

文章编号: 2095-2163(2020)12-0170-04

中图分类号: TP311; TP319

文献标志码: A

基于大数据与图像识别的柑橘病虫害防治系统的设计与实现

罗昱晟, 冉力争, 张小荣, 王 敏

(四川工商学院 计算机学院, 成都 611745)

摘 要: 柑橘作为世界性大宗水果, 每年受病虫害影响巨大。本系统采用 Yolov4 与 Mask-RCnn 框架, 实现柑橘病虫害智能识别功能; 采用 Hadoop 分布式计算平台对大数据进行挖掘分析, 实现基于大数据的病虫害爆发预警分析、智能解决方案推送; 并使用前端技术实现相关数据的可视化展示。该系统解决了病虫害快速识别、柑橘病虫害爆发预警分析、解决方案智能推送等问题, 在较大程度上减少果农损失, 同时推进了农业生产的信息化、智能化的发展。

关键词: Yolov4; Mask-RCnn; Hadoop; 大数据; 柑橘

Design and implementation of diseases and pests prophylaxis and treatment system of citrus based on big data and image recognition

LUO Yusheng, RAN Lizheng, ZHANG Xiaorong, WANG Min

(School of computer, Sichuan Technology and Business University, Chengdu 611745, China)

[Abstract] As the one of world's staple fruits, the production of citrus is heavily affected by diseases and pests. The system adopts the artificial intelligence trained by Yolov4 and Mask-RCNN framework which Realized intelligent identification of citrus pests and diseases; the system adopts Hadoop to mine and analyze big data witch offers distributed cloud storage management, Realized warning analyze of diseases and pests outbreak and the solution offers based on big data; And the system realized the visual presentation of relevant data based on front-end technology. The system can solve the problems includes diseases and pests of citrus fast recognition, warning analyze of diseases and pests outbreak, the solution offers and so on. The system can Minimize losses of growers and it has promoted the developments of informatization and intelligentization of agricultural production at the same time.

[Key words] Yolov4; Mask-RCNN; Hadoop; Big data; Citrus

0 引言

柑橘作为全球第三大粮食作物和第一大类水果, 因受到病虫害的侵害每年造成巨大的经济损失。主要问题在于人工鉴别病虫害具有不及时、不准确的固有缺陷; 果农缺乏专家指导与专业技术; 果农针对柑橘病虫害没有预防意识, 长期采取见虫就打, 见病就救治的态度来对待。

目前, 对于柑橘病虫害识别与大数据信息化防治策略方面研究相对落后, 大多数研究项目停留在病虫害类型认知上, 对于病虫害爆发没有一套基于此的预警预备系统和专业具体的解决方案推送系统。

将人工智能图像识别技术应用于柑橘病虫害识别, 有助于帮助农户快速识别病虫害, 并结合对应的解决方案, 及时治理病虫害减少损失; 将大数据分析技术应用于柑橘病虫害预警分析, 有助于帮助农户

进行病虫害的提前防治, 从源头上减少病虫害威胁, 实现增产增收^[1-2]。

1 系统总体架构设计

本系统总体架构设计如图 1 所示, 主要分为 6 层: 用户交互层、接口层、平台业务层、数据存储层、平台支撑层和硬件设施层。

1.1 用户交互层

用户交互层, 由前端用户输入和输出二部分组成。输入方面由用户提交的图片信息及手机指令相关信息组成; 输出由病虫害信息、智能推送的解决方案、病虫害预警分析结果和数据可视化展示等。

1.2 接口层

接口层, 统一为 Java/Scala 的 API 接口。由于 Hadoop 中使用到 Spark 进行数据处理, 所以为了更好的交互, 需要使用 Scala 的 API 接口接入; 前端数据可视化以及信息传输则使用基础 Java API 接口。

作者简介: 罗昱晟(2000-), 男, 本科生, 主要研究方向: 云计算、大数据; 冉力争(2000-), 男, 本科生, 主要研究方向: 软件工程; 张小荣(1998-), 女, 本科生, 主要研究方向: 云计算; 王 敏(1984-), 男, 硕士, 副教授, 主要研究方向: 人工智能、数据挖掘。

通讯作者: 王 敏 Email: 172128788@qq.com

收稿日期: 2020-09-27

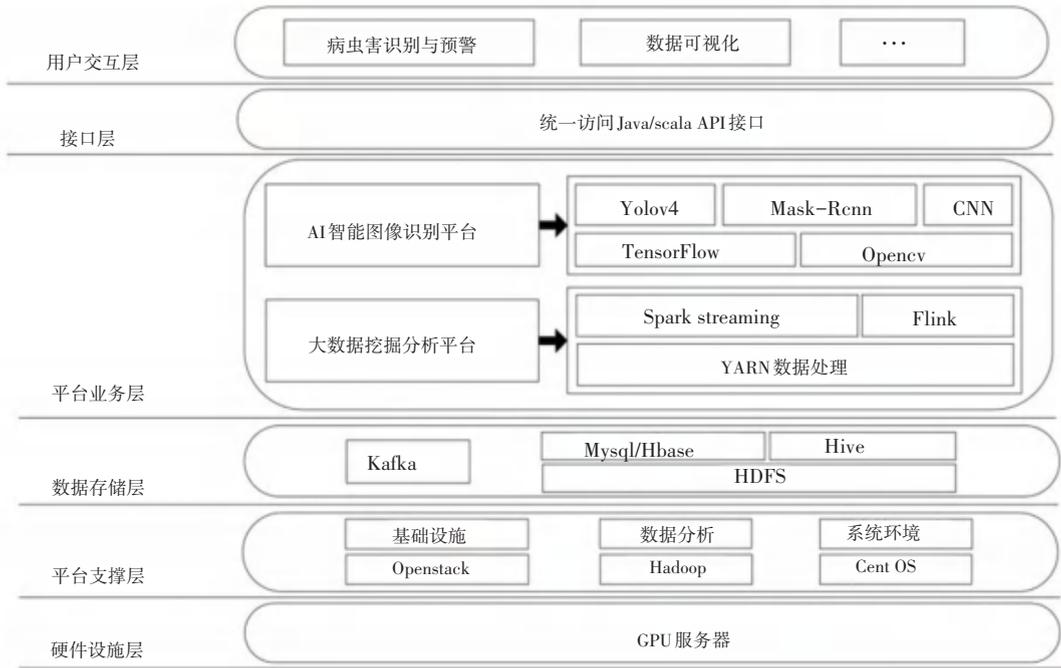


图 1 系统总体设计

Fig. 1 Overall design of system

1.3 平台业务层

平台业务层,由 AI 智能图像识别平台、大数据挖掘分析平台和后端管理系统组成。AI 图像识别平台采用 Yolov4 和 Mask-Rcnn 框架、TensorFlow、CNN 等人工智能图像识别组件实现柑橘病虫害图像识别功能^[3];大数据挖掘分析平台采用 Openstack 作为基础设施管理平台,采用 Hadoop 分布式计算平台对大数据进行挖掘分析^[4],其中包含:Flume、Kafka、Spring Boot、Spark、SSM 等框架组件;后台管理系统主要对数据库信息负责并对项目测试调试提供便捷。

1.4 数据存储层

数据存储层,平台采用 Kafka 和 HDFS 及其上层的 Hbase 和 Hive 组成。此层次实现大数据的存储管理,为大数据挖掘分析提供数据基础。

Kafka 主要作为信息收集管理和日志收集管理组件,其高吞吐量、低延迟、可扩展性强、高可靠性、高容错率和高并发性的特性可承载大量用户同时进行数据输入和输出,确保在柑橘图片信息大量输入输出时服务器稳定;支持将数据持久化到本地磁盘,防止柑橘病虫害数据和解决方案等数据丢失;支持集群中节点失败(即允许副本数量为 n 时,有 $n-1$ 个节点失败),当有集群节点失败时也不影响本系统正常运行;其日志收集功能通过 kafka 以统一接口服务的方式开放推送给 Hadoop 与 Hbase;其用户

跟踪功能能对用户进行实时监控分析,或装载到 Hadoop 中做数据挖掘和分析。

HDFS 分布式文件系统,使用多个节点的数据存储服务器组成集群,是实现 HBase 和 Hive 数据存储服务的基础。Hbase 依赖 HDFS 来作为物理存储的同时需要 Zookeeper 来进行分布式协调服务,Hive 只需要部署 Hadoop 平台即可工作。Hbase 虽具有实时处理的功能但是其不支持 SQL 语句,导致很难对数据库中柑橘病虫害信息等数据进行操作,而 Hive 可以使用 HQL 语句(同属 SQL 语句)。所以在数据存储层需要同时使用 HBase 和 Hive,此时需要 MapReduce 框架来联合使用 Hbase 和 Hive 才能实现完整的数据库存储与控制。

1.5 平台支撑层

平台支撑层,采用 Openstack 作为基础设施 IAAS 管理平台为本系统提供基础硬件设施调度分配服务;Hadoop 作为数据分析处理与存储管理平台,实现大数据的挖掘与分析;CentOS 为平台提供底层运行环境。

1.6 硬件设施层

采用华为云端 GPU 服务器 pi2.4xlarge.4 作为硬件基础。

其包含:

CPU 配置: Intel Cascade Lake 6278 2.6GHz

GPU 配置: NVIDIA T4 16G

内存:64GB

硬盘:50TB

基准/最大带宽:4/10 (Gbit/s)

2 系统总体业务流程设计

2.1 用户流程

通过用户发起信息提交, AI 智能图像识别平台对病虫害进行识别会直接得到病虫害种类信息;收集图片中的时间、地点、病虫害种类和手机信令等信息, 发送至解决方案分析系统和大数据分析处理平台中;系统会根据提交的信息结合爬虫爬取的当地相关天气、生产生活习惯等信息对解决方案推送办法进行分析, 分析完成后向数据管理终端发送请求;

终端接收到请求后, 如符合数据库中存储的解决方案属性, 则会作出反馈;收到反馈后通过前端页面推送解决方案相关信息。

2.2 数据分析流程

系统总体业务流程设计如图 2 所示。数据进入大数据挖掘分析平台后, 首先进行数据处理——剔除无效与损坏的数据后, 对数据进行分类和标签化;然后将数据呈递给分析系统, 通过大数据分析病虫害发生规律, 经数据管理筛选后, 将符合标准的分析结果送入数据库中存储, 最后将数据传递给前端的数据可视化展示界面, 实现病虫害的预警与病虫害数据可视化展示。

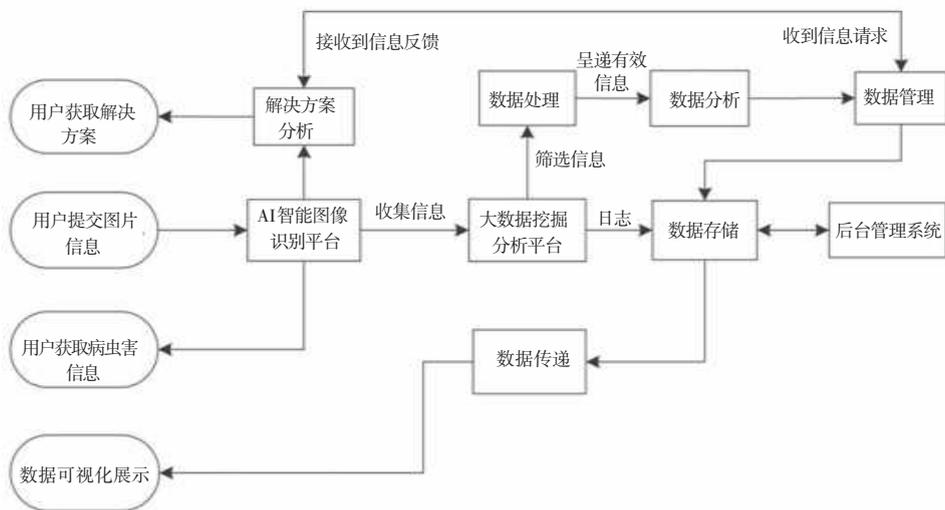


图 2 系统总体业务流程设计

Fig. 2 Overall process business design of system

3 AI 智能图像识别实现

3.1 范围检测

由于病虫害产生时, 不同病虫害在柑橘表面产生的视觉效果不同。部分以表面糜烂或扩散性表面坏死表现, 所以对于病虫害的识别本系统首先对病虫害的视觉效果进行范围识别检测^[5]。

通过自行改善优化的 Yolov4 开源图像识别框架, 实现效果较为准确的范围定位, 确定其产生范围大小可以分析其受病程度, 并且能进一步筛选掉一些非范围致病的虫害, 增加识别速度与效率。

3.2 病虫害识别

3.2.1 数据集

本系统在优化的 Mask-rcnn 图像识别框架的基础上, 采用了标准化的柑橘病虫害训练集样本, 包含: 特征明显的柑橘病虫害识别样本、常见的柑橘病虫害识别样本等。样本通过人工抽样采集汇总而成。

对于地区特有的柑橘病虫害类型, 经过走访调查, 本系统也收录了各种四川特异性的柑橘病虫害样本信息, 实现病虫害四川地区特异性识别。

本系统采用了 Mask 数据集增强, 使数据在原有基础上遍历 20 个 epoch, 确保其在 20 个变换过程中无损。

3.2.2 柑橘病虫害图像识别

使用 Yolov4 进行范围识别, 得到最外侧的矩形范围框架; 框架内的图像进行进一步特征值提取, 使用 Mask-Rcnn 框架训练的权重对其进行识别。柑橘常见病种红蜘蛛病的病症原图与识别效果图如图 3 和图 4 所示。

4 大数据挖掘分析实现

4.1 平台搭建

本系统采用 CentOS 7.4 下的 Linux 环境作为底层环境; 部署 Openstack 基础设施管理平台作为 IAAS 对基础硬件设施进行管理调度与分配; 部署

Hadoop 分布式云存储管理平台对大数据存储进行管理, 其中包含: Flume、Kafka、Spring Boot、Spark、SSM 等框架组件。



图 3 病症原图

Fig. 3 The original image of symptoms

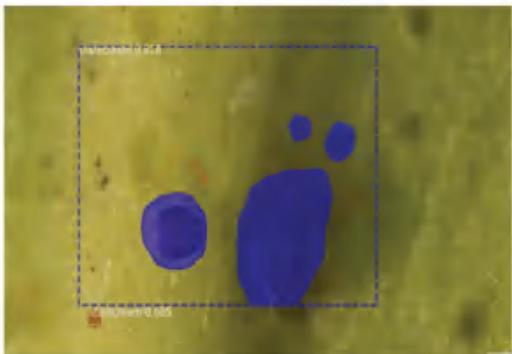


图 4 识别效果图

Fig. 4 The identify renderings

4.2 数据挖掘

4.2.1 数据获取与挖掘

通过用户上传的图片数据、用户手机信令与备注信息等, 可以获取相关病虫害种类、病虫害发生位置、时间、受灾程度等信息。这些信息将一并打包发送至后端数据服务器进行处理。

将获取的数据与标准化数据比对, 将不合格数据剔除后保留有效数据, 再将获取的数据进行分类, 区分其数据种类, 最后将数据标签化, 便于大数据平台处理以及信息可视化展示, 此操作由数据服务器端中 Python 程序与 Shell 脚本完成。

4.2.2 数据分析

将递交上来的数据送入病虫害预警分析平台中, 根据标签化的数据, 结合实时更新的地区数据, 将会预测出近一个月到一周内病虫害的变化趋势, 并在 Web 前端平台上展示, 如图 5 所示; 同时已发生的数据将会送至柑橘病虫害统计平台, 实现 Web 前端平台的数据可视化, 如图 6 所示。实现相关可视化界面主要利用 echarts、spring boots、css 等框架。



图 5 柑橘病虫害预警分析

Fig. 5 Warning analyze of diseases and pests of citrus



图 6 柑橘病虫害数据可视化展示

Fig. 6 The visual presentation of diseases and pests of citrus

5 解决方案智能推送

通过对初步数据的分析, 得到较为标准的分类标签化数据, 系统自动推选出几组对应的解决方案。结合当地地理环境、近期天气、生产生活方式等因素, 并通过设置不同的权值, 筛选出较为适宜的若干套解决方案, 再结合大数据分析其方案的普遍性以及被选择率等, 分优先层级推荐解决方案。

6 结束语

本系统设计通过使用人工智能图像识别技术与大数据分析技术, 实现柑橘病虫害智能识别、地域化的柑橘病虫害识别、病虫害防治解决方案智能推送、针对病虫害发生的预警分析及其相关数据可视化展示。一定程度上提高农村生产生活信息化水平, 降低农户受灾风险; 预警系统能在病虫害发生前提醒农户提前防治, 在最大程度上减少损失。

参考文献

- [1] 范振军. 农作物病虫害图像检索方法研究与实现[D]. 西安: 西安科技大学, 2018.
- [2] 温芝元, 曹乐平. 柑橘病虫害机器视觉[J]. 工程数学学报 2012(5): 641-645.
- [3] 赵建敏, 薛晓波, 李琦. 基于机器视觉的马铃薯病害识别系统[J]. 江苏农业科学 2017(2): 198-202.
- [4] 杜涛. 面向柑橘病虫害预警的 Hadoop 数据挖掘技术研究[D]. 南昌: 华东交通大学, 2017.
- [5] 黎移新. 柑橘病虫害疤痕的计算机视觉识别[J]. 食品与机械 2009(3): 78-81.