

一种面向5G网络的人工智能平台

An Artificial Intelligence Platform for 5G Network

尹俊¹,高有利¹,刘贤松¹,盛刚²(1. 中国联通网络AI中心,上海200050;2. 科大讯飞软件股份有限公司,安徽合肥230000)

Yin Jun¹,Gao Youli¹,Liu Xiansong¹,Sheng Gang²(1. China Unicom Network AI Center,Shanghai 200050,China;2. USTC Sinovate Software Co.,Ltd.,Hefei 230000,China)

摘要:

首先分析了人工智能在运营商网络建设、规划、运营、优化、维护等方面的重要意义,以及使用人工技能技术的痛点,横向对比互联网厂商以及各大运营商的AI平台建设思路后,详细介绍中国联通面向网络的AI平台的建设目标、建设情况以及创新举措,最后展望网络AI平台未来的演进计划。

关键词:

人工智能平台;模型服务;模型生成服务;数据服务;自动学习;深度学习

doi:10.12045/j.issn.1007-3043.2020.10.003

文章编号:1007-3043(2020)10-0010-05

中图分类号:TN915

文献标识码:A

开放科学(资源服务)标识码(OSID):



Abstract:

Firstly, it analyzes the importance of artificial intelligence in the construction, planning, operation, optimization, and maintenance of operators' networks, and the pain points of using artificial skills. After horizontally comparing the AI platform construction ideas of Internet manufacturers and operators, it introduces in detail the construction goal, construction situation and innovation measures of China Unicom's network oriented AI platform. Finally, the future evolution plan of network AI platform is prospected.

Keywords:

Artificial intelligence platform; Model service; Model generation service; Data service; Automatic learning; Deep learning

引用格式:尹俊,高有利,刘贤松,等.一种面向5G网络的人工智能平台[J].邮电设计技术,2020(10):10-14.

0 引言

AI已经成为国家战略,同时也是数字经济发展的重要引擎。AI对于运营商网络的重要性不言而喻,不管是在规划、建设、维护、优化,还是面向市场的支撑方面,都发挥着重要的作用。

当前,运营商的各个部门围绕生产经营中的热点和痛点,纷纷进行了一些人工智能领域的探索和尝试^[1-4]。但是在取得这些成果的同时,我们也看到网络AI的需求比较多,覆盖也比较广,在部署过程中依然存在一些问题。第一是网络AI开发的难度比较大,运

营商当前的人才储备难以满足大规模AI开发的需要;第二是AI的应用场景还有待丰富,相对于传统AI领域,运营商的网络AI应用场景更为广泛;第三是AI部署的协同不足。运营商的特殊组织架构关系,导致了集团公司与省分公司之间协同不足,省分和省分之间协同也不强,可能一个省分自己做了AI应用,没有办法共享经验,没有办法简单地应用到其他省公司,这样可能会造成一定的重复开发与资源的浪费。

所以,急需一个面向运营商网络的人工智能平台去解决上述问题,推动网络AI做深做透做出规模。

1 人工智能平台的发展现状

随着人工智能技术的不断发展,各大厂商都投入

收稿日期:2020-08-28

相当大的资金与精力到AI的开发与研究之中,这其中互联网企业首当其冲。作为展示其AI能力,提供落地化产品的重要手段,AI开放平台成为了各大企业的标配。而作为当前互联网行业公认的三大巨头百度、阿里、腾讯(BAT)的AI开放平台有着相对最为丰富的内容和资源。

百度AI开放平台有着最为完整的AI生态体系^[5]。从AI开放能力,到开发平台以及行业应用,同时包括AI硬件产品以及AI教学支持等,形成了从线上到线下,从研究到开发到落地的一整套体系。

阿里云开放平台,依托阿里云服务器,形成了一套以云服务器为特色的AI开放平台。从基础服务器到上层应用,阿里开放平台能够以具体行业为定制目标,形成一套完成的从部署到应用的AI应用体系。

腾讯AI开放平台^[6],依托其三大AI实验室,AI LAB、腾讯优图、微信实验室,为AI应用提供了强大的技术支持,能够提供较为完善的通用能力体系并提供以硬件为依托的智能应用平台。

除了互联网企业外,通信设备厂商、运营商等也有包含各自特色的AI开放平台。华为NAIE便是一个将AI引入网络领域^[7],解决网络业务预测类、重复性、复杂类等问题,提升网络资源利用率、运维效率、能源效率和业务体验,使能自动驾驶网络的AI平台^[8]。而中国电信和中国移动分别有以“灯塔”和“九天”为主打品牌的AI开放平台,为各自的网络AI业务赋能。

2 中国联通网络AI平台的功能架构

中国联通网络AI平台的核心功能规划紧紧围绕AI开发者和AI服务的使用者构建。平台的主要功能架构如图1所示。

2.1 基础资源层

基础资源层是整个AI平台的算力基础、数据基础

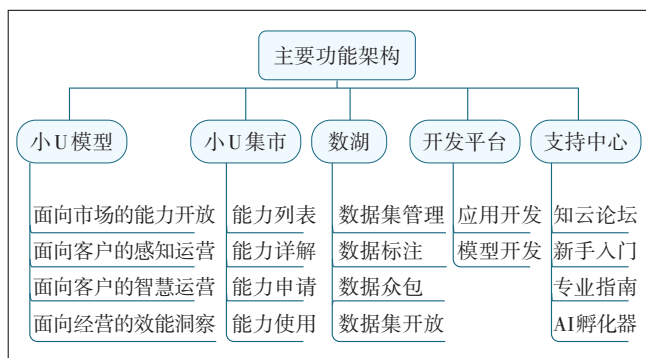


图1 平台主要功能架构图

和算法基础。基础资源层提供了数据挖掘、深度学习等计算所必须的CPU、GPU、内存等IT硬件基础设施。在IT基础设施之上,平台提供了资源虚拟化、环境隔离等资源管理工具,方便模型训练和推理动态地申请、分配以及释放资源,实现整个IT资源利用率的最大化。同时通过环境隔离,解决了不同算法之间的运行环境的冲突问题。数据存储为上层的模型、应用、数据提供了稳定冗余的存储环境,保证平台数据和模型的安全。AI框架为虚拟化的容器资源提供了AI算法所依赖的最基础的算法框架,比如cuda/cuDNN、tensorflow^[9]等。

2.2 模型开发层

模型开发层为AI开发提供了一套完整的可视化的流程。传统的AI模型开发包括数据获取、数据清洗、特征工程、模型构建、参数调优、模型训练、模型评估、模型封装、AI能力开放等环节,而模型开发层则是将这些底层繁琐冗杂的开发细节全部屏蔽,用户只需要关注需要处理的数据、算法以及需要调整的参数即可,模型开发层的功能关系如图2所示。

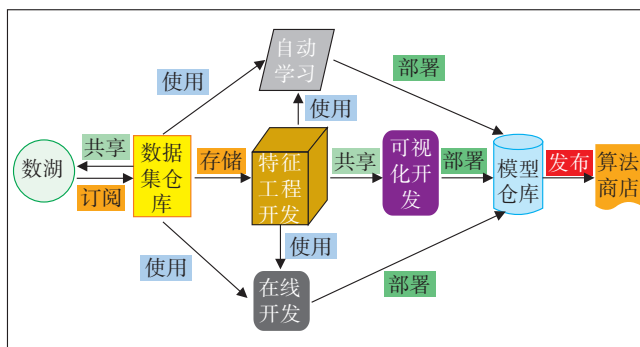


图2 模型开发层的功能关系图

2.2.1 自动学习

自动学习包括了图像分类^[10]、预测分析、时间序列预测、关联规则挖掘等算法的自动学习。图像分类自动学习提供了alexnet、inceptionV4、resnet等预置算法,用户可以根据自己的业务数据选择相应的算法,进行简单的迭代次数调整即可提交训练任务。预测分析针对结构化数据的分类和回归,基于xgboost、svm等主流算法,通过内置贝叶斯超参数搜索算法和模型预评估选举策略,帮助用户自动地从候选算法中选择一组最优的参数进行模型构建。

2.2.2 可视化开发

可视化开发相对于自动学习较为灵活,可视化开发通过托拉拽的形式配置一个完整的算法开发流程。

主要包括数据集加载、数据预处理、特征工程、机器学习、预测、评估等类型的节点。其中,数据预处理和特征工程针对数据提供了缺失值填充、异常值处理、删除重复行、删除列、数据拆分、随机采样、数据过滤、归一化、标准化、数值化映射等功能节点。机器学习则涵盖了分类、回归、聚类等各类别下的逻辑回归、lightGBM、SVM、Xgboost、岭回归、KMeans 聚类算法^[11]。

2.2.3 在线开发

在线开发为专业的AI开发者提供了独立隔离的代码开发、调试、运行环境。在线开发集成了jupyter-Lab工具,预置了sklearn、pandas等常用算法开发包以满足常见的算法开发需求。同时,平台支持自定义开发环境以满足特殊算法环境下的开发和测试。

算法开发调试完成后,平台会接管模型训练,生成的模型会自动按照平台标准封装成模型。平台支持多种模型封装标准,除了平台本身定义的模型封装标准之外,还兼容了AcumosAI、普通SpringBoot微服务等格式。平台的模型以标准压缩包、docker镜像2种方式存储和管理。其中docker模型镜像不仅仅封装了已经训练完成的模型和调用脚本,还集成了运行该模型所依赖的Runtime。模型镜像在通过kubernetes调度拉起为Deployment实例后,经过ingress代理,开放为标准RESTful风格的模型接口服务。服务再经过API网关的注册,对外进行AI能力开放以及鉴权控制。

2.3 应用开发层

应用开发层是以开发层生成的AI能力为基础能力,结合元数据开发、界面定制开发、流程开发等工具,快速地基于平台生成一个创新AI应用。

2.4 数湖

数湖是模型开发的基础。数湖为模型开发提供了丰富的标准网络侧数据集,用户可以通过平台提供的数据集开发网络侧的AI模型而不再需要自己准备数据。平台提供针对图像和结构化数据进行图像分类标注和结构化数据标注的功能。对于平台不具备的原始数据,可以基于标注功能通过多人协作完成数据的人工打标签,然后生成为标准化的数据集。用户自己创建的数据集可以自己使用,也可以经过脱敏后开放给其他用户使用。

2.5 小U集市

小U集市为AI服务使用者提供了面向网络的智慧运营、面向市场的能力开放、面向客户的感知运营、面向经营的效能洞察4个方面的典型网络AI应用和

网络AI能力,包括3G/4G/5G无线小区价值度预测、网络信号预测、OLT上联口流量预测、KPI异常检测^[12]等AI能力。此外,为了满足省分公司开发创新应用,平台还封装了语音识别、OCR^[13]、身份证检测、二维码/条形码识别^[14]、人脸识别和比对等通用的AI能力。用户可以在线查看这些能力和应用的业务场景、接入方式、调用文档,在线体验模型效果,并且申请使用这些模型能力。

3 中国联通网络AI平台的技术架构

人工智能技术和应用的快速发展给人工智能平台带来了强烈需求,也要求平台能够快速解决计算能力的供应、多样化数据的处理、算法模型的迭代开发、以及模型的封装部署等一系列问题。在其技术上,主要分为平台基础技术、算法优化技术以及模型封装开放技术。

3.1 平台基础技术

平台基础技术是作为搭建整个平台的必备条件,从技术架构上可以将基础技术分为数据存储层、资源调度层、AI计算层、能力开放层以及平台DevOps和安全管理功能。具体技术架构如图3所示。

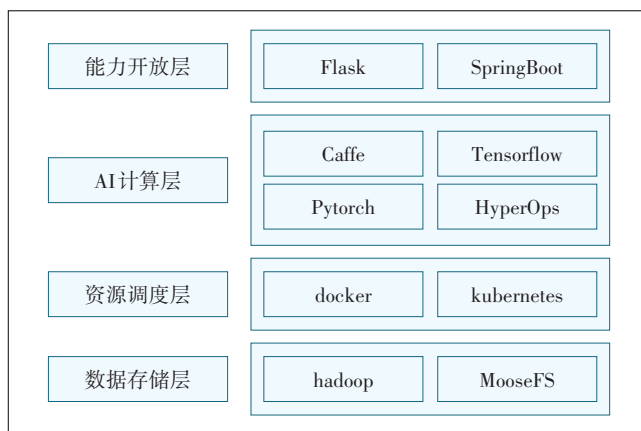


图3 技术架构图

3.1.1 数据存储层

数据存储层为平台提供结构化数据存储、对象存储以及文件存储功能。采用传统的关系型数据库提供对结构化业务数据的查询和存储;依托于大数据技术和分布式数据采集组件,完成对外部数据源的适配接入与存储;使用分布式文件系统MooseFS与虚拟化容器环境结合的形式,实现对平台内标准数据集、模型文件以及训练部署模型中间结果的存放和处理工作,满足AI平台的高扩展性、高可用性等存储需求。

3.1.2 资源调度层

平台基于虚拟化和容器技术,提供弹性资源的计算。采用 docker 技术将模型封装成虚拟镜像,使用 Harbor 为平台提供模型镜像仓库,利用 YARN 的资源调度功能实现对 CPU、内存以及 GPU 等资源分配和智能调度,基于 kubernetes 对容器的管理操作技术,完成对算法模型训练、开发或部署过程中需要的容器进行统一管理和监控工作。

3.1.3 AI 计算层

AI 计算层为平台提供核心的机器学习和深度学习算法框架,平台利用成熟的算法框架和 5G 网络领域大规模的数据集提供预先训练调优好的专业化预置模型,帮助运营商的 5G 网络 AI 应用快速落地。例如用于图像分类的 Inception V4、darknet_53、ResNet_v2_50 等视觉处理预置模型,检测物体类别和位置 yolo V3、faster rcnn、inception_resnet_v2 等图片处理预置模型,还有包括分类 LightGBM、SVM、XGBoost 等众多的机器学习预置模型,这些模型不仅可以直接使用,还可以在其基础上再迭代训练,方便用户根据特定业务数据对模型进行优化调整。

3.1.4 能力开放层

平台采用分布式微服务架构,根据功能划分使用 springcloud 微服务套件将每个功能模块做成独立的微服务对外开放。使用 Nacos 作为注册中心和位置中心,用于微服务的注册、发现和动态配置组件;Feign 为微服务提供声明式 WebService,让微服务调用更加简单;Ribbon 实现微服务调用的负载均衡;Hystrix 解决底层服务的故障有可能引发级联故障的雪崩问题;SpringCloud gateway 作为所有微服务的统一网关路由,为每个微服务提供身份认证与安全控制,保证整个平台的安全稳定。

3.1.5 DevOps 和安全管理

平台采用 Gitlab 管理代码、Jenkins 持续集成、Selenium 自动测试、禅道问题反馈与项目管理、Ansible 自动部署等组件相结合的方式,实现系统的 DevOps。同时在平台层面,使用包括 shiro 安全认证、RSA 签名加密完成自身安全验证控制,在服务器层面,使用 IDS、WAF、RASP 等技术实现系统整体的安全防御。

3.2 算法优化

近年来,随着人工智能的不断发展,相关机器学习和深度学习框架层出不穷,特别是 Tensorflow、Caffe、Torch、MXNet、PyTorch 等主流深度学习框架广泛用于

各个领域,这使得平台需要适应多样化的算法框架。平台根据目前市场上常用的算法框架内置了一批已经优化后的基础运行镜像,用户在开发、训练、部署算法时,无需再考虑其环境资源问题,只需根据算法框架选择对应的运行镜像资源即可完成算法开发工作,使得平台更高效易用,大大提升模型优化速度。

3.3 模型封装开放

为了进一步降低使用门槛,平台针对传统的模型开放 API 进行了系统封装,制定了模型开放入口和请求参数规范,用户只需按照规范编写算法代码,系统会自动采用 Flask 完成对模型的封装,并且适配多种请求参数类型,如普通文本、文件流以及文本+文件流组合形式。针对市场上广为流传的 AcumosAI 模型,平台也做了无缝兼容,无需任何改造即可直接加载至平台中使用,并且平台提供与 AcumosAI 模型相互转换的快捷方式,便于与第三方人工智能平台的交流合作。

4 平台的创新点

本文针对中国联通面向网络的 AI 平台的设计与实现进行了探讨,中国联通网络 AI 平台的建设初衷并不是只为拥有深厚 AI 经验的用户提供服务,而是通过降低 AI 建模门槛,为广大非专业 AI 开发者和 AI 服务使用者提供一个易上手、体验好的人工智能平台。

考虑到 AI 建模流程的严谨性、复杂性,如果不具备一定 AI 基础的话,想训练一个自己的建模十分困难。本平台为了降低用户的 AI 建模门槛,采用“自动学习”和“可视化开发”的方式帮助用户快速建模。如果用户不具备 AI 基础的话,可使用“自动学习”的建模方式,只需设置一些基本的参数,平台便会自动进行模型训练并最终得到一个最优模型;如果用户具备一定的 AI 基础,“可视化开发”的建模方式将建模过程中涉及到的各个环节均封装成了“黑盒”形式,用户根据实际需要进行各个环节的自由组合,在快捷建模的同时,帮助用户更加深刻理解建模流程。

许多用户都面临着手中有数据但自己进行 AI 建模困难大的问题,使用平台的“模型生成服务”,可以参考平台已有的相同或相似模型。用户只需按照已有模型的要求将数据上传到平台,平台便可基于上传的数据并仿照已有的模型训练得到用户自己的模型,实现定制化建模。

建模不是目的,利用模型解决生产生活中的问题才能体现模型的价值。但不同的 AI 开发者使用的算

法、编程语言以及拥有的模型运行环境不尽相同。通常,用户在本地训练的模型可能换个环境就无法运行,利用模型解决实际问题便无从谈起。针对这个问题,平台提供了标准化的模型封装和部署功能。如果用户是在平台中进行建模,建模完成后便可立即部署模型,平台会根据模型功能生成相应的API接口以供调用;如果用户在本地建模,只需按照平台的要求,在本地对模型进行封装后上传到平台进行部署以生成API接口,如此,不论模型需要什么样的运行环境,其他用户都可通过平台提供的API接口使用模型提供的能力。本平台通过这种方式打通了算法开发和工程应用之间的壁垒,更好地体现了模型的价值。

5 平台上线成果与未来展望

网络AI平台在2020年5月15日举办的智网创新中心智网中台发布会上正式发布。截至2020年7月21日,平台累计访问量23万次,用户量1115个。

5.1 模型研发与上线

自平台上线以来,已累计研发上线AI模型78个(含前期模型自研成果)。涵盖了面向网络的智慧运营、面向市场的能力开放、面向客户的感知运营、面向经营的效能洞察四大维度。高热度模型如表1所示。

表1 高热度模型列表

模型名称	需求量	模型名称	需求量
2G高价值用户识别	19	离网用户预测	6
告警关联规则挖掘	12	投诉工单智能分类	5
携号转网用户预测	6		

5.2 开发平台使用

自平台上线以来,吸引了大量算法开发人员以及AI技术爱好者使用。具体项目数量如表2所示。

表2 开发平台项目数量表

模块	数量	模块	数量
数据标注	35	可视化开发	97
特征工程	58	在线开发	105
自动学习	275		

从项目数量上来看,在运营商的人员体系下,简化的自动学习功能更受开发者的欢迎。依托平台的架构以及强大的CPU&GPU算力支持,极大地提升了开发者的开发效率。

5.3 未来展望

本次开发的基于5G网络人工智能平台主要面对

运营商常见问题,另外数湖中的行业数据是人工离线上传到平台中,下一步将研究如何将数湖中数据规模扩大,如实时在线连接到更多行业数据库,解决更多领域、更多用户切实需要解决的问题,从而服务更多的人群,实现更大的价值,也实现AI的服务能力、解决能力。

参考文献:

- [1] 张青. 基于AI的5G基站节能场景智能识别[C]//5G网络创新研讨会(2019)论文集. 2019.
- [2] 严斌峰,加雄伟,闫硕. 面向智慧城市应用的人工智能服务运营方案探讨[J]. 信息通信技术,2019,13(1):7-13.
- [3] 陈红星. 人工智能在通信运营商网络中的应用[J]. 中国新通信,2020,22(11):38.
- [4] 刘漳辉,郭文忠,陈羽中,等. 基于大数据与人工智能的疫情防控平台建设方案[J]. 福州大学学报(哲学社会科学版),2020,34(2):19-24.
- [5] 王秋蓉,于志宏. 担负起AI造福人类的时代之责——访百度集团副总裁、深度学习技术及应用国家工程实验室副主任吴甜[J]. 可持续发展经济导刊,2020(4):21-23.
- [6] 佚名. Qualcomm、vivo、腾讯王者荣耀和腾讯AI Lab强强联合[J]. 中国电子商情(基础电子),2019(4):46.
- [7] 李建东,盛敏,文娟. 基于非完备大数据的业务预测[J]. 中兴通讯技术,2019,25(1):63-66.
- [8] 佚名. 华为发布自动驾驶网络解决方案[J]. 智能城市,2019,5(21):66.
- [9] A ABADI, MARTÍN. TensorFlow: learning functions at scale [J]. Acm Sigplan Notices,2016,51(9):1-1.
- [10] KRIZHEVSKY A, SUTSKEVER I, HINTON G E, et al. ImageNet Classification with Deep Convolutional Neural Networks [C]//neural information processing systems,2012:1097-1105.
- [11] 朱天元. 机器学习算法在数据挖掘中的应用[J]. 数字技术与应用,2017,000(003):166-166.
- [12] 纪勇. 基于频繁模式的KPI异常检测研究[J]. 无线互联科技,2016(15):115-118.
- [13] 郭军. 基于数字图像处理的文本型数字图像OCR识别准确度提高策略研究[J]. 网络安全技术与应用,2017(6):143-145.
- [14] 杨勇. 基于机器视觉的条形码及二维码识别系统的设计[J]. 微型电脑应用,2018,34(11):67-70.

作者简介:

尹俊,毕业于中国传媒大学,中国联通网络AI中心算法工程师,硕士,主要从事面向运营商网络的算法开发工作;高有利,毕业于东北农业大学,中国联通网络AI中心高级软件开发工程师,学士,主要从事网络应用项目管理工作;刘贤松,毕业于武汉大学(武汉大学),硕士,中国联通网络AI中心副经理,主要分管网络服务智能化产品研发和网络服务智能化产品推广管理工作;盛刚,工程师,硕士,主要从事面向运营商的企业级分布式应用的架构设计及研发工作。