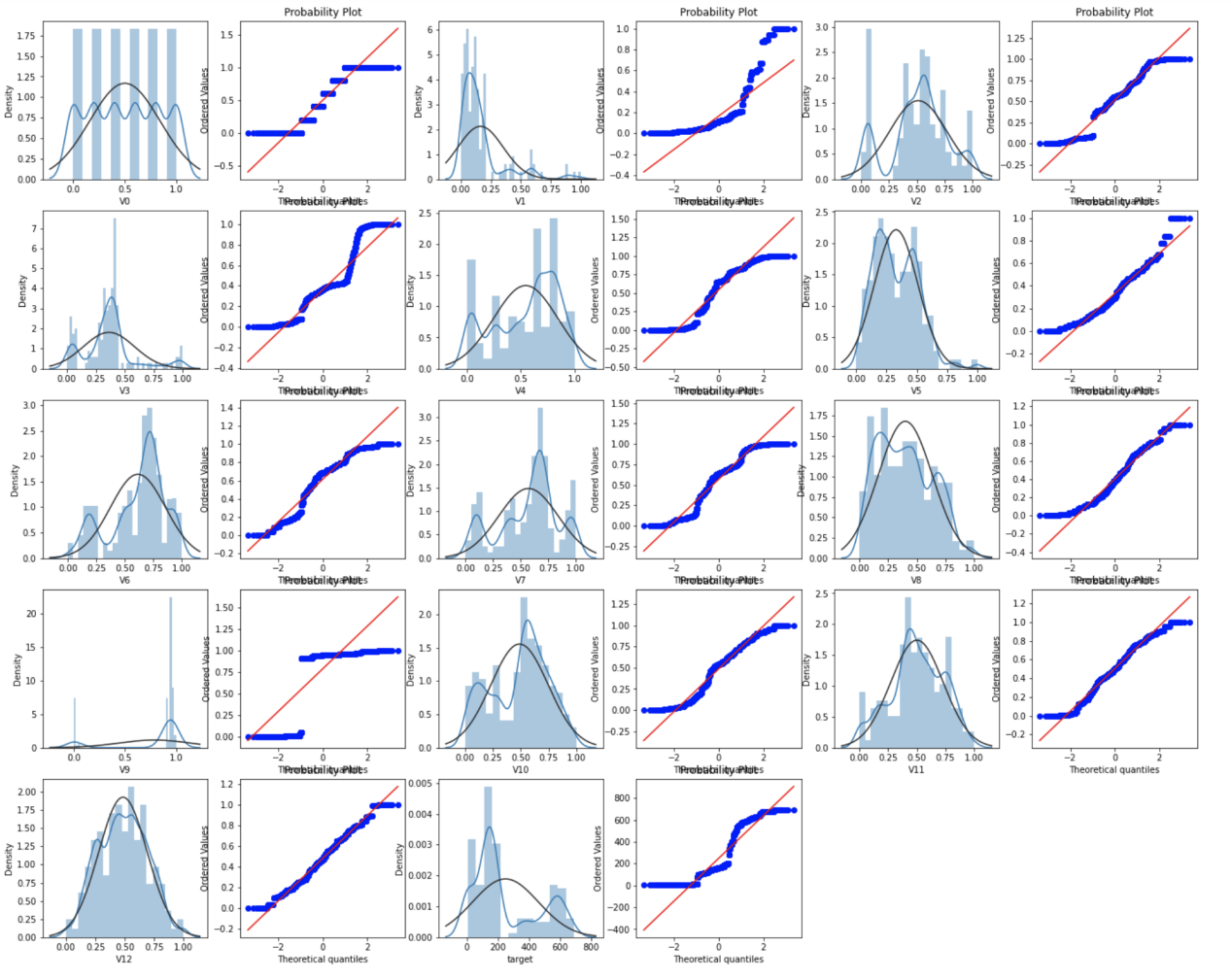


图5-7特征变量的数据分布直方图

如图5-7可以发现部分数据不随对角线分布，后续可以使用数据变换。

接下来对数据的相关性分析，可以发现特征变量和目标变量及特征变量之间的关系，为提取特征做准备，画出相关热力图，结果为所有特征变量和target变量之间的相关性系数，对角线数据一样。

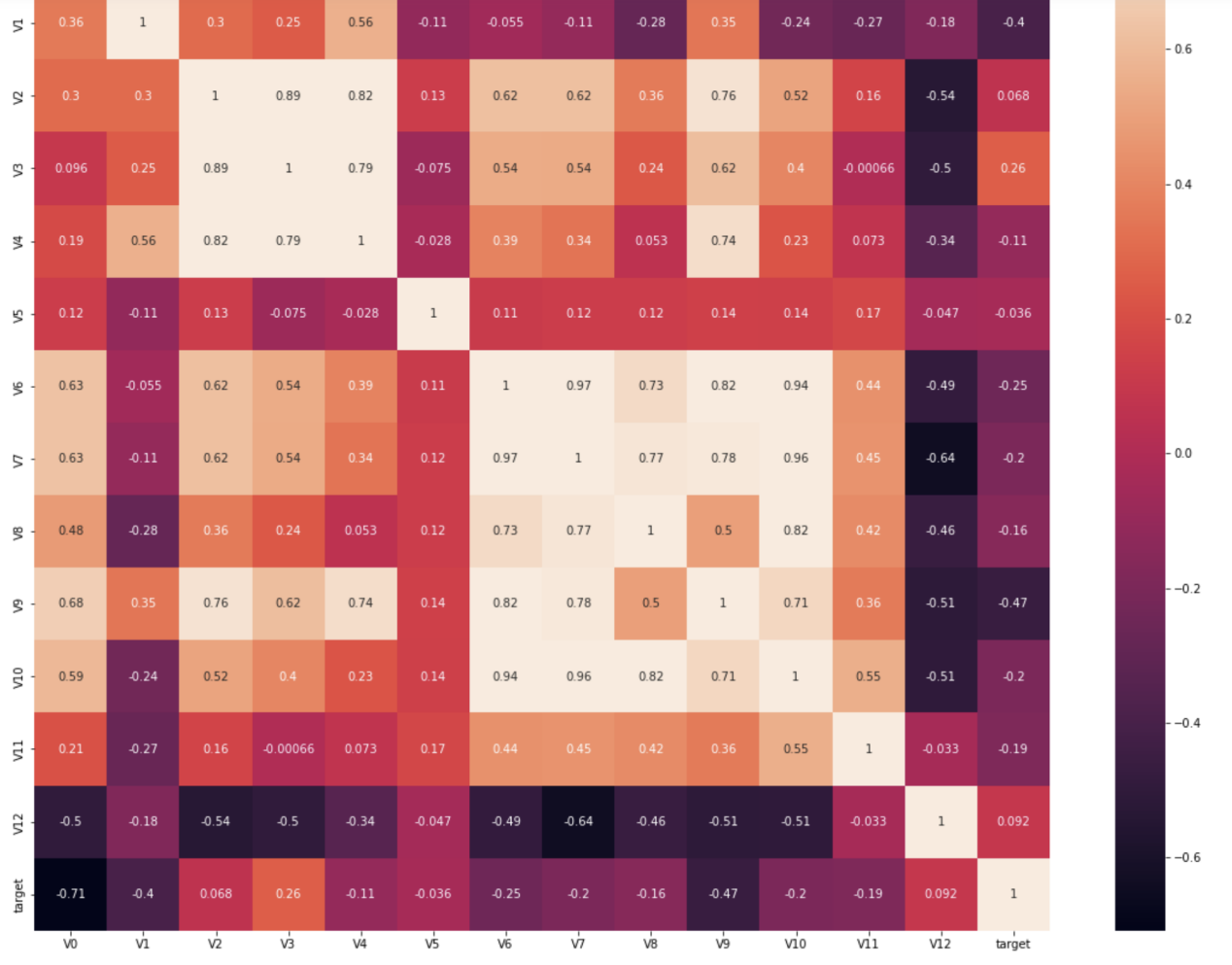


图5-8 热力图

只保留梯形热力图，如图5-9所示，找出相关度。

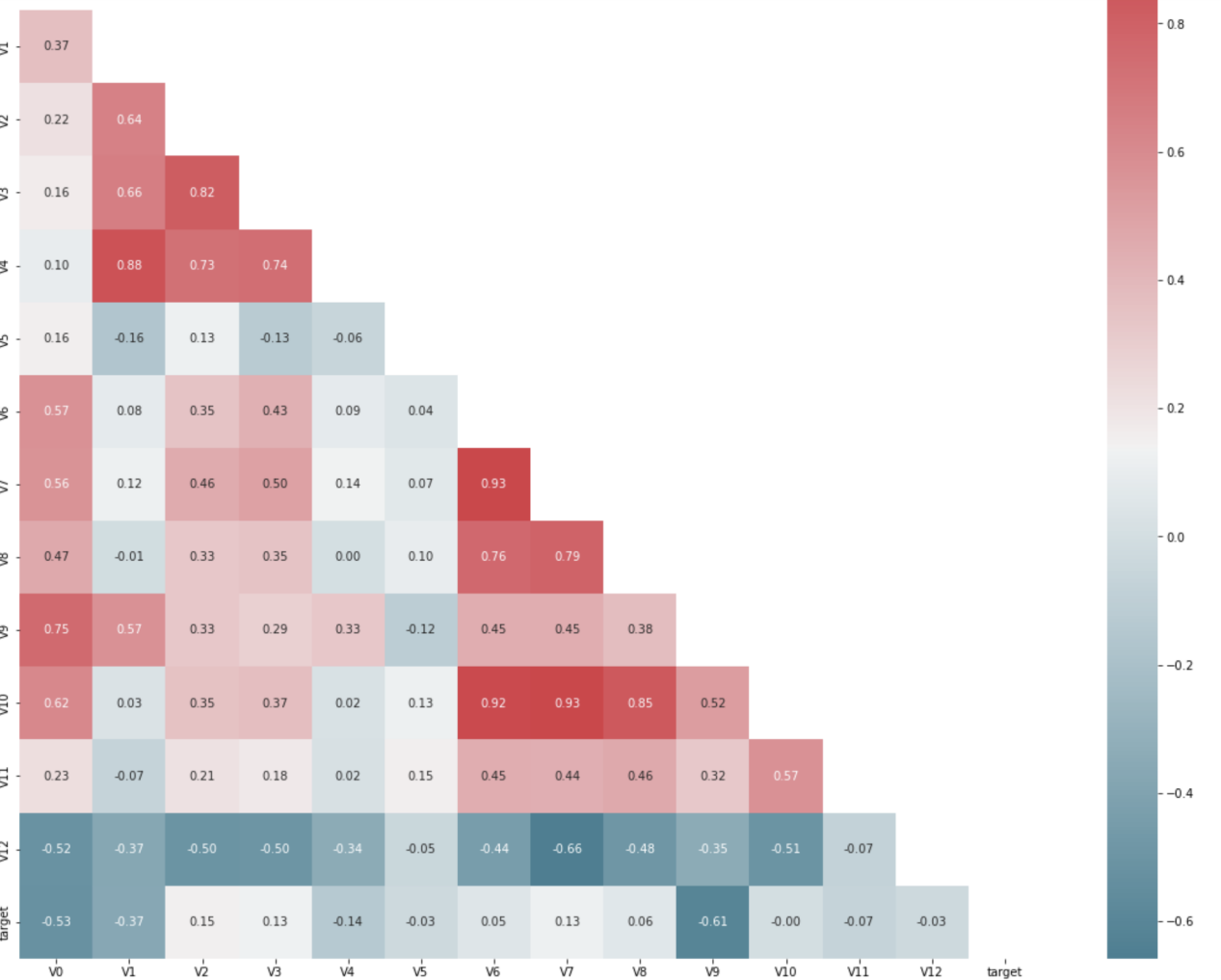


图5-9 梯形热力图

筛选出相关度大于0.1的特征，即会有相关性的特征，即使呈现弱相关。如图5-10所示。

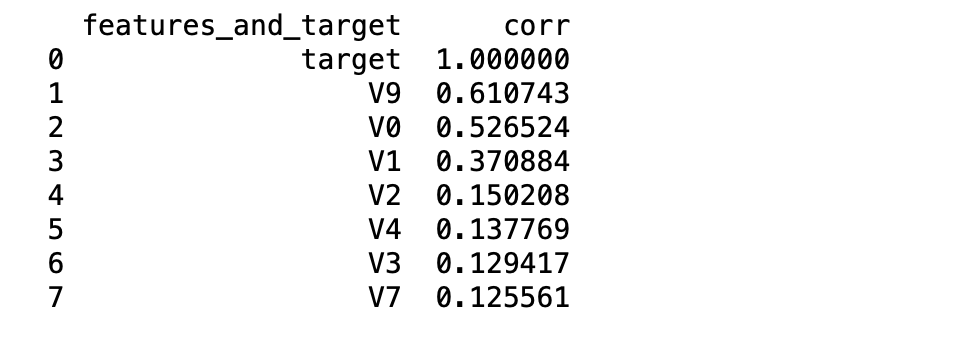


图5-10 相关性呈现

接下来划分数据集，分为训练集和测试集。

首先过滤异常值，此处过滤异常样本的基本思想为：利用岭回归先对train 数据进行建模，建模后的获取 x\_train的预测值y\_train\_predict，若真实值与预测值相减的绝对值 |y\_train-y\_train\_predict| >α\*y\_train.std()，则判定该样本为异常值；此处α参数经过多次测试，取了0.8 较为合适。

将其过程可视化如下：

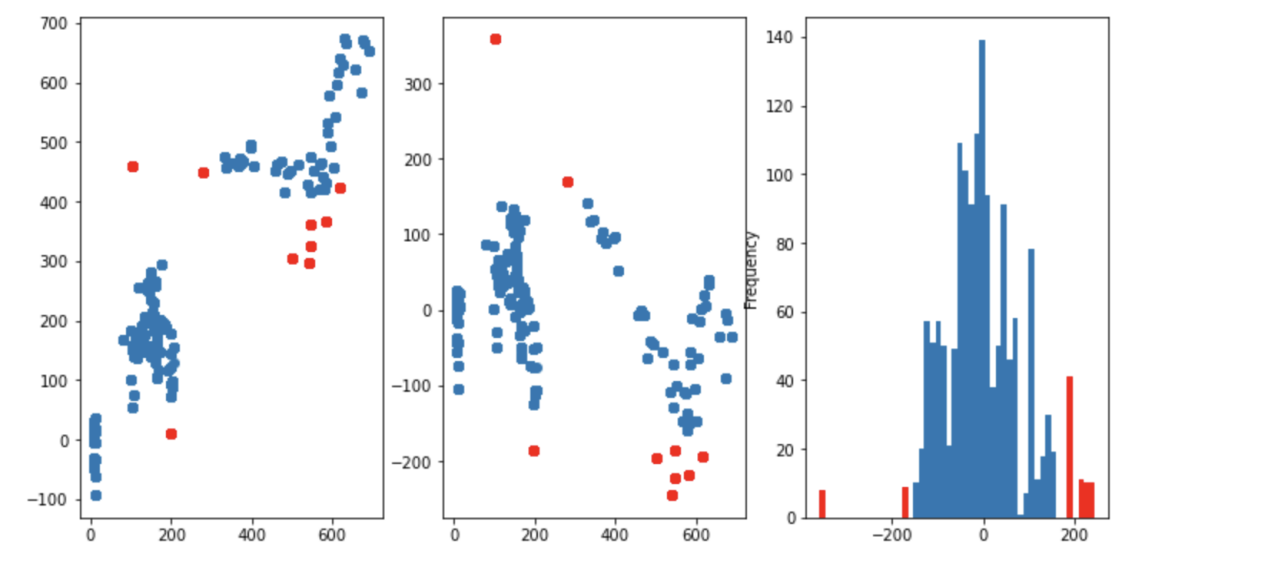


图5-11 岭回归预测

获取行索引，异常值过滤，如图5-12和5-13所示：

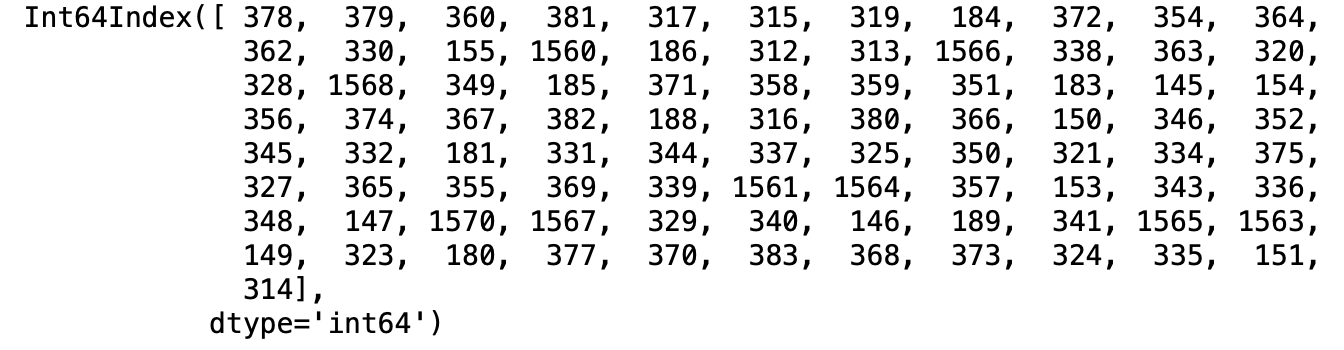


图5-12 异常值索引（1）

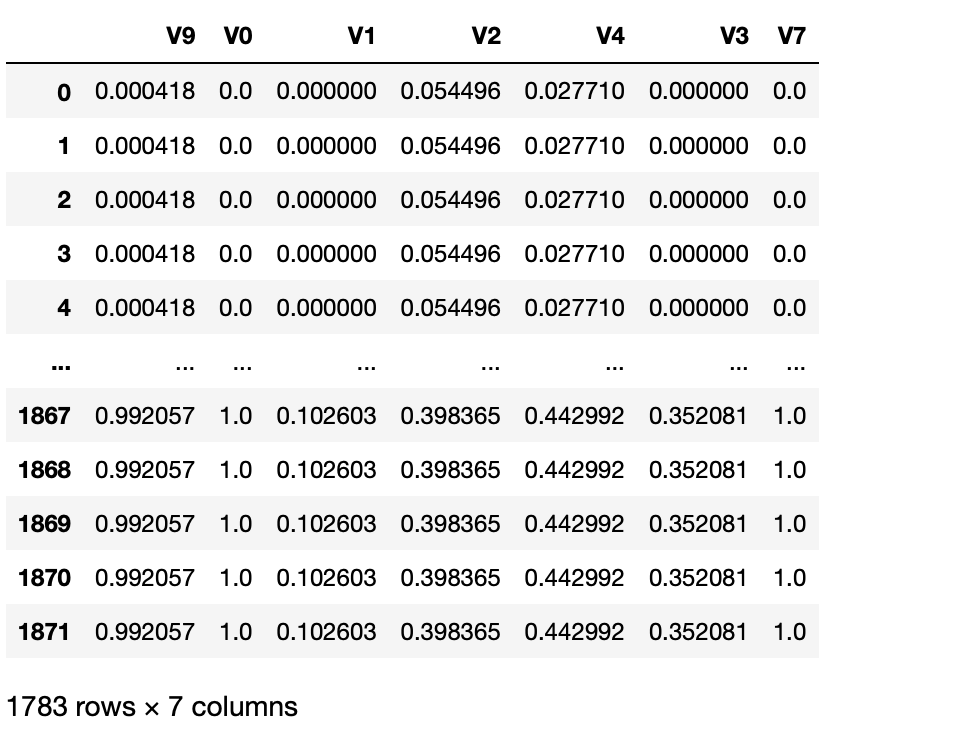


图5-12 异常值索引（2）

## 4.3 分析与结论

### 4.3.1信度分析

**1.水稻产量预测**

针对水稻的各地区的年总作物产量，我们使用spss软件进行了回归分析。以下对实验结果进行分析，对实验的结果进行解读。

**由结果得：**本研究纳入的自变量中，柴油使用量，冷冻灾成灾面积，近地面短波辐射，农用氮肥施用折纯量^2，农用氮肥施用折纯量^2，播种面积，温度，近地面风速，近地面空气湿度，近地面气压，近地面长波辐射，降水率，旱灾成灾面积，风雹灾受灾面积，旱灾受灾面积，水灾受灾面积，有效灌溉面积，农用氮肥施用折纯量，农用化肥施用折纯量的P值具有统计学意义，即这些因素是影响作物产量的主要因素。

理论模型：

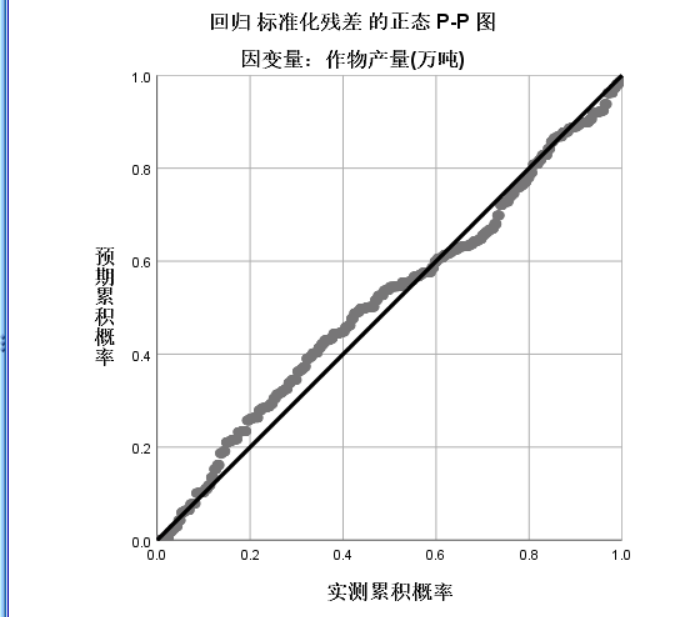
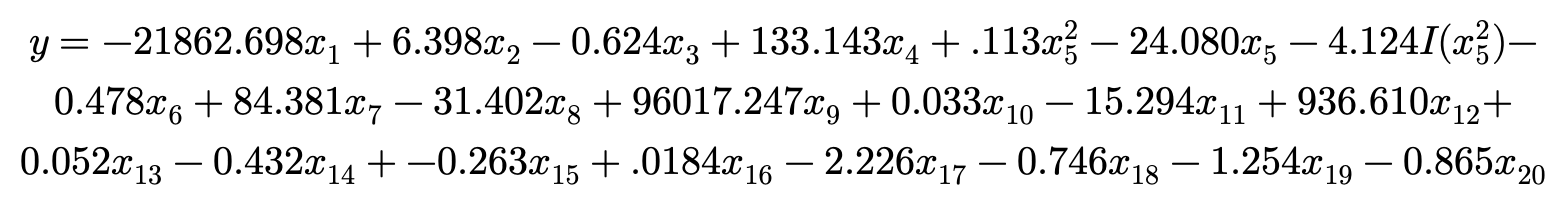


图5-13 回归标准化残差的正态P-P图（水稻）

由回归标准化残差的正态P-P可以看出：残差效果较好，所有的点大致都在一条直线上，可以认为残差符合正态分布的要求。

表5-2 ANOVA表（水稻）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 模型 | 平方和 | 自由度 | 均方 | F | 显著性 |
| 回归 | 125706653.6 | 22 | 5713938.801 | 477.885 | .000b |
| 残差 | 22107961.98 | 1849 | 11956.713 |  |  |
| 总计 | 147814615.6 | 1871 |  |  |  |

a.因变量：作物产量（万吨）

b.预测变量：（常量），农用化肥施用折纯量（万吨），近地面风速（米每秒），冷冻灾成灾面积（千公顷），风雹灾受灾面积（千公顷），水灾受灾面积（千公顷），近地面短波辐射（瓦每平方米）,降水率（毫米每小时），温度/K,冷冻受灾面积（千公顷），有效灌溉面积（千公顷），风雹灾成灾面积（千公顷），旱灾受灾面积（千公顷），农用化肥施用量折纯量^2（万吨^2），水灾成灾面积（千公顷），近地面气压（帕）,近地面空气湿度（千克^2）,农用氮肥施用折纯量^2（万吨^2），S近地面长波辐射（瓦每平方米），农用氮肥施用折纯量（万吨）

该表格可以用来检验模型的统计学意义，由结果可得F=477.885，P<0.001，因此拒绝原假设，认为回归模型通过了置信水平为0.05的F检验，即所拟合的方程具有统计学意义。

表5-3 变量解释程度与独立性检验（水稻）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 模型摘要b | | | | | |
| 模型 | R | R方 | 调整后R方 | 标准估算的错误 | 德宾-沃森 |
| 1 | .922a | .850 | .849 | 109.3467549 | 2.072 |

a.预测变量：（常量），农用化肥施用折纯量（万吨），近地面风速（米每秒），冷冻灾成灾面积（千公顷），风雹灾受灾面积（千公顷），水灾受灾面积（千公顷），近地面短波辐射（瓦每平方米）,降水率（毫米每小时），温度/K,冷冻受灾面积（千公顷），有效灌溉面积（千公顷），风雹灾成灾面积（千公顷），旱灾受灾面积（千公顷），农用化肥施用量折纯量^2（万吨^2），水灾成灾面积（千公顷），近地面气压（帕）,近地面空气湿度（千克^2）,农用氮肥施用折纯量^2（万吨^2），S近地面长波辐射（瓦每平方米），农用氮肥施用折纯量（万吨）

b.因变量：作物产量（万吨）

由表格可以看出，调整后的R方为0.849，即总体变量对因变量的解释程度达到84.9%。拟合效果较好，说明模型比较稳定；本研究的得宾-沃森检验值为2.072，一般来说，得宾-沃森检验值分布在0—4之间，越接近于2检验值互相独立的可能性越大。可以认为本研究中的观测值具有相互独立性。

**2.小麦产量预测**

针对小麦的各地区的年总作物产量，我们使用spss软件进行了回归分析。以下对实验结果进行分析，对实验的结果进行解读。

**由结果得：**本研究纳入的自变量中，播种面积，冷冻灾受灾面积，农用氮肥施用折纯量^2，农用化肥施用折纯量^2，降水率，近地面短波辐射，近地面风速，近地面空气湿度，近地面长波辐射，近地面气压，温度，冷冻灾成灾面积，旱灾成灾面积，水灾受灾面积，旱灾受灾面积，有效灌溉面积，农用化肥施用折纯量，柴油使用量，农用氮肥施用折纯量的P值具有统计学意义即：

播种面积，冷冻灾受灾面积，农用氮肥施用折纯量^2，农用化肥施用折纯量^2，近地面短波辐射，近地面风速，近地面空气湿度，近地面长波辐射，近地面气压，温度，冷冻灾成灾面积，旱灾成灾面积，水灾受灾面积，旱灾受灾面积，有效灌溉面积，农用化肥施用折纯量，柴油使用量，农用氮肥施用折纯量是影响作物产量的主要因素理论模型：

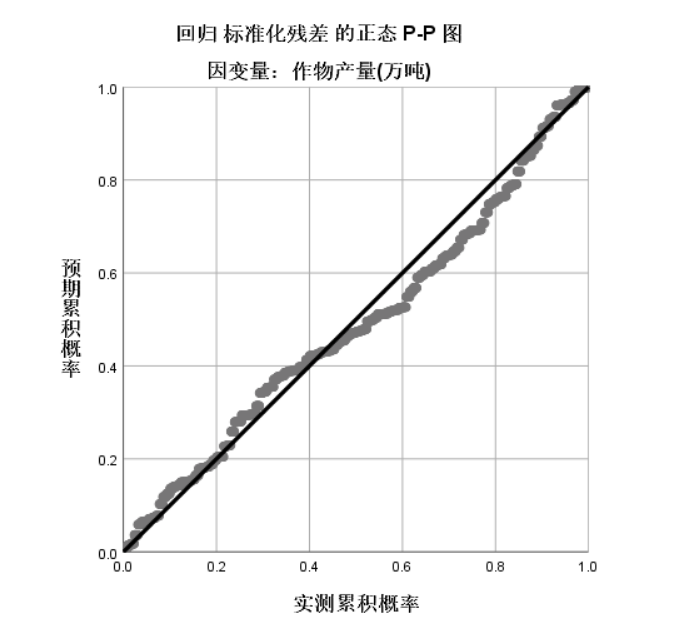
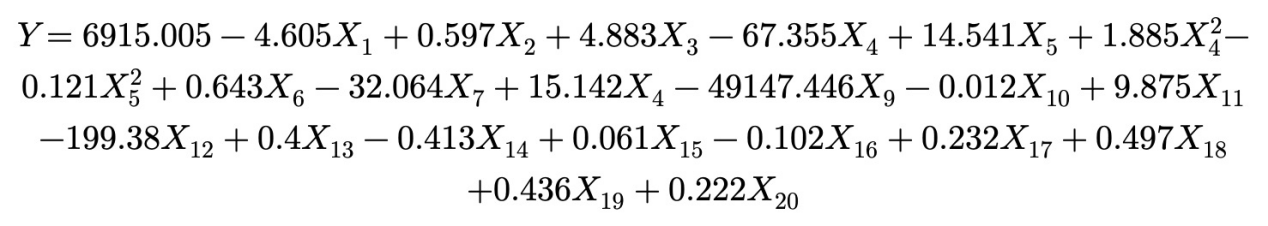


图5-14 回归标准化残差的正态P-P图（小麦）

由回归标准化残差的正态P-P可以看出：残差效果较好，所有的点大致都在一条直线上，可以认为残差符合正态分布的要求。

表5-4 ANOVA表（小麦）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 模型 | 平方和 | 自由度 | 均方 | F | 显著性 |
| 回归 | 18675750.67 | 22 | 848806.849 | 427.736 | .000b |
| 残差 | 3050049.507 | 1537 | 1984.417 |  |  |
| 总计 | 21723800.18 | 1559 |  |  |  |

a.因变量：作物产量（万吨）

b.预测变量：（常量），农用化肥施用折纯量（万吨），近地面风速（米每秒），冷冻灾成灾面积（千公顷），风雹灾受灾面积（千公顷），水灾受灾面积（千公顷），近地面短波辐射（瓦每平方米）,降水率（毫米每小时），温度/K,冷冻受灾面积（千公顷），有效灌溉面积（千公顷），风雹灾成灾面积（千公顷），旱灾受灾面积（千公顷），农用化肥施用量折纯量^2（万吨^2），水灾成灾面积（千公顷），近地面气压（帕）,近地面空气湿度（千克^2）,农用氮肥施用折纯量^2（万吨^2），S近地面长波辐射（瓦每平方米），农用氮肥施用折纯量（万吨）

该表格可以用来检验模型的统计学意义，由结果可得F=427.736，P<0.001，因此拒绝原假设，认为回归模型通过了置信水平为0.05的F检验，即所拟合的方程具有统计学意义。

**3. 变量解释程度与独立性检验**

表5-5 变量解释程度与独立性检验

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 模型摘要b | | | | | |
| 模型 | R | R方 | 调整后R方 | 标准估算的错误 | 德宾-沃森 |
| 1 | .927a | .860 | .858 | 44.54679984 | 2.081 |

a.预测变量：（常量），农用化肥施用折纯量（万吨），近地面风速（米每秒），冷冻灾成灾面积（千公顷），风雹灾受灾面积（千公顷），水灾受灾面积（千公顷），近地面短波辐射（瓦每平方米）,降水率（毫米每小时），温度/K,冷冻受灾面积（千公顷），有效灌溉面积（千公顷），风雹灾成灾面积（千公顷），旱灾受灾面积（千公顷），农用化肥施用量折纯量^2（万吨^2），水灾成灾面积（千公顷），近地面气压（帕）,近地面空气湿度（千克^2）,农用氮肥施用折纯量^2（万吨^2），S近地面长波辐射（瓦每平方米），农用氮肥施用折纯量（万吨）

b.因变量：作物产量（万吨）

由表格可以看出，调整后的R方为0.858，即总体变量对因变量的解释程度达到85.8%。拟合效果较好，说明模型比较稳定；本研究的得宾-沃森检验值为2.081，一般来说，得宾-沃森检验值分布在0—4之间，越接近于2检验值互相独立的可能性越大。可以认为本研究中的观测值具有相互独立性。

### 4.3.2 实践案例仿真

以武汉市彭盛源家庭农场为例，经小云鹰系统测得土地面积大小为273.3亩，2020年当地年均亩投入为1352元，预测年产量籼稻共996千克每亩，预估净收入三十一万六千余元，违约距离1.823，预期违约率3.416%。

根据精算模型计算，在损失赔付率为90%、赔付触发条件为综合收入下降超10%的条件下，我们给出的参考保费为每亩30.4元，总保费为8308.3元。

经项目组成员实地考察，与农场主沟通确认实际土地面积268.2亩，2020年均亩投入为1324元，实际产量约976千克每亩，净收入三十万七千余元，经验证发现误差在百分之三左右。

# 作品总结及开发感想

本项目旨在为农村弱信用群体征信提供合理解决方式，最终以基于计算机视觉和机器学习的遥感农业大数据新应用的形式呈现。在开发过程中，团队成员齐心协力紧密合作，多次实现跨领域跨学科式的突破自我，汲取相关知识。在巩固和掌握原有计算机技术的基础上，学习了机器学习、web开发、ui设计等新技术，并运用到实践中。在各个学科交叉领域跨域交流的过程中，提升了自己的交流能力，学习到了其他学科思维和解决问题的方法。在团队合作的过程中虽然有过争吵，但更多的是无数思维火花的碰撞，是一起熬夜也要解决问题的友谊。

目前本项目还有许多有待实现的拓展功能：针对商业模式，开辟新的To C方业务；扩充自有数据集，改良算法与模型等等；在商业推广方面运营自媒体，通过线上线下相结合的营销方式实现裂变式吸纳流量。这次项目开发经历令我们受益匪浅，或许比赛只是我们的起点，我们的团队合作、团队精神永远不会改变。我们将践行自强弘毅、求是拓新的精神，将更多的精力投入后续功能开发以及性能提升中，与小云鹰一起不断进步。