

基于大数据与人工智能的疫情防控平台建设方案

刘漳辉^{1 2} 郭文忠¹ 陈羽中¹ 陈锋情²

(1. 福州大学数学与计算机科学学院, 福建福州 350108;

2. 福州市联创智云信息科技有限公司, 福建福州 350002)

摘 要: 针对疫情防控的迫切需要, 依托团队在数据采集、数据预处理、数据融合、数据存储、数据管理、数据计算、数据分析以及可视化技术方面的研究成果, 可以建设基于大数据与人工智能技术的疫情防控平台, 平台能够为各级政府机构提供及时准确的疫情动态信息及疫情防控决策支持, 为落实“突出重点、统筹兼顾, 分类指导、分区施策”的重大方针提供技术支撑。

关键词: 人工智能; 大数据; 疫情防控; 疫情态势; 精准防控

中图分类号: TP18 文献标识码: A 文章编号: 1002-3321(2020)02-0019-06

2020 年新年伊始, 在湖北省武汉市发现的新
型冠状病毒^[1]所引发的肺炎疫情, 在全国范围内
蔓延。该病毒具有强传染性和传播力, 截止 2020
年 2 月 22 日, 我国累计确诊人数 76000 多人, 累
计死亡人数 2300 多人, 严重威胁我国人民的生命健
康, 给国民经济造成极大损失。因此, 及时掌握疫
情变化趋势, 识别高危群体、高危区域, 防止疫情
扩散等成为了疫情防控重点。

本团队长期从事大数据与人工智能前沿技术
的研究, 针对当前疫情防控的需要, 团队基于大数
据^[2]与人工智能^[3]技术, 提出疫情防控平台的建
设思路和实施方案, 利用团队在数据采集、数据预
处理、数据融合、数据存储、数据管理、数据计算、
数据分析以及可视化技术的研究成果, 构建“基于
大数据与人工智能技术的疫情防控平台”, 为各级
政府机构提供及时准确的疫情动态信息及决策支
持, 依托大数据与人工智能技术, 为落实习近平总

书记提出的“突出重点、统筹兼顾, 分类指导、分区
施策”的重大要求提供技术支撑, 做到有的放矢、
精准防控, 推动疫情防控与经济生产发展两手抓、
两不误。

一、平台架构

平台包括云平台^[4]、数据中台^[5]、AI 中台、
API 网关以及疫情大数据应用等模块。云平台为
疫情防控平台提供基础设施管理; 以数据中台进
行数据采集、数据预处理、数据融合、数据存储、数
据管理, 采集汇聚全国、本省、本市、疫情高发地
市的确诊、疑似、治愈、死亡、新增隔离人员与解
除隔离人员等数据, 实现分区域、分维度统计疫情
最新情况, 掌握疫情变化趋势; 以 AI 中台进行疫
情建模分析; API 网关提供分析结果汇聚服务, 为
顶层具体疫情大数据应用提供大数据与 AI 赋能,
实现高危群体、高危区域的精准识别, 疫情态势
的精确分析、疫情防控物资的精准把控, 为各级政
府机构

收稿日期: 2020-02-20

作者简介: 刘漳辉, 男, 福建惠安人, 福州大学数学与计算机科学学院副教授、硕士生导师, 硕士;

郭文忠, 男, 福建惠安人, 福州大学数学与计算机科学学院教授、博士生导师, 博士;

陈羽中, 男, 福建连江人, 福州大学数学与计算机科学学院教授、硕士生导师, 博士;

陈锋情, 男, 福建南安人, 福州市联创智云信息科技有限公司总经理, 硕士。

提供决策支持(见图 1) 。

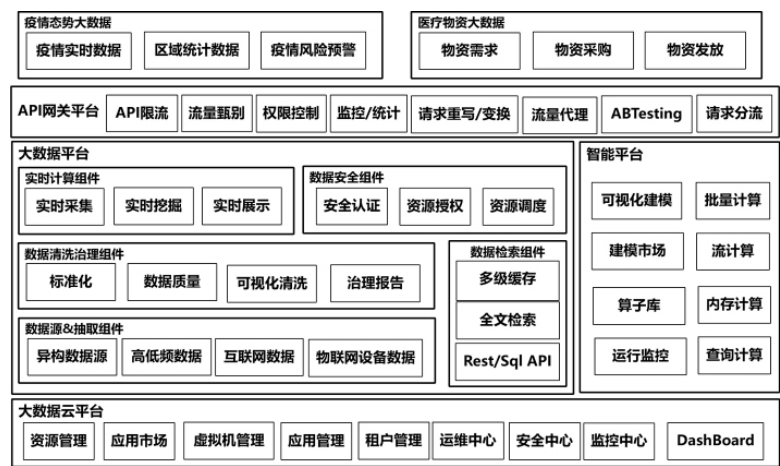


图 1 疫情防控平台架构

二、云平台

云平台基于企业级框架 SSM+Dubbo 构建 ,让系统的可靠性及健壮性得到进一步的提升和保证 ,本平台为 B/S 架构 ,可通过各种 WEB 浏览器访问客户端 ,而服务端是部署于 Linux 系统上的服务器容器中。

云平台底层兼容当今最为流行的三种虚拟化: Vmware ,OpenStack ,Docker。可在三种虚拟化方式中自由切换 ,用户可根据不同业务的特点 ,结合不同虚拟化方式的特性 ,创建所需的虚拟机。

云平台具有良好的用户体验 ,用户只需要简单几步操作 ,就可以完成创建虚拟机的操作。用户不再需要去创建及配置相应的框架环境 ,只需要在申请的过程中选择自己需要的框架版本即可 ,隐藏环境配置的复杂性。

对平台的应用包括纯净虚拟机、应用中间件、成型产品类三种类型 ,将疫情防控平台上架到云平台的产品中心 ,通过云平台可实现疫情防控平台的快速部署以及可视化运维(见图 2) 。

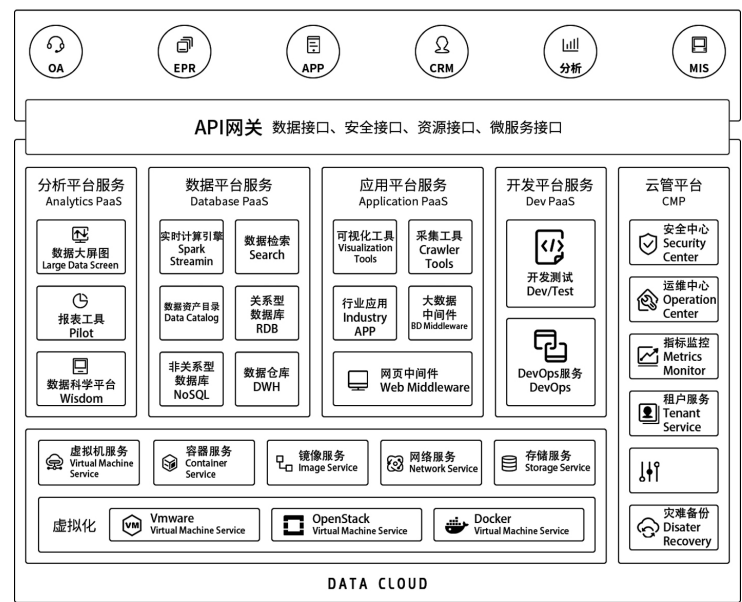


图 2 云平台架构

三、数据中台

数据中台基于团队自主研发的统一 SQL 解析引擎 ,实现数据源管理、数据抽取、数据清洗、数据

存储、数据共享一站式全流程数据治理^[6] ,以及数据安全、运维监控等管理。数据源管理模块实现各类数据源(MySQL、Oracle、PostgreSQL、MongoDB

等) 的接入; 数据抽取提供离线数据抽取、文件数据采集和实时数据采集三种数据抽取方式, 数据清洗模块对各个业务系统内的数据进行统一的标准化的, 去除重复数据、脏数据, 建立包含业务系统及信息化建设需要的标准共享数据库; 针对全数据类型和多样计算需求, 数据存储模块实现以海量规模存储、快速查询读取为特征, 实现面向多源异构数据的分布式数据存储; 数据质量管理模块通

过数据质量检测算法提供数据质量检测能力; 数据检索模块提供面向多模式异构数据的统一分布式数据检索能力; 数据实时计算模块实现时间窗口的统计和在线数据分析能力; 数据安全模块基于 Kerberos 等实现平台业务数据的多级访问控制; 数据运维管理模块实现集群的自动化部署以及管理, 同时提供操作系统级数据监控、JVM 级数据监控、SQL 级监控、应用级数据监控(见图 3) 。

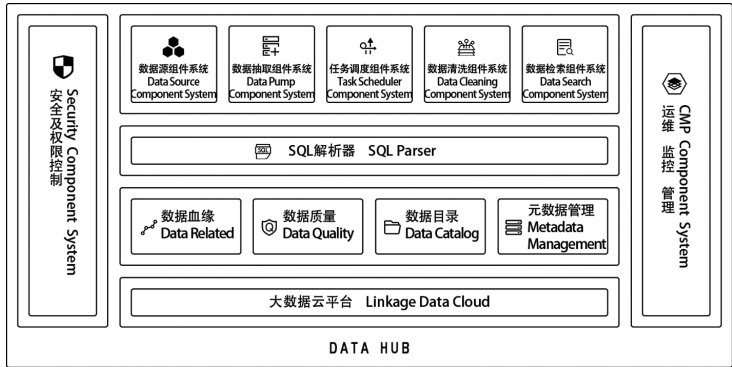


图 3 数据中台架构

(一) 疫情数据汇聚

通过数据中台, 可实现从各部门快速汇集疫情防控所需的多维度数据, 以实时指标和动态空间分布的模式, 集中汇总与关联, 做到实时性、准确性、全面性, 实现多维度、全感知的数据汇聚能力, 为疫情态势感知与监控、高危群体识别、多维度广角度决策分析提供基础数据能力。疫情防控平台所汇聚的数据如下:

- (1) 病患数据: 确诊病例数据、疑似病例数据、接触人员数据;
- (2) 医疗数据: 药品器械数据、留观隔离数据、医护人员数据、医疗物资数据;
- (3) 关联防控资源数据: 交通资源、仓储资源、社会治安执法资源数据;
- (4) 防控关联数据: 包括人群数据、运营商数据、交通出行数据、公安户籍数据、酒店/学校/体育场馆等可供人员临时安置的场所数据、社治排查数据。公安户籍数据提供各级区域常住人口、流动人口数据, 包括个人基本信息、群体关系数据。运营商数据提供基于运营商的基站定位数据, 获取区域人群密度、人群流动态势数据、个人活动轨迹数据。交通出行数据提供火车站、机场、汽车站的交通出行数据, 获得跨区的人员流动情况数据。酒店/学校/体育场馆数据提供酒店/学

校/体育场馆等人员临时安置的场所位置、场所面积、可容纳人数等数据。家庭用电数据提供基于电力公司统计的家庭用电数据, 获取居民短暂和长期外出、举家返回、隔离人员异动情况等。社治排查数据提供社治排查渠道获取的人群信息, 如外来人员户籍、外来人员出发地、到达地、居住地点、体温情况等(见图 4) 。

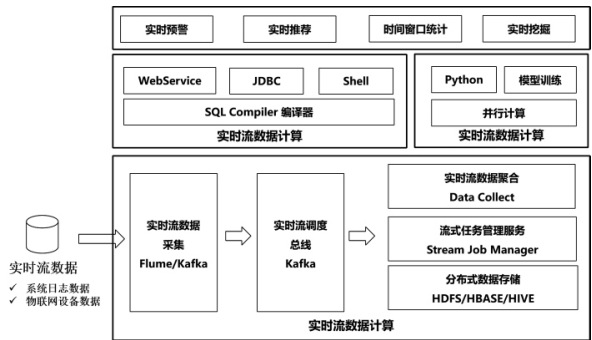


图 4 实时数据汇聚

(二) 疫情数据预处理

数据预处理模块负责对采集到的疫情数据的预处理: 通过双线性插值、异常点检测等数据预处理算法, 对数据采集模块传上来的数据进行异常分析和数据解析, 实现异常/噪声数据检测清洗、缺失数据补全以及异常数据去除; 通过数据转换组件, 实现不同类型、不同格式疫情数据的转换(见图 5) 。



图5 数据清理

(三) 数据管理模块

数据管理模块包括数据存储引擎与数据检索引擎两部分。数据存储引擎针对疫情防控平台的各类数据与多样化统计分析需求,以海量规模存储为特征,以 HBASE、Hive、HDFS、MySQL、Redis 等多种异构数据存储组件为核心,实现面向多源异构数据的分布式数据存储,可基于数据特征(来源、类型、长度等)选择不同存储策略,在读写效率与空间利用率之间达到平衡。数据检索引擎以 ElasticSearch 为核心,构建分布式数据检索引擎,实现多种异构数据的高性能检索。

四、AI 中台

AI 中台支持各类主流的关系型与非关系型数据库,基于 JBPM 工作流引擎实现数据分析流程的自定义,通过可视化流程配置,便捷完成数据处理、模型训练以及模型上线。中台内置丰富的预处理、分类、聚类、回归与验证算子以及一些专用领域的算子等,同时提供标准化接口,用户可以自定义拓展算子。中台提供基于 R 语言、Spark 的数据统计分析能力,基于 Mahout 与 Spark MLlib 框架的批量数据分析能力以及基于 Storm、Spark Streaming 的实时数据分析能力(见图6)。

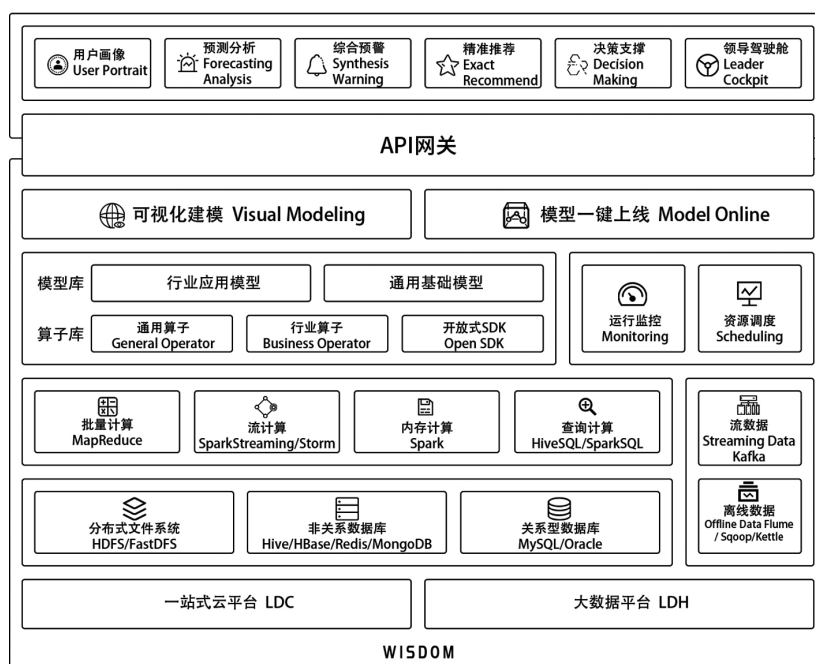


图6 AI中台架构图

AI 中台主要包括数据计算、流程配置、算子管理、建模任务管理、流程 API 发布等模块。数据计算模块包括数据离线处理引擎与数据实时处理引擎,实现基于并行处理技术和流计算技术的分布式、高可用、可扩展的离线数据与流数据的数据分析。流程配置模块通过使用 GooFlow 流程设计器实现 web 可视化拖拉设计模型流程,并保存相应

流程,通过 JBPM 进行流程执行。算子管理模块实现各类数据分析算子的添加、删除、启用、禁用、查看和授权管理。建模任务管理模块实现数据建模任务的管理与建模进度跟踪;流程 API 发布模块实现 API 创建、删除、启用/禁用、授权的全生命周期管理,可通过配置来源、目的源和选择可用流程实现 API 的发布(见图 7)。



图 7 数据建模

五、疫情防控分析

(一) 疫情态势分析

疫情防控分析模块从宏观角度,结合全国、本省、本市、疫情高发地市的确诊、疑似、治愈、死亡、新增隔离人员、解除隔离人员等数据,分区域、分维度统计疫情最新情况,掌握疫情变化趋势,为各级政府部门提供及时准确的疫情动态信息。

1. 疫情传播分析

疫情传播分析提供如下功能:

分析密切接触者转化为确诊病例的数量,以及感染途径(家庭传播、接触传播、同行传播、同住传播、同单位传播等),分析不同感染途径的转化率;通过可视化手段展示确诊人员与其密切接触者的接触场所、关系网络;帮助各级政府部门了解疫情的主要传播途径,调整防控策略。

2. 疫情风险分析

疫情风险分析模块提供以下功能:

潜在风险分析:基于社治排查数据、交通出行数据、运营商提供的人员轨迹数据,结合疫情暴发地以及其他疫情高发地市的确诊人数,通过跨区域人员流动分析定位重点潜在风险区域,并以可视化形式展示,实现疫情防护工作的有的放矢。

高危区域分析:根据确诊病例、疑似病例的户籍信息、居住场所信息、工作单位信息、以及运营

商提供的活动轨迹数据,获取其在不同地点、场所的停留时间,密切接触群体信息,通过热力分析,生成高位区域分布图,以实现高危区域的更高级别社区防控。

3. 高危人群监测

高危人群监测模块提供以下功能:

高危人群轨迹监测:根据运营商提供的活动轨迹数据以及人员的关系群体数据,实现对确诊病例、疑似病例的历史轨迹进行回溯监测,快速定位各层级的密切接触人员,可生成活动轨迹通报、人员分布热力图等。

高危人群精准排查:根据航班客运、列车客运、客车客运数据,以及疫情暴发地、其他疫情高发地市的手机号段,结合运营商提供的基站轨迹数据,快速定位来自疫情高发区域的人员群体,帮助各级政府部门确定重点排查群体、找到疑似排查缺口,实现及时精准排查。

人员筛查:基于数据规则引擎,可基于出发地、到达地、车次、航班号、出发时间、到达时间、户籍地、居住地等各项属性,自定义人员筛查规则,进行重点监控人群的灵活、快速筛查。

4. 病患分析

病患分析提供以下功能:

定点医院病例分布:按定点医院统计上报人

数、疑似病例数、确诊病例数、死亡病例数、危重病例数、重症病例数、轻症病例数以及相应比例,统计各定点医院的使用床位、未使用床位、使用床位比例。

病患报告医院确诊人员数:按病患报告医院统计确诊人员数量。

感染者年龄、性别分布:按照病例性别、不同的年龄段、感染时间、死亡时间、治愈时间、轻症转重症、重症转轻症等,进行多样化的分析统计。

(二) 医疗物资需求分析

医疗物资需求分析模块提供以下功能:

(1) 医疗物资供需分析:按照医疗机构,对疫情防控所需的各类物资如防护口罩、防护服、护目镜、呼吸机等,进行整合统计分析,分析库存量、消耗量、供应量等指标。

(2) 医疗物资采购分析:按照医疗机构,对疫情防控所需的各类物资的采购情况进行分析跟踪。

(3) 医疗物资下发情况:按照医疗机构,分析疫情防控所需的各类物资的下发情况、紧缺程度。

(三) 人员分布分析

人员分布分析模块提供如下功能

基于移动电话关联的数据资源,进行多维度的展现。通过运营商或通管局提供的武汉籍本地驻留人员市内分布情况,按照移动、联通、电信数据进行统计分析展现;通过运营商或通管局提供武汉入网用户在各个行政区划的人数分布情况,按照移动、联通、电信数据进行统计分析展现;运营商提供的武汉前来人员移动数据信息,统计驻留基站的相关内容。

针对列车客运统计:按照日期天,按照每日下车人数和到站人数进行统计、分析和运用;针对航

班客运统计:按照日期天,按照每日出港人数和入港人数进行统计、分析和运用;针对客车客运统计:按照日期天,按照每日出站人数和入站人数进行统计、分析和运用。

(四) 客运关联人员分析

客运关联人员分析模块提供如下功能:

(1) 基于疾控部门的确诊病例地址以及密切接触者地址信息,从互联网平台的地图服务中获取经纬度信息,基于经纬度数据结合 UGIS 形成新型冠状病毒高危区域热力图。

(2) 将社区排查的疫情暴发地来本地人数与运营商提供的暴发地手机用户进行比对分析,找出疑似排查缺口,为下一步排查提供指导。

(3) 从确诊病例的密切接触者着手分析,一旦密切接触者(主要为本地人员)转化成确诊病例,即开始分析该人员的密切接触者,构建其关系图谱,将其密切接触者纳为重点监控对象。

(4) 疑似病例和密切接触者的区域分布统计,监控可能爆发的风险点,为未来的医学观察点设置找点提供指导。

(5) 社区排查的疫情暴发地流入本地人数与运营商提供的暴发地手机用户进行比对分析,找出疑似排查缺口,为下一步排查提供指导。

六、结语

疫情防控期间,本文基于大数据与人工智能技术,构建疫情防控平台,为各级政府、卫健委、应急部门等提供疫情防控大数据服务。平台实现疫情数据的快速汇聚,一站式对疫情数据进行预处理,搭建疫情数据中台,同时结合 AI 中台进行数据建模,实现疫情分析。

注释:

- [1] Prof Nanshan Chen MD, "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study" *Lancet*, vol.395, no.10223(February 2020), pp.507-513.
- [2] Andrew McAfee, Erik Brynjolfsson, "Big Data: The Management Revolution" *Harv Bus Rev*, vol.90, no.10(2012), pp.60-68,128.
- [3] Antoni Ligeza, "Artificial Intelligence: A Modern Approach" *Neurocomputing*, vol.9, no.2(1995), pp.215-218.
- [4] Brian Hayes, "Cloud Computing" *Communications of the Acm*, vol.51, no.7(2008), pp.9-11.
- [5] 李炳森、胡全贵、陈小峰《电网企业数据中台的研究与设计》,《电力信息与通信技术》2019年第7期。
- [6] 张宁、袁勤俭《数据治理研究述评》,《情报杂志》2017年第5期。

[责任编辑:石雪梅]