# 第三章 技术方案

## 1.系统概述

### 1.1技术方案概述

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。

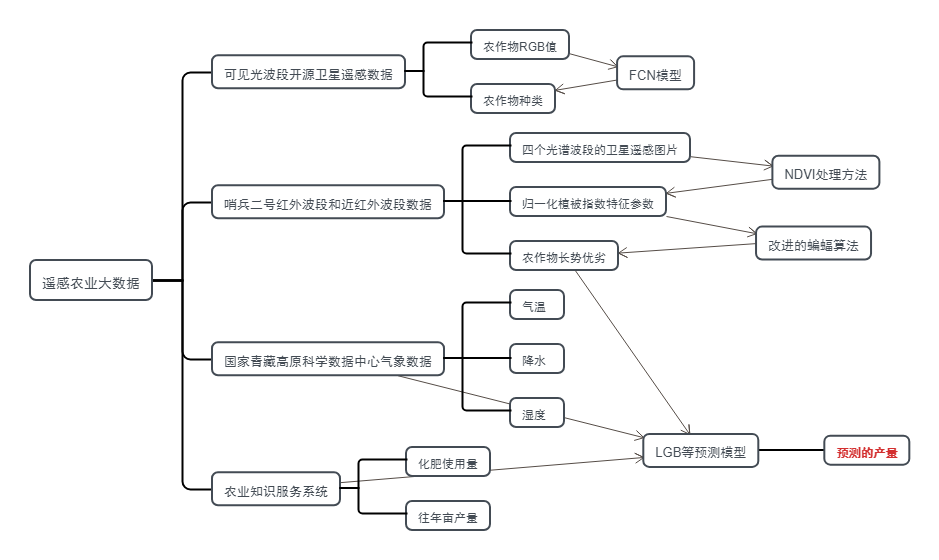
紧接着，我们从国家青藏高原科学数据中心得到实时的气象数据如气温、降水、湿度等，从农业知识服务系统搜索得到化肥使用量和往年亩产量的数据，与我们得到的农作物长势共同作为变量输入LGB模型建立最优预测产量模型，并运用深度学习的VGG、残差网络模型对比验证精确度。

图1-1 技术路线思维框架图

### 1.2 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。

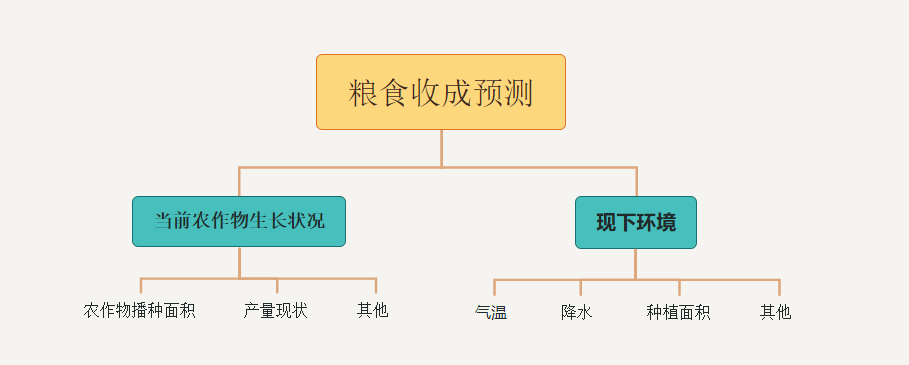


图1-2 收成预测影响因素

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1-1 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1-2 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。

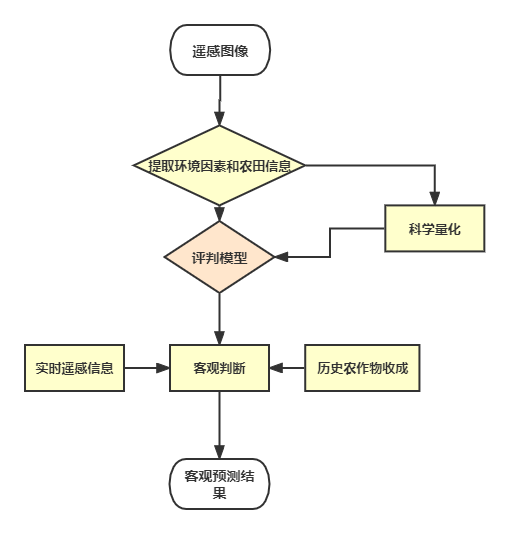


图1-3 农作物收成预测流程图

## 运行环境

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。

## 3.软件架构及说明

### 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。

### 3.2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）

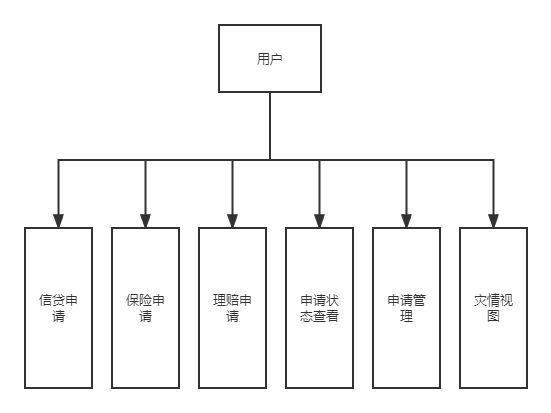


图1-4 用户功能模块

b.后台管理人员：用户申请管理、用户管理。（如图1-5）

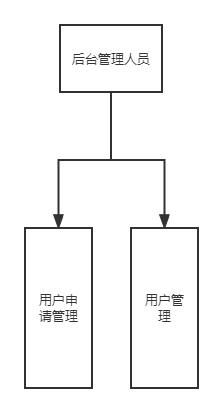


图1-5 管理人员功能模块

c.系统业务流程分析（如图1-6）

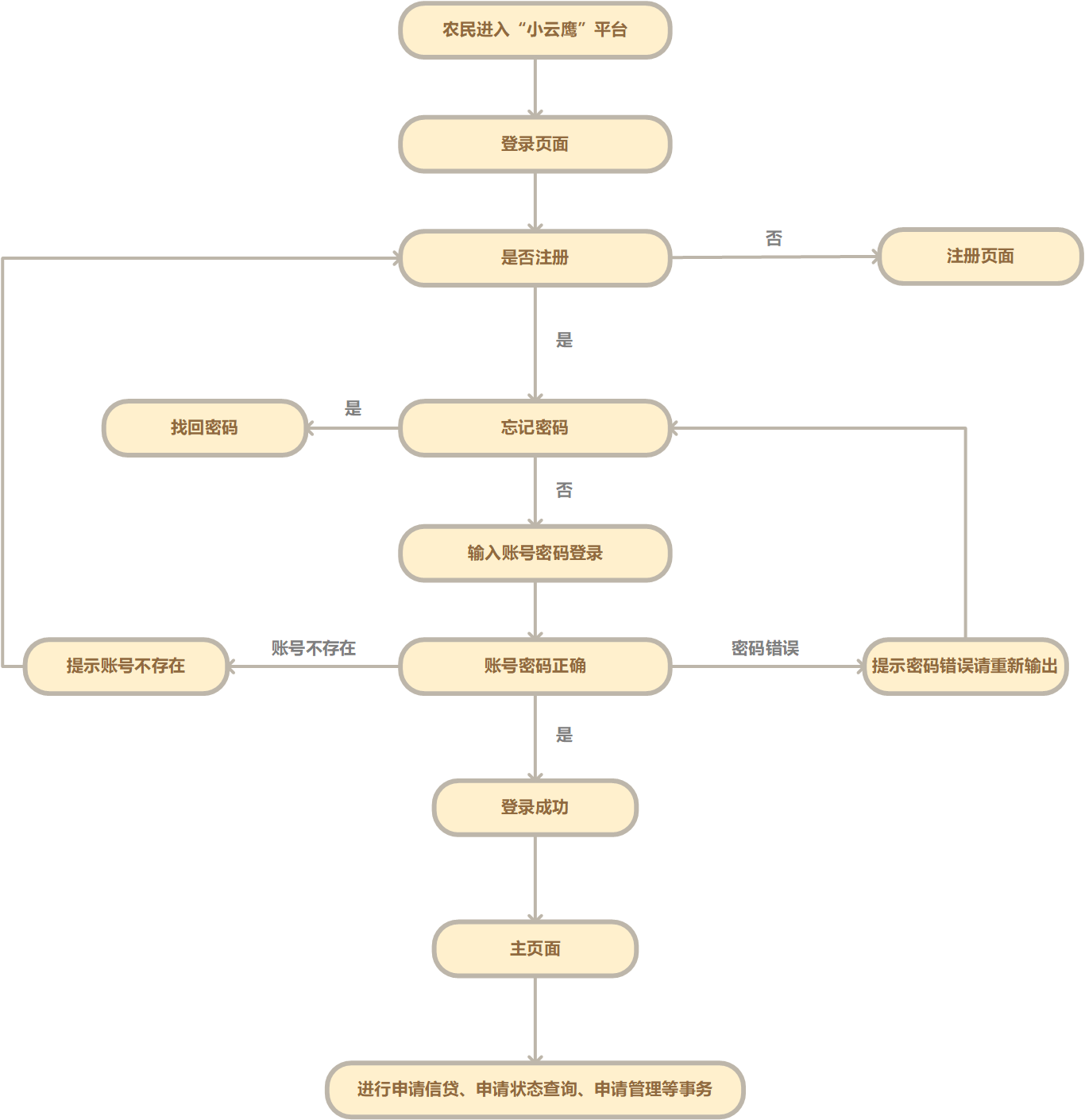


图1-6 用户登录及服务使用流程

d.平台业务流程图（1-7）

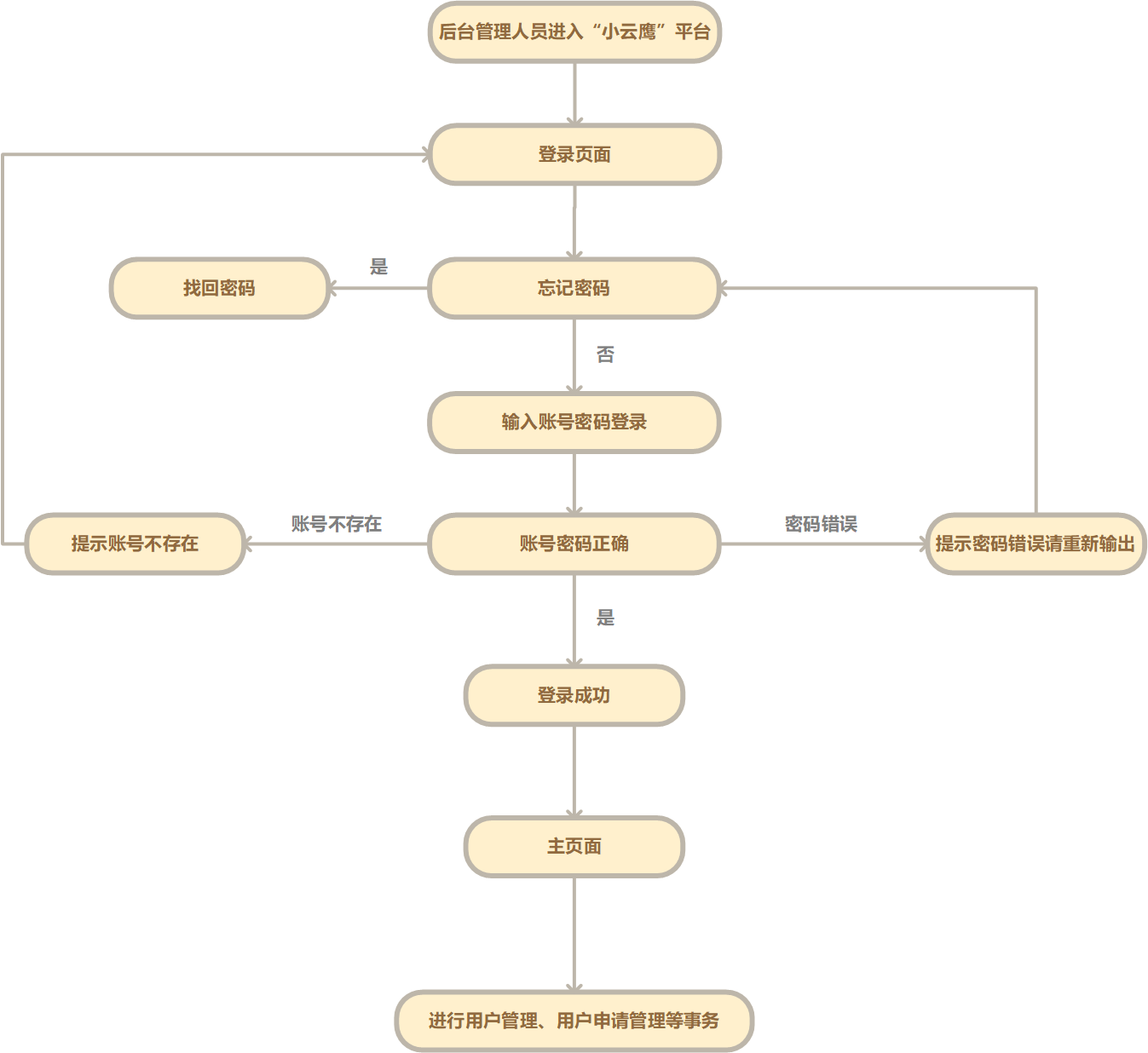
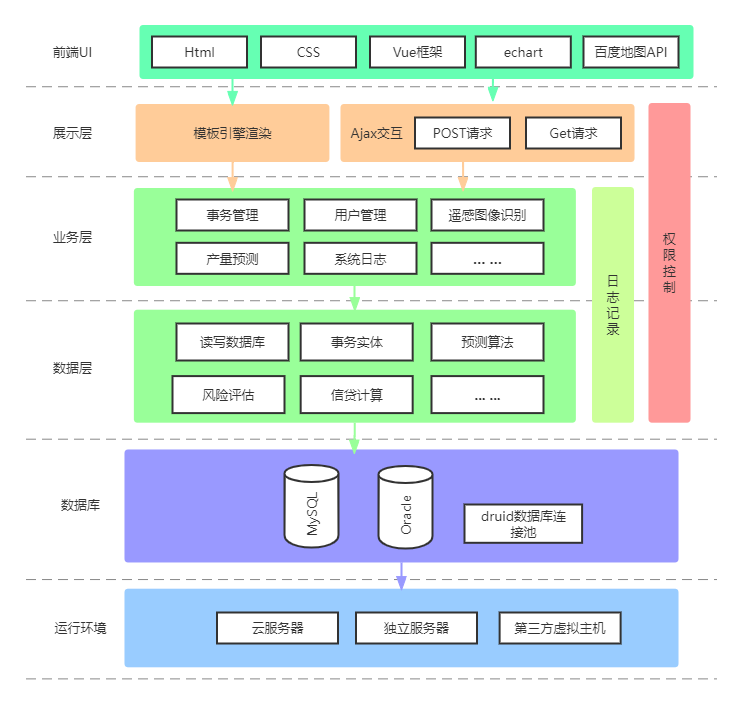


图1-7 后台管理人员登录及系统操作流程

## 4.硬件架构及说明



## 5.关键技术与算法

### 5.1 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1-5 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1-6 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。

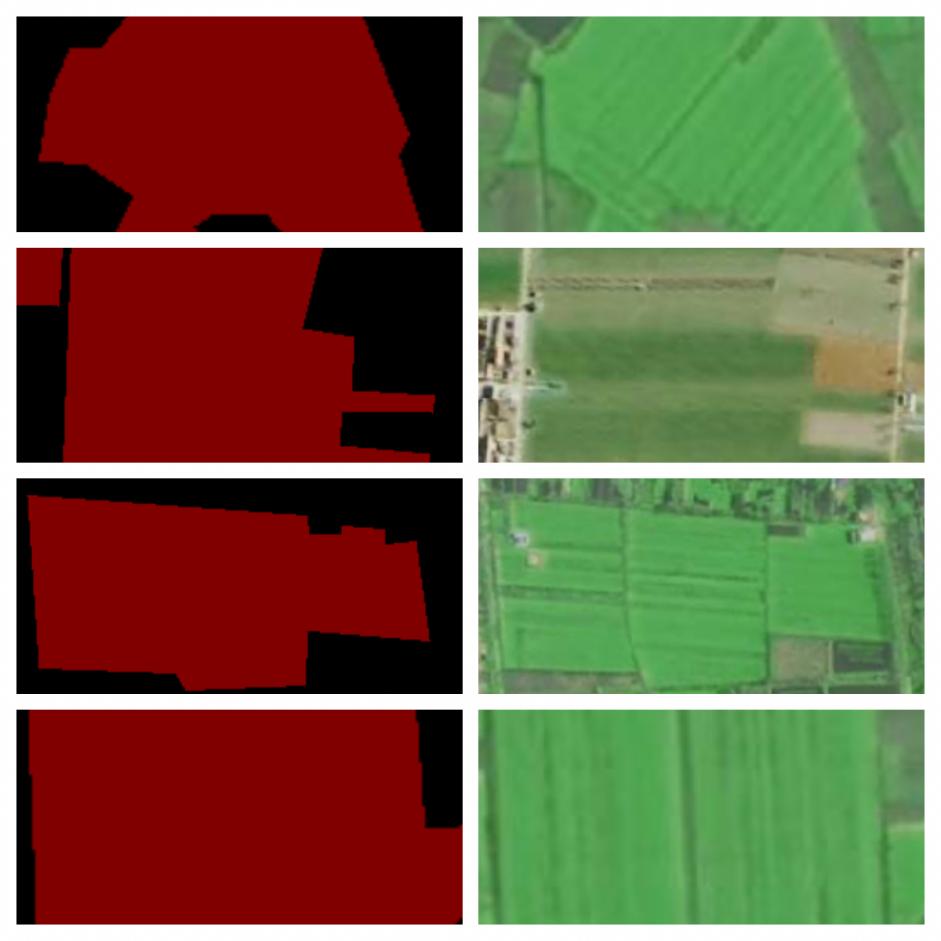


图1-9 遥感图像数据加强对比效果

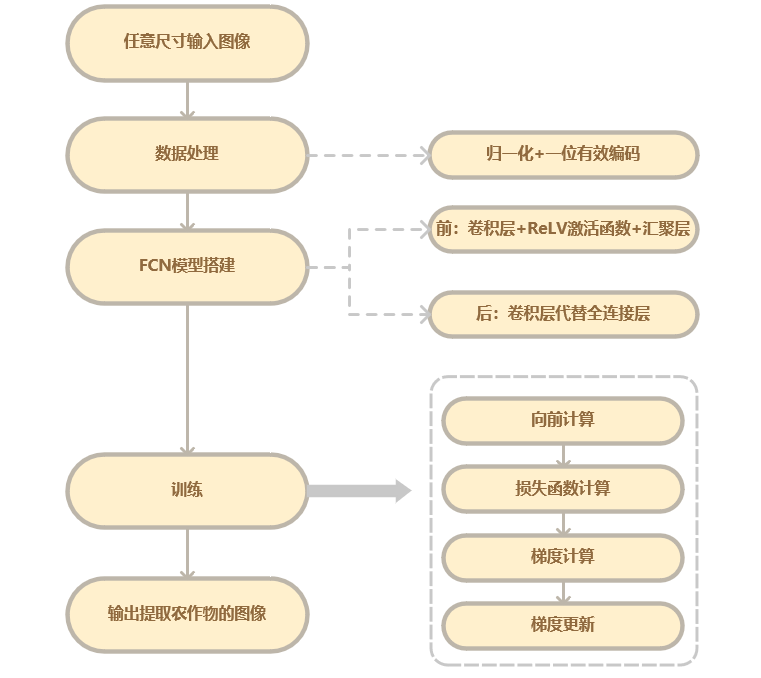


图1-10 语义分割提取农作物流程图

### 5.2 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。

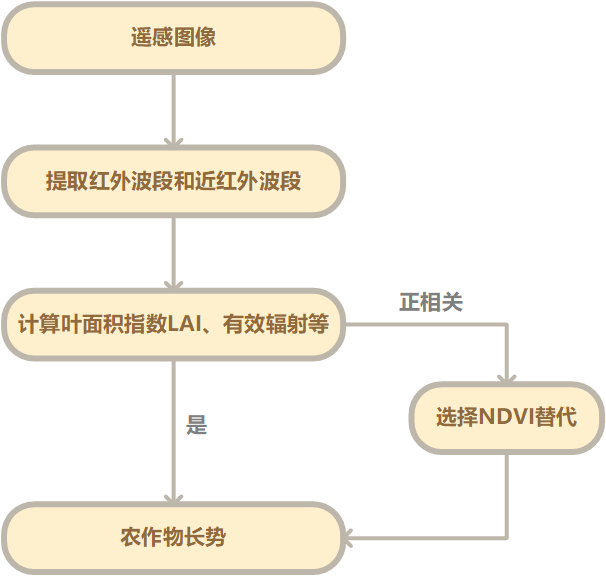


图1-11 NDVI分析流程图

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1-12 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1-13 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：

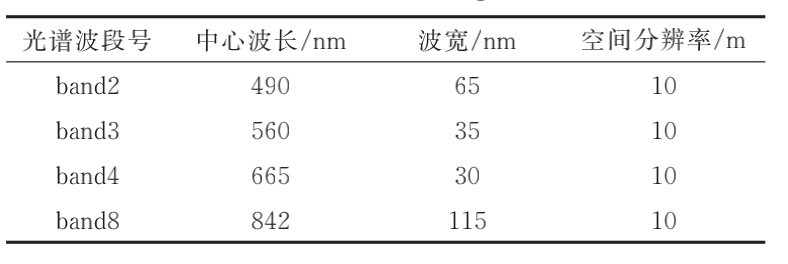


图1-14 光谱波段号对应数据

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。

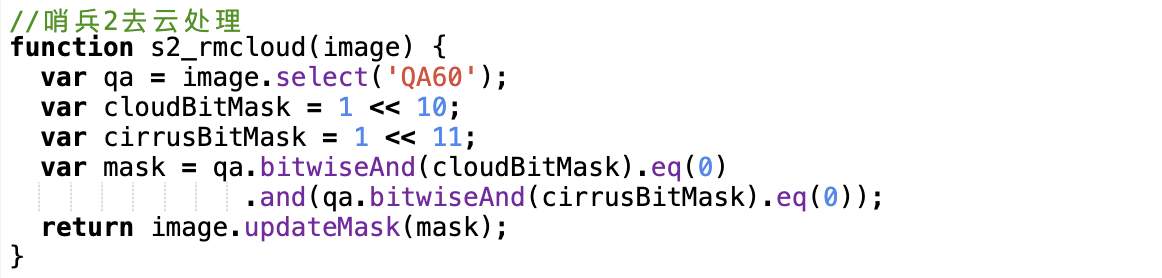


图1-15卫星光谱图去云处理核心代码



图1-16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。

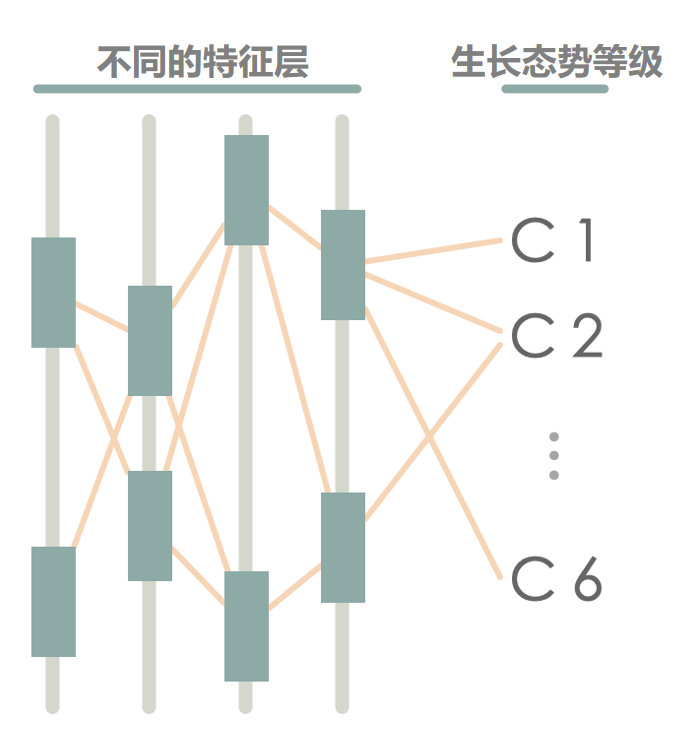


图1-17 B-E规则挖掘原理

* 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。

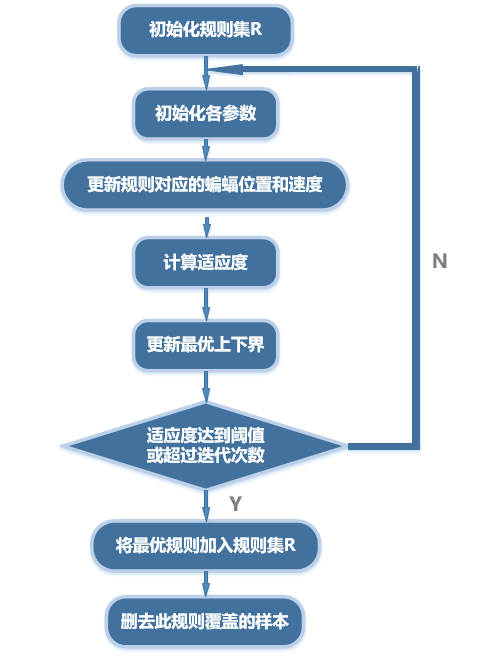


图1-18 蝙蝠算法

* 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1-7 农作物影像经过一系列图像处理、分析所得的最终结果



### 5.3 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。

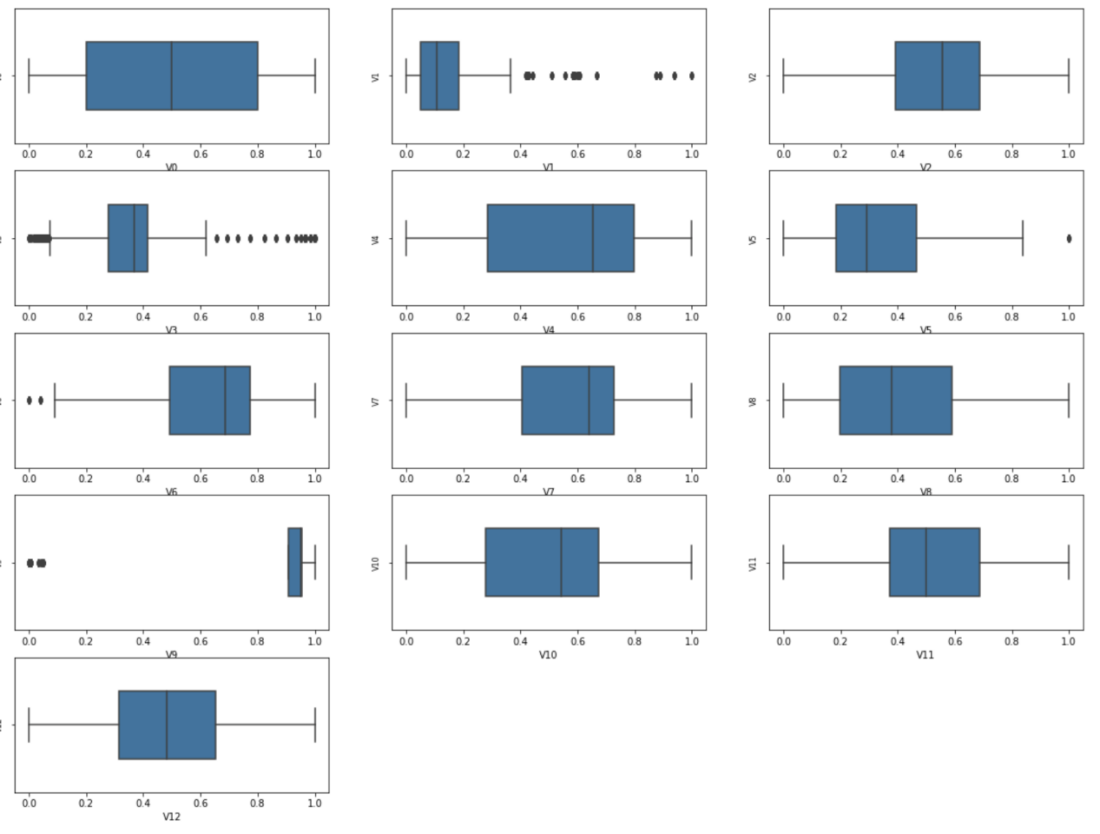


图1-19特征箱式图

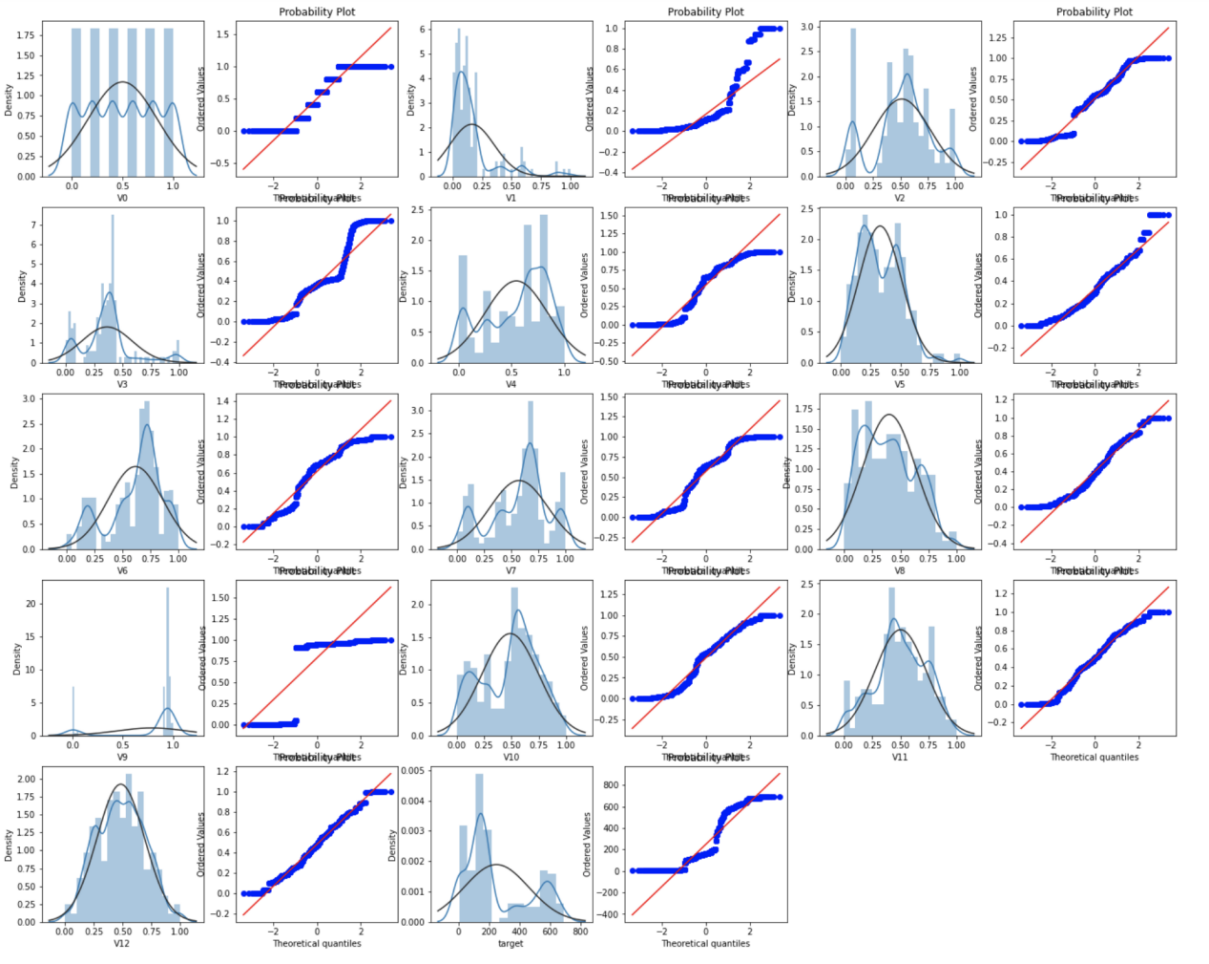


图1-20 数据分布直方图和Q-Q图

在对数据的相关性分析中，画出梯形热力图（图1-21），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。

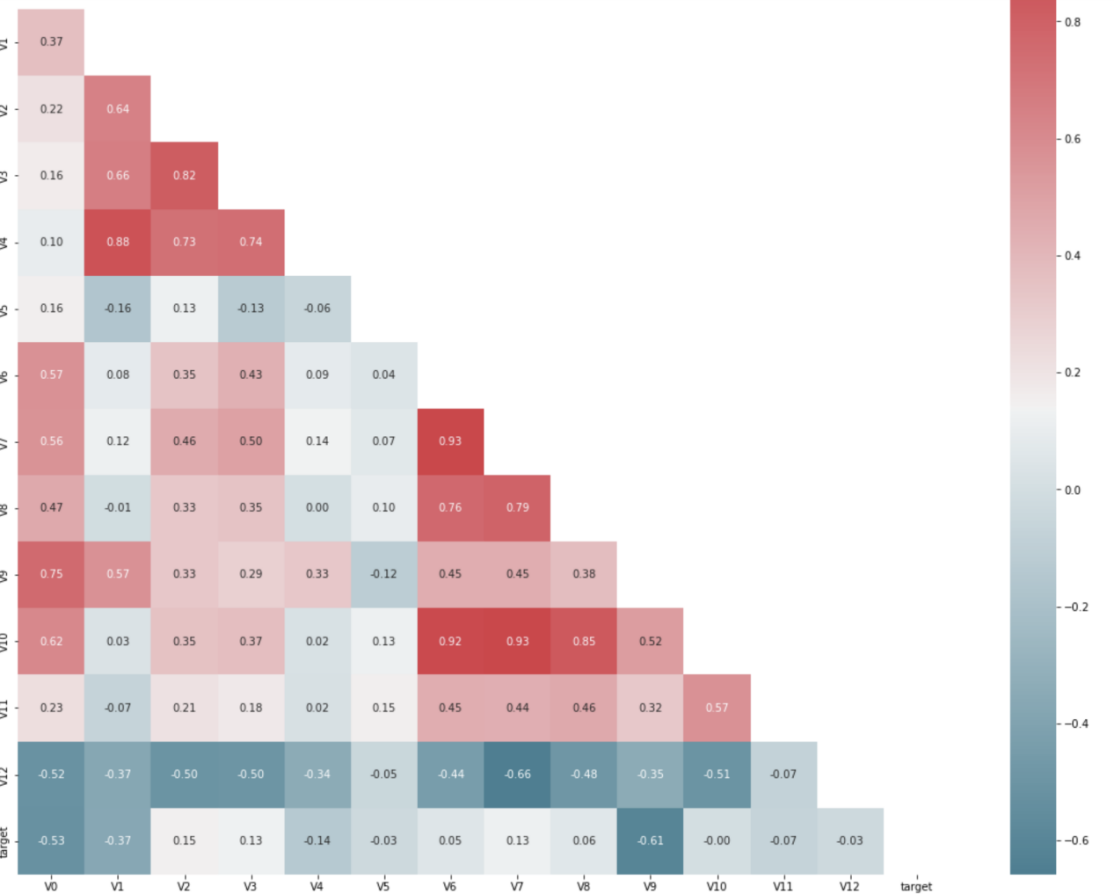


图1-21 热力梯度图

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。

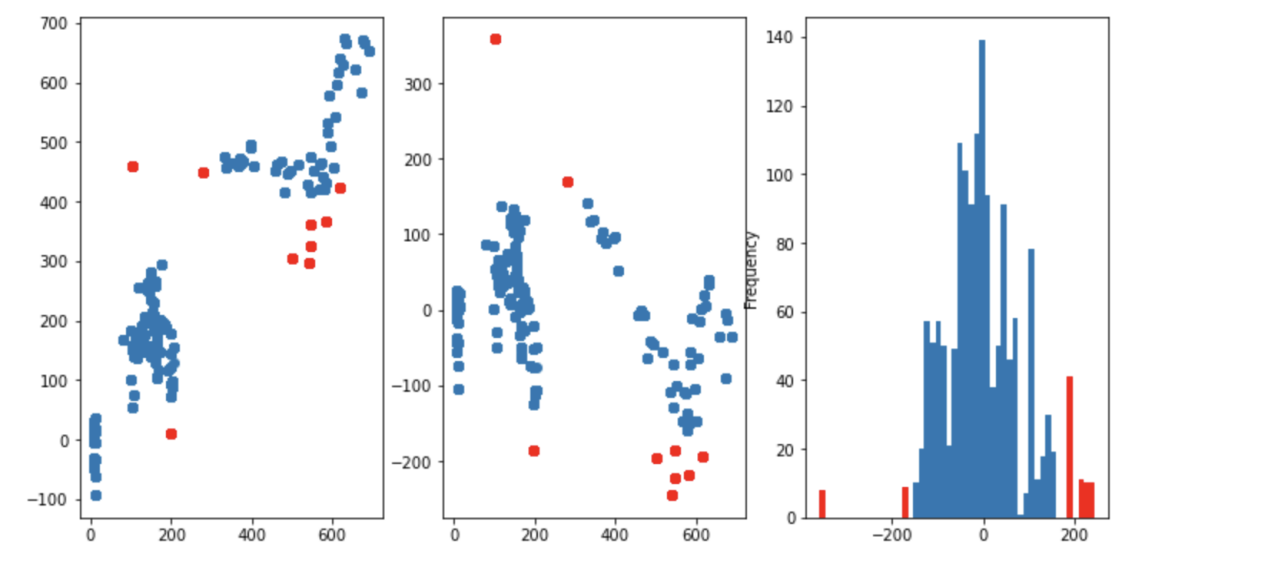
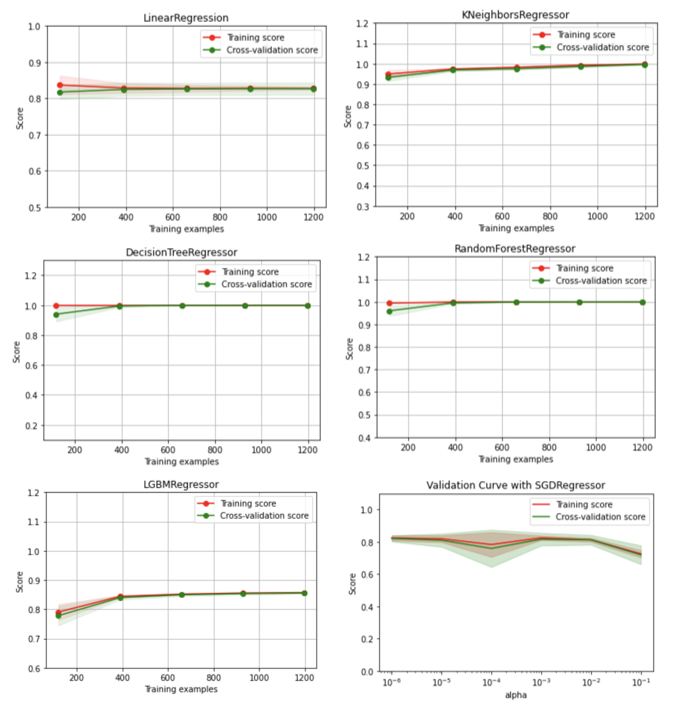


图1-22 岭回归样本异常值可视图

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1-23）。

图1-23 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。

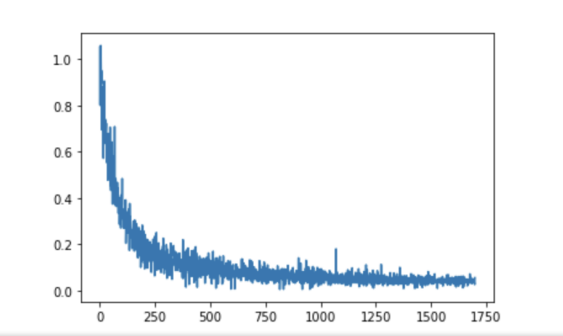


图1-24 模型训练过程



图1-25 产量预测流程图

## 系统功能设计

### 6.1 主界面

小云鹰涉农平台首页界面如下图所示。



图1-26 小云鹰涉农平台主界面UI

### 6.2 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1-27 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1-28 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1-29 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1-30 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1-31 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1-32 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1-33 理赔申请视图

## 7.数据库设计

### 7.1 遥感图像数据集

表1-8



### 7.2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1-8的E-R模型图。

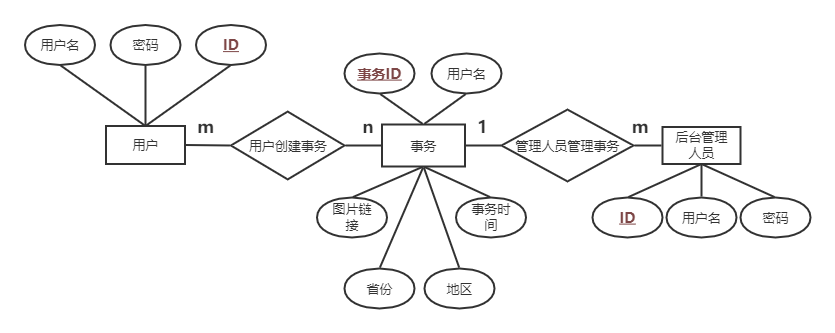


图1-8 E-R图设计

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1-3 transaction表

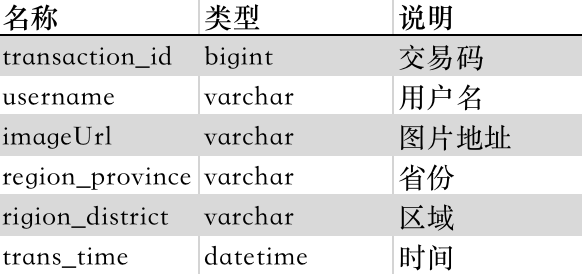


表1-4 coordinate表

