

引用格式: 柏艺琴, 陈新锋, 原军锋. 基于大数据的无人机云交换平台统计分析技术研究[J]. 地球信息科学学报, 2019, 21(4): 560-569. [Bai Y Q, Chen X F, Yuan J F. Statistical analysis technology of UAS cloud data exchange platform based on big data[J]. Journal of Geo-information Science, 2019, 21(4): 560-569.] DOI:10.12082/dqxxkx.2019.180395

基于大数据的无人机云交换平台统计分析技术研究

柏艺琴¹, 陈新锋¹, 原军锋²

1. 中国民航科学技术研究院, 北京 100028; 2. 北京享飞网络科技有限公司, 北京 100016

Statistical Analysis Technology of UAS Cloud Data Exchange Platform based on Big Data

BAI Yiqin^{1*}, CHEN Xinfeng¹, YUAN Junfeng²

1. China Academy of Civil Aviation Science and Technology, Beijing 100028, China;

2. Beijing Xiangfei Network Technology Company Limited, Beijing 100016, China

Abstract: After several years of development, light and small unmanned aircraft systems (UASs) have been widely used in various industries both in China and many other countries. However, the UASs have many models, with scattered equipments and no systematic management. Meanwhile, some safety issues exist. This urgently requires the relevant regulatory authorities to regulate, supervise, and maintain safe flight operations by taking their operating rules and characteristics into account. In order to standardize the operation of light and small civil UASs across the country and promote the industry development, the Civil Aviation Administration of China has issued the "Provisions for the Operation of Light and Small Unmanned Aircraft (for Trial Implementation)" advisory circular and the "Specification for Interface Data of Unmanned Aircraft System Cloud System". The UAS cloud data exchange platform was developed in 2016, and the data sharing of multiple UAS cloud systems in China was realized. With this platform the UASs registered in different UAS cloud systems are visible to each other in the same airspace, which improved the flight safety of China's low-altitude airspace. However, with the rapid development of the application of the unmanned aerial vehicle (UAV) industry, the number of UAVs supervised by the Civil Aviation Authority and the data on the operation of the UAVs has increased dramatically, which has also brought great challenges to the traditional data management methods. In this paper, we will describe the current situation of the operation of big data from the UAS cloud data exchange platform in China followed by discussion on the technical bottle necks in the statistical analysis of the operation data of the traditional UAS. Then we will propose a statistical analysis method for the UAS operation data, and establish a framework of statistical analysis of big data from the cloud data exchange platform. In the end, we will outline how to use Apache Spark and Cassandra database to quickly process, store, count, and analyze the massive data generated by the UAS cloud data exchange platform. The research situation of implementing various statistical index algorithms based on the platform is introduced in detail. This research not only improves the efficiency of statistical analysis of UAS operation data, but also provides the operation management rules of China's light and small UASs from multiple dimensions. We highlight that the UAS has significant operational

收稿日期: 2018-08-24; 修回日期: 2019-01-06.

基金项目: 国家重点研发计划项目(2017YFB0503005). [**Foundation item:** National Key Research and Development Program of China, No.2017YFB0503005.]

作者简介: 柏艺琴(1979-), 女, 湖南长沙人, 硕士, 主要从事无人机运行数据分析和无人机运行管理研究。

E-mail: baiyq@mail.castc.org.cn

characteristics, which are different from general and transportation aviations. This paper provides reference for government and industry decision-making, which has strong practical significance.

Key words: big data analysis; big data architecture; UAS cloud data exchange; UAS operation analysis; Civil Aviation Administration of China

***Corresponding author:** BAI Yiqin, E-mail: baiyq@mail.castc.org.cn

摘要:随着无人机在各行业应用的迅速普及和大规模应用,民航局监管的无人机数量和无人机运行数据呈现海量增长,在给政府和行业监管带来便利的同时,也给传统的数据分析方法带来了极大的挑战。本文阐述了民航局无人机云交换平台的运行大数据的现状,分析了传统无人机运行数据统计分析存在的技术瓶颈,提出了基于大数据进行无人机运行数据统计分析方法,建立了无人机云数据交换平台大数据统计和分析系统的设计框架,并概述了如何基于 Apache Spark 和 Cassandra 数据库的大数据分析方法将无人机云交换平台生产的海量数据快速处理、存储、统计和分析的研究思路和方法,以及实现各类无人机运行特征分析的情况。目前基于本文的研究成果不仅可以满足行业主管部门对无人机运行数据统计的需求,还对掌握和了解中国无人机运行特征,探索更合适的无人机管理措施和健全无人机监管体制机制,有非常重要的意义。

关键词:大数据分析;大数据架构;无人机云数据;无人机运行分析;中国民用航空局

1 引言

低空轻小型无人机具有成本低、易操作和高度灵活性的特点,能够携带相关任务设备在空中执行一些特殊任务。经过近几年的发展,其技术已经较为成熟,并广泛应用在电力巡检^[1]、交通流量分析^[2]、国土资源监测^[3-4]、农林遥感^[5]和灾害应急服务^[6]等领域。然而,这类无人机机型多,装备分散,缺乏统一管理,存在严重安全隐患。随着国家逐步对低空的开放,无人机在各行业应用的迅猛发展,迫切需要监管部门对无人机进行统一监管^[6-7],掌握其运行规律,并使无人机运行数据共享,实现安全飞行^[8-9]。

为了规范全国轻小无人机的运行,2015年12月,中国民用航空局(简称民航局)飞行标准司下发了《轻小无人机运行规定(试行)》(AC-91-31)咨询通告^[1],该咨询通告明确了民用无人机的运行监管(使用电子围栏、无人机云等^[10])、对无人机云提供商的要求等,对Ⅲ类以上无人机以及部分Ⅱ类无人机提出了接入无人机云系统的要求^[10]。该文件的出台,推动了无人机监管行业的发展。依据该文件要求,截止2018年12月,全国已有包括优云(U-Cloud)在内的9家无人机云运营商取得了相应的运行资质。

为了更好地促进行业发展,民航局颁布了《无人机云系统接口数据规范》MH/T 2009-2017行业标准^[11],并研制了无人机云数据交换平台(简称交换平台),实现了全国多个无人机云系统数据共享,实现了同一空域下注册无人机互相可见,从而提升

了中国中低空飞行安全。

从国内和国际上已经公布的有关民用无人机的量化信息来看,主要以无人机注册量和驾驶员合格证数量为主,相关的无人机运行数据一直缺乏,由于该行业发展较快且时间较短,运行数据量较大,政府和行业亟需相应的大数据分析和统计数据为管理决策提供支撑。

随着无人机在各行业应用的迅速普及和大规模应用,民航局监管的无人机数量和无人机运行数据呈现海量增长,在给政府和行业监管带来便利的同时,也给传统的数据管理方式带来了极大的挑战。本文研究的大数据分析技术和统计方法,是改进最初使用的无人机运行数据低效统计算法和方法,对交换平台产生的海量全国民用无人机运行数据快速处理分析和统计,使用图表对其进行汇总和分析,并按照时间周期形成相应的统计报告。一方面,可为民航局提供无人机运行管理大数据参考和决策依据,便于管理无人机,同时对监管部门和行业管理机构掌握和了解中国轻小无人机运行规律和特点也有重要意义;另一方面,对无人机运行数据的周期性分析也能够对云数据交换过程中的问题数据分析和统计,及时提醒各个云系统改进和完善功能,提升数据的有效性,并适时改进管理规则,进而为政府和行业决策提供参考。

2 无人机数据

本文以无人机作为研究主体,基于交换平台的运行数据对无人机运行特征进行统计分析研究。

2.1 数据源

本文不对实时动态运行数据进行分析研究,采用的原始数据均来源于交换平台的真实数据,主要包括以下4类数据源(图1):① 无人机云数据:各无人机云系统按照数据交换协议每天实时上报无人机运行数据给交换平台,每月定期从交换平台获取原始运行数据;② 无人机云运营商数据:每月定期从交换平台获取;③ 无人机实名注册数据:每月定期从民航局无人机实名注册系统获取数据,用于提取无人机注册数据,可以知道是否合法飞行,清洗和抽取无人机运行数据;④ 无人机驾驶员执照数据:每月定期从民航局无人机驾驶员执照系统获取数据,用于提取驾驶员数据,可以知道是否合法飞行,抽取和清洗无人机运行数据。

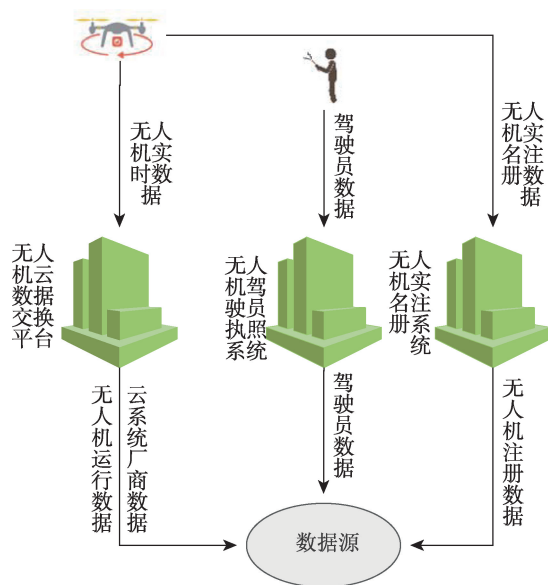


图1 无人机云数据分析数据源

Fig. 1 Data source of UAS exchange data

2.2 无人机云数据交换平台

依据民航局咨询通告AC-91-31的要求^[10],对于III、IV、VI和VII类的民用无人机应接入无人机云,在人口稠密区报告频率最少每秒一次,在非人口稠密区报告频率最少每30 s一次,并且具备数据续传功能。接入云系统的无人机飞行过程中应实时上传飞行数据^[10]。

受民航局委托,中国民航科学技术研究院按照行业标准MH/T 2009-2017^[11]研制的交换平台可以接收各个云系统实时上报的无人机运行数据,通过数据分发使各云系统之间数据互相交换,并记录存储,实现无人机数据交换和共享。

2.3 无人机数据交换协议

按照民航行业标准MH/T 2009-2017^[11]的要求,无人机云系统必须按照标准中规定的无人机数据交换协议实时向民航局无人机云数据交换平台上报每架无人机运行数据,实现对无人机运行数据监测和数据交换。

无人机数据交换协议内容包括:① 无人机产品序列号(MSN)、飞控系统序列号(FCSN)、国籍登记标志或实名注册编码(REG)、无人机云运营商(以下简称运营商)在云系统中为无人机生成的编号(CPN)等静态信息;② 经度、纬度、高度、地速、航向、时间等动态信息^[11]。其中,无人机动态数据交换协议如表1所示。每个无人机云系统通过上述这些接口规范实现无人机运行数据的交换和共享。

3 基于大数据的云数据统计分析方法

大数据(Big Data)已经成为普遍关注的话题,其指的是数据集和流,这些数据集和流的规模足够大,当使用常规工具和基础结构在可容忍的时间量内收集、管理和处理数据时都面临重大挑战^[12-13]。大数据分析(Big Data Analytics, BDA)是大数据理念与方法的核心,是指对海量类型多样、增长快速、内容真实的数据进行分析,从中找出可以帮助决策的隐藏模式、未知的相关关系以及其他有用信息的过程^[14]。官思发等^[15]认为大数据分析是根据数据生成机制对数据进行广泛的采集与存储,并对数据进行格式化清洗,以大数据分析模型为依据,在集成化大数据分析平台的支撑下,运用云计算技术调度计算分析资源,最终挖掘出大数据背后的模式或规律的数据分析过程。

本文重点利用大数据、云计算等相关技术架构和数据分析算法工具,对无人机运行数据相关分析方法进行研究。

3.1 无人机云数据分析研究的特征指标

本文从以下5个维度研究中国轻小无人机运行规律和特征:

(1)不同无人机云系统的无人机运行数据交换量频次。从各个云系统提供的无人机运行数据记录信息可分析每个云系统监视的无人机数量、无人机类别、实名注册率、不同时间段的无人机的运行小时数等。

(2)无人机运行的高度分布。了解无人机运行

表1 无人机动态信息所包含数据信息

Tab. 1 The structure of UAS dynamic information

数据名称	数据描述	数据处理
CPN	可采用匹配的方式获取	
经度/°	度	精确到小数点后7位
纬度/°	度	精确到小数点后7位
高度/m	星基高度(GNSS高度),GNSS全球卫星导航系统(Global Navigation Satellite System)	精确到小数点后2位
时间/s	UTC 世界时(Coordinated Universal Time)	精确到小数点后3位
地速/(m/s)		精确到小数点后1位
航向/°	真航向	精确到整数位
定位精度/m	水平定位精度(Hdop)	精确到小数点后2位
有效数据长度	—	—
系统状态位	0—无人机处于正常状态,异常置相应位	—
保留字节长度	描述保留字段的长度	—
保留字段	自行定义	—

的高度区间范围,以及在不同高度区间的运行频次,掌握不同高度区间的无人机类别分布情况。

(3)无人机运行的速度分布。在某一选取条件下,按照飞行时长信息筛选出不同区间的无人机运行速度分布,了解运行的各类别无人机所占比例,以及运行速度分布的普遍规律。

(4)无人机运行的时间分布。利用频率直方图统计出某一选取条件下无人机在不同时段的分布,可观察不同时段无人机运行数量,了解无人机的日常运行规律^[16]。

(5)无人机运行的空间分布。在某一选取条件下,按照时间信息筛选出不同时段无人机运行空间分布,进而了解不同时段无人机运行区域分布规律,掌握无人机运行集中分布区域^[16]。

3.2 传统数据库分析方法存在的问题

前期由于接入民航局无人机云交换平台的无人机云系统和无人机数量较少,无人机运行数据量小,使用传统的MySQL关系数据库系统和Excel软件即可对云数据处理,汇总统计和BI分析。随着无人机云系统和无人机接入数量的增加,平台无人机运行数据呈海量增长,各无人机云运营管理不到位也使上报的运行数据中脏数据比前期增多,数据辨析、抽取和清洗的复杂度和计算量增大,直接导致处理流程臃肿,数据存储和数据处理的性能成为瓶颈,统计分析效率降低。主要表现在以下3个方面:

(1)交换平台部署在民航局信息中心,采用的是MySQL数据库系统,由于机房安全管理的要求无

法直接使用数据库,每次分析统计前需要将数据从信息中心机房服务器导出,再导入用于统计的MySQL数据库系统,对于海量数据记录,像MySQL这种关系型数据库系导入速度很慢,利用MySQL数据库分析统计速度更慢,且不是简单提升硬件配置可解决的。在很多任务中,MySQL表现非常不好。MySQL的限制之一在于:1次查询等于1个CPU内核,即便采用多内核,也无法充分利用所有的计算能力,相反采用Spark计算却能充分使用CPU内核。

(2)无人机在高空飞行,运行速度快,对安全性要求很高。无人机云交换数据存储频率要求为每秒一次,随着民航局批准的云系统和无人机数量的不断增加,交换平台的云数据记录海量增加。其中2017全年无人机云系统数据总量约200 G,提供的有效数据为12 684万条,换算成飞行时间为176 169 h,其中1个云系统批准的晚,所上报数据可以忽略,无人机云交换数据量增长趋势见图2。截止到2018年6月,民航局已经批准了8个无人机云系统,根据从2017年1月到2018年9月的云数据交换量增长趋势分析,并考虑即将扩展的新的云数据交换协议将会增加指令、飞行计划、驾驶员等信息,预测2020年的云数据交换量将达到10 TB以上,并且民航局要求所有无人机运行数据都要永久存储,传统关系数据库的处理计算能力和统计效率已经无法适应无人机云数据的快速增长。

(3)无人机云运营商的研发能力层次不齐,无人机云系统上报数据包含了很多脏数据。为了真实反映无人机运行情况,首先在分析过程中需要对

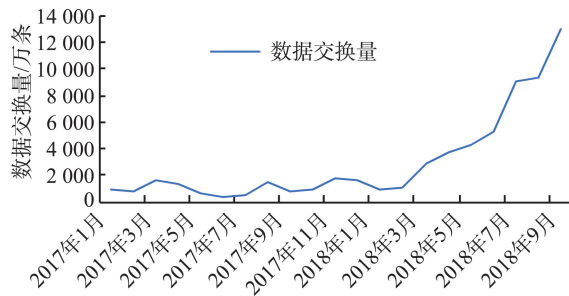


图2 2017年1月-2018年9月无人机云数据交换量增长趋势

Fig. 2 The growth trend of UAS exchange data from January 2017 to September 2018

云数据进行清洗,然后仅对有效数据进行分析统计,如果直接采用MySQL数据库来清洗数据,其效率很低,同时,鉴于SQL语句非常复杂,其无法满足统计所要求的灵活修改需求。

3.3 云数据大数据分析系统信息架构

大数据处理的信息架构覆盖了数据的捕获、分类、存储、建模、计算和分析等不同活动。对于不同来源的数据流,应按照数据类型特征将文档、图表、日志和邮件等不同类别的数据归类存储,通过构建数据模型实现数据的物理存储与数据应用之间的分离,使更便捷的数据管理成为可能^[17]。

Apache Spark、NoSQL和大数据可视化分析工具等是支撑大数据处理和分析的基础技术,也是大数据信息架构的核心。

3.3.1 大数据处理计算框架

针对集群环境的大数据计算框架有很多,包括MapReduce、Hadoop、Storm等,但是一些计算框架支持的处理模式单一,存在很多不足,往往需要一系列不同的处理框架才能很好完成,如MapReduce支持批处理计算模型较好,但对交互式 and 流式计算模型支持支持不好,而Hadoop的机制和MapReduce一样。针对这些不足,伯克利大学推出了全新的统一大数据处理框架Spark,后捐赠给了Apache软件基金会。Apache Spark是用于大规模数据处理快速通用计算引擎,它提供了一种抽象的弹性分布式数据集(RDD),可以有效执行分布式计算,可以支持同时处理3种计算模型。另外,Spark有优秀的多任务调度模型,并且允许将数据加载到集群内存中反复查询,大大缩短计算时间^[18]。

Apache Spark提供的Spark SQL可以高效使用SQL处理结构化数据,提供了查询结构化数据及计

算结果等信息的接口。本文在分析无人机云数据时使用Spark读取原始压缩TXT文件,然后用弹性分布式数据集(RDD)转化格式,用SQL进行查询,根据分离规则进行筛选、清洗和分组,根据特征分析维度按小时、天、月、季度、年分区,最后将管道结果(聚合数据)存入Spark。

3.3.2 大数据管理系统

NoSQL的全称Not Only SQL,是一种非关系型数据库管理系统,是为了解决大规模多种数据集合带来的挑战以及大数据应用难题应运而生的^[17]。

Apache Cassandra是一个分布式、可扩展、容错的NoSQL数据库系统。它最初由Facebook开发,集Google BigTable的数据模型与Amazon Dynamo的完全分布式架构于一身。由于Cassandra良好的可扩展性和性能,被Apple、Instagram、Netflix等知名网站所采用,成为了一种流行的分布式结构化数据存储方案。

美国Datastax公司比较了4个最具竞争力的NoSQL系统,比较结果显示Cassandra的工作负载和负载过程的吞吐量具有明显优势^[19],Cassandra具有最佳的速度和可靠性,Cassandra在快速读写性能方面享有盛誉,在数据存储方面具有无可否认的可靠性。

中国无人机产业和应用发展飞速,民航局对无人机管理也越来越规范,无人机云数据的增长只会量级增长,大数据分析应用场景只会更复杂,综合考虑Cassandra数据库较好的性能和可靠性,本文分析时选用Cassandra数据库作为大数据管理系统来存储分析后的大数据集。

3.3.3 无人机云数据可视化分析方法

大数据可视化是大数据分析不可或缺的重要手段和工具。可视分析能够有效地弥补计算机自动化分析方法的劣势与不足,大数据可视分析将人面对可视化信息时强大的感知认知能力与计算机的分析计算能力优势进行有机融合,在数据挖掘等方法技术的基础上,综合利用认知理论、科学/信息可视化以及人机交互技术,辅助人们更为直观和高效地洞悉大数据背后的信息、知识与智慧^[20]。

时空数据可视化与地理制图学相结合,重点对时间与空间维度以及与之相关的信息对象属性建立可视化表征,对与时间和空间密切相关的模式及规律进行展示^[20]。基于大数据可视化分析技术研究和挖掘无人机运行信息分布特征,可从客观上反映无人机日常飞行在行业应用、时间和空间上的分布规律,可有效、精准掌握中国无人机活动的时空

分布特征和规律,为民航局管理无人机提供最及时和客观的参考。

本文在大数据可视化分析中需要解决的最大难题是如何基于大数据生成无人机运行的空间分布特征图,如在2018年第3季度的飞行轨迹数据为31 368万条,这么大数据量目前在Web浏览器根本无法将数据流传到前端数据可视化组件生成热力图。通过分析研究,决定选用GeoMesa地理空间分析工具,在云服务器端实现热力图的绘制,然后再推送到前端客户端展示。

GeoMesa是一个分布式时空数据库系统,支持在Apache Spark引擎上进行大规模地理空间查询和分析,支持自定义分布式地理空间分析,可在Cassandra数据库之上提供时空索引,用于大量存储点、线和多边形数据。GeoMesa可通过GeoServer GIS服务引擎,以标准OGC(开放地理空间联盟)API和协议(如WFS和WMS)与无人机云数据大数据分析系统集成,支持时间分析、性能评估、直方图、热力图等数据地图可视化分析处理^[21],采用该方法,30 000万条轨迹数据大约消耗10 s的时间完成热力图的绘制。

3.4 无人机云数据大数据分析处理模式及流程

大数据的处理流程可以定义为在合适工具的辅助下,对广泛异构的数据源进行抽取和继承,结果按照一定的标准统一存储,利用合适的数据分析技术对存储的数据进行分析,从中提取有益的知识并利用恰当的方式将结果展现给终端用户。具体来说,可分为数据抽取与集成、数据分析以及数据解释^[12]。其中在数据抽取和集成时需要对数据进行清洗,保证数据质量和可信性。数据分析是大数据处理的核心,但如果采用适当的解释方法,极端情况下可能会误导用户,可以考虑从可视化技术和人机交互2个方面设计提升数据解释能力。

无人机云数据大数据处理分析流程如图3所示:①为了提升源数据导入的速度和效率,避免采用MySQL数据库系统处理数据,源数据不采用数据库原始数据备份文件,而是从无人机云交换平台的MySQL数据库系统将原始数据批量导出为TXT格式文件;②调用数据预处理服务程序,使用Spark RDD方式将原始数据取出放在缓存中,利用Spark计算引擎对原始数据进行清洗、抽取和分块处理,构建出5个分析维度的细小粒度(秒级)的大数据集存储在Cassandra数据库;③统计分析时利

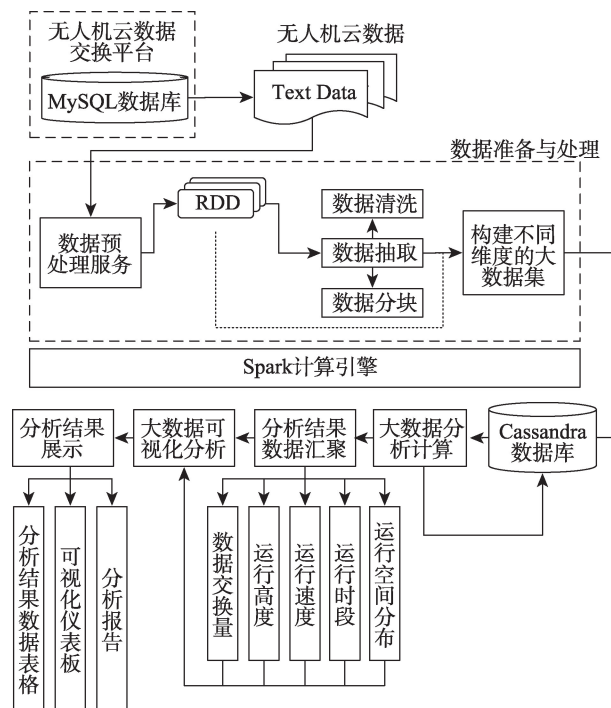


图3 无人机云数据大数据处理分析流程

Fig. 3 The flow chart of UAS exchange big data processing and analysis

用Cassandra数据库快速查询并聚合出分析结果,并利用大数据可视化分析工具生成不同维度的数据图表分析和展示。

本文基于上述大数据处理分析架构研制开发了无人机云数据大数据分析统计系统(以下简称统计分析系统),基本设计思路是将系统分为下述3个模块,其中云服务器端模块部分基于Spring Boot框架,采用Java开发,Web前端部分采用Vue框架开发。

(1)云数据导入模块:相当于系统的数据生产者角色,负责将TXT格式原始数据文件批量处理,按照Spark RDD方式将原始数据转换后放入Spark缓存中。

(2)云数据分析处理模块:利用Spark计算引擎对缓存中原始数据进行清洗(重复数据、位置跳点数据等)^[22-23]和抽取。系统分析处理模块在设计上充分利用Spark基于缓存和集群处理的优势来提高整个分析的效率和缩短分析时间消耗。

(3)云数据统计分析模块:此模块对分析特征指标进行计算分析后汇聚生成各类分析结果,该结果存储在MySQL数据库。分析结果在Web端使用表格或图表的形式展现,并可生成无人机云数据统计报告。由于使用了聚合方法,本文的分析数据结果集将会小很多,从而将其存放在MySQL数据库

中并较容易实现与许多其他标准程序一同进行协作处理。利用本文研究的大数据分析方法研制开发的统计分析系统实现了系统化高效快速分析无人机云数据和生成统计分析结果。整个统计分析系统采用了分布式设计,具有良好的扩展性,极大缩短了无人机云数据统计分析的周期,可以较好地满足快速增长的无人机云数据的大数据分析要求。

4 结果及分析

为了更全面分析中国民用无人机运行的时空分布和变化特征,采用本文研究的无人机云数据统计分析方法,对中国2017年和2018年参与民航局无人机云数据交换和共享的无人机云系统中共计62 274万条飞行数据进行了处理和计算,分别对无人机数据交换总量、运行高度、运行速度、运行时段、运行区域时空分布等特征进行多维度分析。其中在大数据可视化分析方面,对无人机云数据交换总量和运行时段变化特征按时间序列采用直方图分析,对无人机运行速度和高度变化特征按时间序列采用折线图分析,并采用帕累托图对不同运行高度区间的占比情况进行分析,对无人机运行时空特征采用热力地图进行不同时间段分析。

4.1 无人机云数据交换总量特征分析

图4是研究无人机云数据交换总量分布直方图,本文通过总量分布图既可研究无人机数据交换量的增长趋势,也可研究无人机业务活动的变化趋势以及业务爆发的黄金时间。从图4可看出,无人机的业务活动分布呈现季节性的变化特征。

4.2 无人机运行高度特征分析

考虑到无人机云系统中注册的无人机以轻小

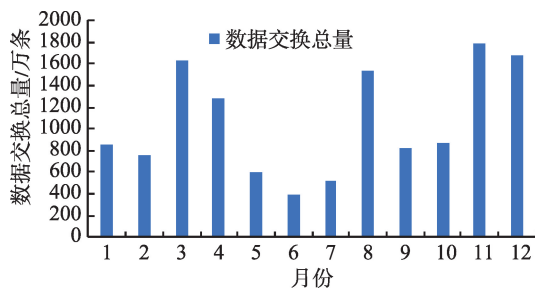


图4 2017年各月份无人机数据交换总量分布
Fig. 4 Distribution of monthly total number of exchange data for UAS in 2017

型为主,故本文研究的飞行高度主要以120 m以下运行数据为主。

将无人机飞行高度区段划分为0~20 m、20~40 m、40~60 m、60~80 m、80~100 m、100~120 m共6个区段进行分析研究全年无人机在不同高度区段内运行频次和占比率。图5是高度占比图,由图中分布情况可以看出,无人机主要在40 m以下的空域飞行,其中20 m以下空域飞行的频次最高,高度占比率大约75%,运行高度在120 m以下的高度占比率将近98%,说明现阶段无人机基本在现行离地120 m高度隔离空域范围内合规飞行。

以10 m为间隔,分为12个高度区段,分析每天无人机在不同高度区段内运行频次,如图6所示。由图6可看出,无人机在0~5 m和35~45 m高区间内的活动频次较高,与图5的高度占比图研究分析的结果一致。

4.3 无人机运行速度特征分析

本文计算不同速度区间的无人机在空飞行时间,并将无人机的飞行速度划分为5段:0~5 m/s、5~15 m/s、15~25 m/s、25~35 m/s、35 m/s以上。同时,按照划分的等级统计各速度区间段的无人机在空

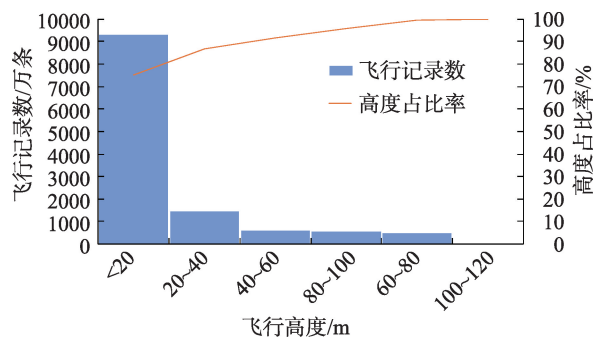


图5 2017年无人机运行高度占比
Fig. 5 The ratio of UAS operation altitude in 2017

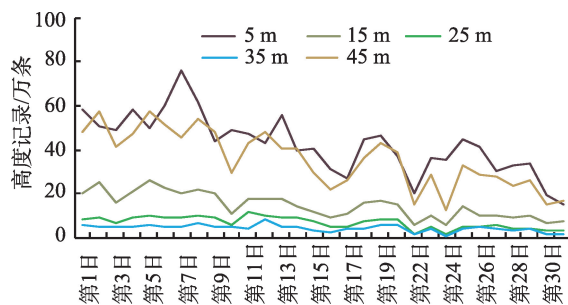


图6 2017年无人机日平均运行高度频次分布
Fig. 6 Distribution of daily average operation altitude of UAS in 2017

时间,结果如图7所示。从图中可看出,无人机运行速度主要集中于0~5 m/s和30~40 m/s之间。

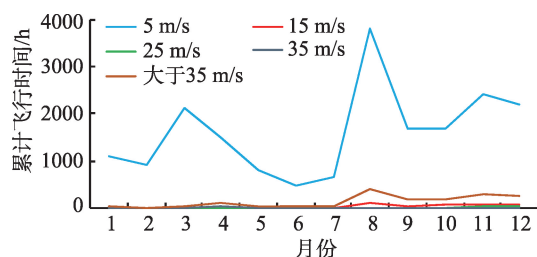


图7 2017年无人机各月份运行速度分布

Fig. 7 Monthly flight hours distribution of UAS operation altitude in 2017

4.4 无人机运行时段特征分析

以30 min为间隔,划分为48个时间段,分析无人机在不同时间段的作业运行频率,如图8所示。图中横坐标表示一日中的时间段,如12:00表示11:30~12:00之间。纵坐标表示某时间段内无人机作业累计飞行时长分布。从图8可知,2017年无人机作业飞行的最高峰值上午出现在10:30~11:00之间,下午出现在15:00~15:30之间,从而可以分析得知目前我国的无人机运行时段主要是昼间运行,与空军航管申报飞行计划要求的只允许白昼飞行时间一致。

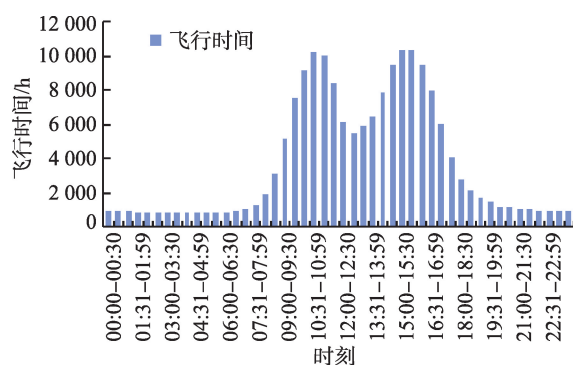


图8 2017年无人机运行时段分布

Fig. 8 Distribution of UAS operation time in 2017

4.5 无人机运行空间特征分析

热力图作为一种直观的时空数据可视化方法,具有综合展示数据地理空间特征和属性特征的良好特性,在空间大数据挖掘和知识发现的研究中具有良好的展示效果^[24]。为了更为直观地掌握无人机运行分布区域数据以及分析未来的发展趋势,选取不同时间段的无人机位置数据,使用核密度估计法^[25],利用GeoMesa时空数据地图分析引擎,计算并绘制无人机运行空间分布热力图,揭示无人机在运行的空间分布特征,并在数据地图上可视化展现

分析结果,以观察不同时间下无人机运行空间分布的动态过程。

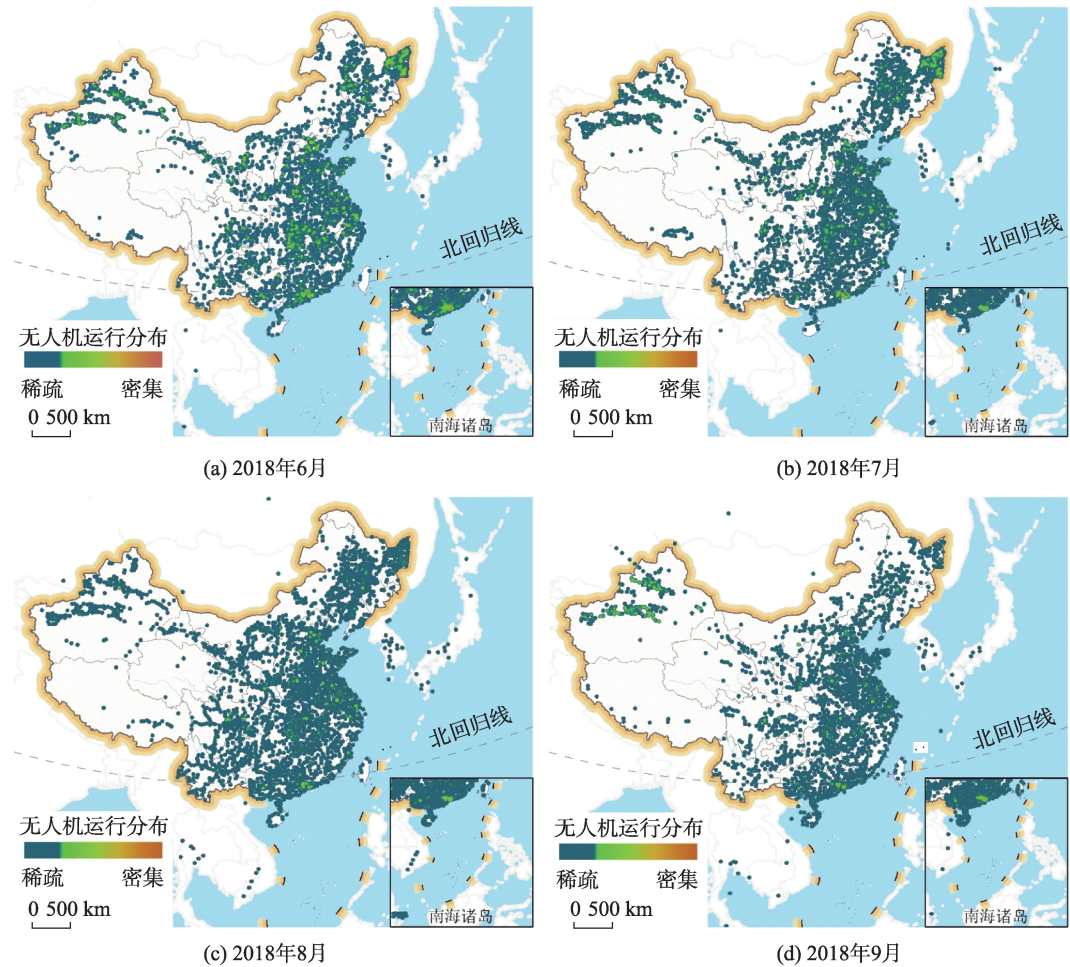
图9为2018年6、7、8、9月的无人机运行空间分布热力图,由红到蓝,颜色的变化代表着无人机活动分布密度的变化,密度值表示落入某一圆形搜索区内的点所占的权重值,其大小与核密度搜索区半径有关^[8]。图中红色区域表示无人机运行活动的密度大,蓝色区域表示无人机运行活动的密度小。从图9可看出:①中国无人机企业的业务范围覆盖了中国、日本、韩国以及东南亚等国家和地区;②在中国东、中部地区和华南地区无人机的运行频次很高,西南地区以四川省最多,东北、新疆、西藏等西北地区无人机运行频次较低,主要活动区域分布在新疆和黑龙江的经济作物规模种植区域,从中可看出无人机的应用发展与地区经济有着很大的关联关系。

由于部分无人机云系统上传的云数据没有进行过滤,把海外业务的无人机飞行数据也上传到云系统,为了充分研究无人机运行特征,本文的研究数据没有剔除这部分数据。本研究发现位置数据的坐标精度也对统计分析结果影响较大,部分无人机云系统提供的位置坐标精度较低,没达到标准MH/T 2009-2017的要求(保留小数点后7位的精度),导致可视化亮点很大程度的重叠,影响分布效果和可视化效果,这也是下一步分析系统清洗数据算法需要考虑的因素。

5 结论与展望

本文以无人机云数据交换平台海量运行数据为基础,建立了基于大数据的无人机运行数据特征分析方法,概述了如何利用Apache Spark大数据处理计算引擎和Cassandra数据库构建大数据统计分析系统将无人机云数据交换平台生产的海量数据快速处理、存储、统计和分析的研究思路和方法,并详细介绍了利用该方法对无人机运行数据5类特征指标的分析情况。

无人机因其独有优势,将在传统农林作业、工业测量和土地资源调查、智慧城市、数字考古、抢险救灾、三维实景、城市规划、新农村建设、物流配送、地理国情监测等行业应用中保持强势增长与快速拓展^[6],尤其是随着物联网和5G网络商业化的快速推进,无人机应用场景和使用频率将不断加大。我国民用无人机产品销售和服务总体市场规模到2020年将达到465亿元,2025年将达到750亿元,民



注：该图基于国家测绘地理信息局标准地图服务网站下载的审图号为GS(2016)1603的标准地图制作，底图无修改。

图9 中国2018年6-9月无人机运行区域热力分布

Fig. 9 Thermodynamic distribute map of UAS operation from June to September 2018 in China

用无人机井喷式发展,无人机大规模数据增长和数据处理为无人机大数据和运营分析带来了巨大的技术挑战,其中包括各种数据挖掘方案和从物联网挖掘数据^[13],也为中国无人机管理提出了很多挑战和难题,为政府和行业监管带来了众多新的研究课题。本研究发现,无人机具有显著运行特性,与通用航空、运输航空有本质的区别,因此也希望能够更多地结合无人机驾驶员操作数据、无人机任务数据、空域申报数据等和人工智能算法等技术进行更深层次的、更高效和精准分析无人机运行特征,以便通过更多的无人机运行大数据分析成果来探索更合适的管理措施来管理无人机,健全无人机监管体制机制,释放其市场活力,贡献国民经济发展。

参考文献(References):

[1] 王万国,田兵,刘越,等.基于RCNN的无人机巡检图像电力小部件识别研究[J].地球信息科学学报,2017,19(2):

256-263. [Wang W G, Tian B, Liu Y, et al. Study on the electrical devices detection in UAV images based on region based convolutional neural networks[J]. Journal of Geo-information Science, 2017,19(2):256-263.]
[2] Khan M A, Ectors W, Bellemans T, et al. Unmanned aerial vehicle-based traffic analysis: A case study for shockwave identification and flow parameters estimation at signalized intersections[J]. Remote Sensing, 2018,10(3):458-474.
[3] 宋晓阳,黄耀欢,董东林,等.融合数字表面模型的无人机遥感影像城市土地利用分类[J].地球信息科学学报, 2018,20(5):703-711. [Song X Y, Huang Y H, Dong D L, et al. Urban land use classification from UAV remote sensing images based on digital surface model[J]. Journal of Geo-information Science, 2018,20(5):703-711.]
[4] 杨海军,黄耀欢.化工污染气体无人机遥感监测[J].地球信息科学学报,2015,17(10):1269- 1274. [Yang H J, Huang Y H. Evaluating atmospheric pollution of chemical plant based on unmanned aircraft vehicle(UAV)[J]. Jour-

- nal of Geo-information Science, 2015,17(10):1269-1274.]
- [5] 李德仁,李明.无人机遥感系统的研究进展与应用前景[J].武汉大学学报·信息科学版,2014,39(5):505-513. [Li D R, Li M. Research advance and application prospect of unmanned aerial vehicle remote sensing system[J]. Geomatics and Information Science of Wuhan University, 2014,39(5):505-513.]
- [6] 廖小罕,周成虎,苏奋振,等.无人机遥感众创时代[J].地球信息科学学报,2016,18(11):1439-1447. [Liao X H, Zhou C H, Su F Z, et al. The mass innovation era of UAV remote sensing[J]. Journal of Geo-information Science, 2016,18(11):1439-1447.]
- [7] 任丽艳,李英成,薛艳丽,等.基于北斗技术的无人机飞行监管系统开发与应用[J].国土资源遥感,2018,30(2):238-242. [Ren L Y, Li Y C, Xue Y L, et al. Development and application of the UAV flying supervisory system based on Beidou technology[J]. Remote Sensing for Land & Resources, 2018,30(2):238-242]
- [8] 高国柱.中国民用无人机监管制度研究[J].北京航空航天大学学报(社会科学版),2017,30(5):28-36. [Gao G Z. On China's regulatory regime for civil unmanned aerial vehicle[J]. Journal of Beijing University of Aeronautics and Astronautics Social Sciences Edition, 2017,30(5):28-36.]
- [9] Spousta R, Chan S. Hold the drones: Fostering the development of big data paradigms through regulatory frameworks[J]. Journal of Communication and Computer, 2015, (12):135-145.
- [10] 中国民用航空局.AC-91-31.轻小无人机运行管理规定(试行)[S].2015:5. [CAAC.AC-91-31. Provisions for the operation of light and small unmanned aircraft (for trial implementation)[S]. 2015:5.]
- [11] 中国民用航空局.MH/T2009-2017.无人机云系统接口数据规范[S]. 2017:2-6. [CAAC.MH/T2009-2017. Specification for interface data of unmanned aircraft system cloud system[S]. 2017:2-6.]
- [12] 孟小峰,慈祥.大数据管理:概念、技术与挑战[J].计算机研究与发展,2013,50(1):146-169. [Meng X F, Ci X. Big data management: Concepts, techniques and challenges [J]. Journal of Computer Research and Development, 2013,50(1):146-169.]
- [13] Big data analytics - a research and innovation agenda for sweden[EB/OL]. <https://www.vinnova.se/globalassets/mikrosajter/strategiska-innovationsprogram/agendor/analys-av-stordata.pdf>.
- [14] 李广健,化柏林.大数据分析情报分析关系辨析[J].中国图书馆学报,2014,40(5):14-22. [Li G J, Hua B L. Relationship between big data analysis and intelligence analysis [J]. Journal of Library Science in China, 2014,40(5):14-22.]
- [15] 官思发,孟玺,李宗洁,等.大数据分析研究现状、问题与对策[J].情报工程,2015,34(5):98-104. [Guan S F, Meng X, Li Z J, et al. Big data study on the current situation, problems and countermeasures[J]. Journal of Intelligence, 2015,34(5):98-104.]
- [16] 孙飞,张霞,唐炉亮,等.基于GPS轨迹大数据的优质客源时空分布研究[J].地球信息科学学报,2015,17(3):329-335. [Sun F, Zhang X, Tang L L, et al. Temporal and spatial distribution of high efficiency passengers based on GPS trajectory big data[J]. Journal of Geo-information Science, 2015,17(3):329-335.]
- [17] 洪漪,赵栋祥,赵一鸣.大数据环境下的信息架构与数据模型[J].信息资源管理学报,2018,8(1):29-38. [Hong Y, Zhao D X, Zhao Y M. Information architecture and data modeling in a big data environment[J]. Journal of Information Resources Management, 2018,8(1):29-38.]
- [18] 夏俊鸾,等.Spark大数据处理技术[M].北京:电子工业出版社,2015. [Xia J Y, et al. Spark big data processing technology[M]. Beijing: Publishing House of Electronics Industry, 2015.]
- [19] Benchmarking NoSQL databases: Cassandra vs. MongoDB vs. HBase vs. Couchbase[EB/OL]. <https://www.datastax.com/nosql-databases/benchmarks-cassandra-vs-mongodb-vs-hbase>.
- [20] 任磊,杜一,马帅,等.大数据可视分析综述[J].软件学报,2014,25(9):1909-1936. [Ren L, Du Y, Ma S, et al. Visual analytics towards big data[J]. Ruan Jian Xue Bao/Journal of Software, 2014,25(9):1909-1936.]
- [21] Atufaali S, Chauhan P, Potdar M B. Analysis of big data using geomesa[J]. International Journal of Advance Research and Innovative Ideas in Education, 2018,4(2):4604-4612.
- [22] 李雪丽,盛勇,兰小机.基于Spark的并行化出租车轨迹热点区域提取与分析[J].计算机科学与应用,2018,8(9):1482-1489. [Li X L, Sheng Y, Lan X J. Extraction and analysis of hotspot region of parallel taxi trajectory based on spark[J]. Computer Science and Application, 2018,8(9):1482-1489.]
- [23] Kwak J, Park J H, Sung Y. Unmanned aerial vehicle flight point classification algorithm based on symmetric big data [J]. Symmetry, 2017,9(1):1-19.
- [24] 王胜开,徐志洁,张健钦,等.逆向热力图的绘制方法[J].地球信息科学学报,2018,20(4):515-522. [Wang S K, Xu Z J, Zhang J Q, et al. A reverse rendering method of Heatmap [J]. Journal of Geo-information Science, 2018,20(4):515-522.]
- [25] 陈卓然,黄翀,刘高焕,等.基于出租车GPS数据的居民就医时空特征分析[J].地球信息科学学报,2018,20(8):1111-1122. [Chen Z R, Huang C, Liu G H, et al. Analysis of spatial-temporal characteristics of resident travel for hospitals based on taxi GPS data[J]. Journal of Geo-information Science, 2018,20(8):1111-1122.]