

一种使能5G的AI深度学习训练 与服务平台

An AI Deep Learning Training and Service Platform Driving 5G

杨平¹,裴霁²(1. 中博信息技术研究院有限公司,江苏南京210012;2. 中国移动江苏公司常州分公司,江苏常州213000)
Yang Ping¹,Pei Ji²(1. Zhongbo Information Technology Research Institute Co.,Ltd.,Nanjing 210012,China;2. Changzhou Branch of China Mobile Jiangsu Company,Changzhou 213000,China)

摘要:

以深度学习为代表的AI技术可以广泛用于5G网络的故障分析、网络覆盖、性能优化、网络容量预测、精准网络建设、能耗管理、资源调度、网络切片、客户服务等领域,从而提升网络运维效率,降低网络运维成本。研究并实现了一种端到端的AI深度学习训练和服务平台,从而为5G各类应用服务的快速构建提供支持。

关键词:

AI;5G;深度学习

doi:10.12045/j.issn.1007-3043.2020.12.014

文章编号:1007-3043(2020)12-0061-04

中图分类号:TN929.5

文献标识码:A

开放科学(资源服务)标识码(OSID):



Abstract:

AI technology represented by in-depth learning can be widely used in 5G network fault analysis, network coverage, performance optimization, network capacity prediction, precise network construction, energy consumption management, resource scheduling, network slicing, customer service and other fields, so as to improve the efficiency of network operation and maintenance and reduce the cost of network operation and maintenance. It aims to realize an end-to-end AI deep learning training and service platform, so as to provide support for the rapid construction of 5G various application services.

Keywords:

AI;5G;Deep learning

引用格式:杨平,裴霁.一种使能5G的AI深度学习训练与服务平台[J].邮电设计技术,2020(12):61-64.

0 引言

近年来,AI技术应用于通信网络的研究取得了重大进展,在网络运营编排、运维管理、智能化网元3个层次上,挖掘出越来越多的场景。但在针对5G所要求的快速响应业务变化、简化网络运维管理、提升网络资源利用效率等爆发性需求,运营商还要能够持续增加业务收入、降低运维成本、优化投资收益,AI则成为电信运营商在应对这些挑战时必须利用的工具之一。

因此,如何最大化地利用AI工具使能5G,推行一

种使能5G的AI深度学习训练与服务平台已成当务之急。

1 AI使能5G的应用策略

1.1 基于AI的智能编排和运营

5G端到端切片智能编排,即AI工具通过采集数据掌握网络的实时运行状况,根据历史数据和实时数据对网络业务以及相应的资源需求进行预测和评估,给出适当的建议措施(如网络切片的扩容、缩容、变更等)。

5G端到端切片智能运营,即可以通过AI完成智能化的端到端切片业务开通和按需变更。同时引入

收稿日期:2020-10-13

智能客服,能够提供智能化的交流、咨询、切片套餐推荐服务。

1.2 基于AI的智能天线优化

5G引入Massive MIMO技术后,无线侧配置参数的pattern组合有了指数级的增加,从3G的几百种配置到5G的上万种配置。AI可以实现5G大规模天线复杂参数的智能化配置。

a) 智能权值搜索和监控。基于UE的分布情况,根据覆盖用户数最多的原则,搜索和预测最优的水平/垂直波瓣宽度、方位角和下倾角。

b) UE位置估算和预测。基站周期性收集一段时间内小区内所有UE的位置信息,基于收集的信息可以估算UE的位置及分布。

c) 场景自学习。利用测量信息描绘出终端的大致分布,进一步利用分布识别场景。根据不同场景,推荐最优权值,并将最终的权值反馈到场景识别模块,使得推荐不断进化。

1.3 基于AI的智能边缘计算

边缘计算具备支持AI运算的能力,使得可以在边缘节点上,配合中心的DC以及用户的终端来做AI业务的智能优化。

a) 本地缓存。基于AI对用户的业务流和用户移动模式进行预测分析,有针对性地确定预存内容和推送内容,从而提高内容分发效率。

b) 智能定位。通过位置已知的终端测量的各无线通信系统信号特征,借助AI、大数据收集分析和边缘计算节点的实时计算能力,利用指纹信息指导实际应用中的终端定位。

c) 频谱感知。边缘计算节点基于不同无线系统的频谱测量结果,利用AI技术对各无线系统在不同区域的无线环境特征、用户行为特征进行分析建模,支撑具体应用场景。

d) 业务感知。在边缘节点上部署高算力的硬件解析资源,结合AI和大数据能力,分析挖掘数据、业务和无线环境之间的内在关联,提供更为准确的业务特性识别。

1.4 基于AI的智能运维

a) 网络健康度检查。基于大数据和人工智能技术的网络健康度分析,预测网格内未来某周期内的小区网络质量并进行预警,针对质差小区,分析引起质差的关键指标及可能的原因。

b) 网络告警关联和故障定位。传统网络运维管

理人员分析网络警告、判断告警原因、查找告警根源、定位并排除故障,耗时耗力。神经网络系统通过不断学习和训练,计算、翻译和调整分布于神经网络当中的连接权值,以整体的方式表达关联规则和故障诊断结果,准确定位网络故障。

c) 基于AI的智能基站节能。在5G基站中,通过AI辅助的业务分析、场景识别建立一个流量变化模型,来控制载波的智能关断,从而降低基站功耗。

2 深度学习训练与服务平台

2.1 总体技术框架

基于AI的深度学习训练与服务平台利用AutoML等相关技术,通过硬件虚拟化、算法服务化以及软件平台化等相关技术,在支持Tensorflow、Torch/Pytorch、MxNet、PaddlePaddle等深度学习框架的基础上,提供通用的网络模型(如VGG、LSTM、seq2seq、Inception-Net、mobileNet)支持,并支持多种不同数据结构的数据预处理和标注。

整个平台包括硬件层、数据/算法层、模型训练层以及模型部署层,以及贯穿各个不同层次管理相关功能,如图1所示。

硬件环境层主要通过虚拟化的方式提供本地的GPU计算环境,或者直接访问云端的相应计算环境。

数据与模型层主要包括样本数据中心和模型算法管理中心。样本数据中心从样本的种类来说,支持图像样本数据和文本样本数据,从数据模型角度来说,支持文档数据、图数据库、对象数据库等。模型算法管理主要包括模型库和相关的模型配置库。模型库中的模型支持tensorflow、pytorch、Caffe以及CNKT等多种不同的深度学习框架,并涵盖了多种常用的网络结构。平台将提供对样本数据以及模型的相关管理功能,如查询、更新、新建、导入导出等。

模型训练与优化层主要是算法的开发人员根据应用需求,检索相应的模型以及模型配置信息,采用相应的数据标注工具,生成特定的标注数据,并基于此数据,选定特定的模型以及相应的实现框架,通过人工或自动的方式对模型的超参数进行选择,并对模型训练结果进行相应的评估,来实施对模型超参数(如学习率、隐含层层数,卷积核大小等)的调节,从而构建出最优化的深度学习模型。

模型部署与服务提供层主要针对最终的模型用户。该层主要将模型训练层所得到的最佳模型进行

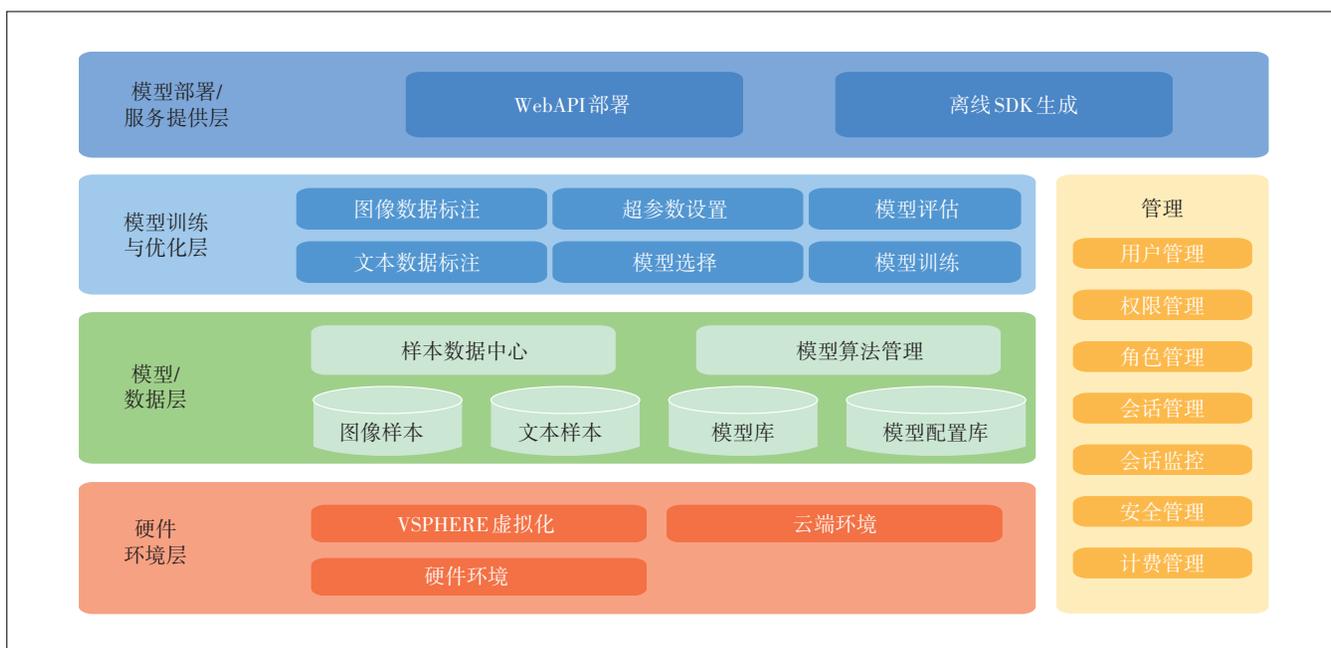


图1 平台总体框架

打包部署。部署的模式有2种:基于WebAPI的部署和离线SDK部署。基于WebAPI的部署将模型部署在平台环境中,并利用平台所提供的软/硬件环境和相关的计算资源,为用户提供服务,用户可通过OpenAPI的方式来调用和使用服务。离线SDK将和模型运行相关的所有软件环境打包,并通过SDK的方式部署到用户的生产环境中,利用用户自身的硬件资源来进行计算。离线SDK可以脱离本平台环境独立使用。

2.2 平台主要功能

端到端的深度学习训练和服务平台将在分析现有模型和框架的基础上,为用户提供统一的模型和框架选择结构,并进一步给出模型所需要设定的超参数集合以及参数的选择范围。主要包括以下功能。

a) 数据处理与数据标注。完成对样本数据的标注处理,并将标注后的数据根据模型需要转换成特定的格式。深度学习模型的训练和应用依赖于海量的数据积累,特别是对于模式有监督学习,还需要对海量数据进行标注。深度学习所采用的网络模型和算法根据确定样本数据的格式和标注方法,我们将通过分析现有的模型和样本数据的关系,建立起常用的数据格式,并提供相应的样本标注工具来生成相应的标注数据。项目将根据业务的需求构建多种不同的样本标注工具,包括图像的标注工具,文本的标注工具(包括文本的意图标注、文本的情感标注、文本的命名实体标注、文本的语言模型标注等)。

b) 模型选择与优化。根据应用需求,选定相应的深度学习框架和网络模型结构,并对模型的超参数进行调优,内容涉及模型的训练、评估、对比以及模型调参。随着深度学习技术的不断发展,深度学习框架越来越多,不同的框架所生成的模型都有所不同,需要支持这些不同框架所提供的深度学习模型。另一方面,为了满足业务的需求,研究人员已经提出了大量的网络模型。针对不同的业务需求,开发人员需要对网络机构和深度框架进行选择,该选择需要开发人员具有全面的网络和框架基础。平台在分析现有模型和框架的基础上,为用户提供统一的模型和框架选择结构,并进一步给出模型所需要的设定的超参数集合以及参数的可能选择范围。

c) 模型的部署和服务。根据应用的需求,将训练好的模型进行一键部署,为用户提供WebAPI或者离线SDK供客户端调用。

2.3 重点解决问题

平台针对特定的5G应用领域,提供端到端的深度学习训练和服务,降低深度学习模型构建和训练的门槛,重点解决了以下问题。

a) 标准化的数据管理与数据标注,为深度学习提供标准的数据样本。

b) 智能化的模型选择与模型优化,实现深度学习模型超参数的自动选择和优化。

c) 自动化的模型部署与服务提供,提供模型的部

署意见,并按需提供模型服务功能。

3 实际应用案例

平台使用户可以快速地基于自己的数据来构建并训练得到满足用户需求的最佳深度网络模型,最终开发5G领域的人工智能应用。

实际应用案例为某省运营商的基于训练服务平台的智能电话机器人系统。

本系统是智能机器人+呼叫系统+报表展示三位一体的智能调研应用,助力企业提升用户回访效率,降低成本,实现降本增效。模型的核心是让机器人正确理解用户所希望表达的意图,并基于理解的结果根据业务逻辑作出相应的反馈,目前在意图识别中采用了基于CNN、RNN/LSTM等多种不同的模型结果,如图2所示。

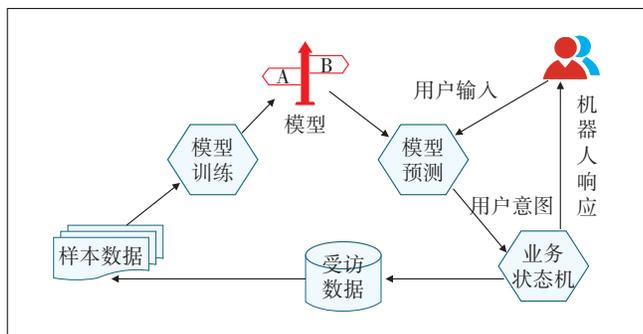


图2 智能机器人模型

为了实现业务逻辑和算法之间松散耦合性,我们利用有限状态机的方式来对不同的业务场景进行建模,如图3所示。

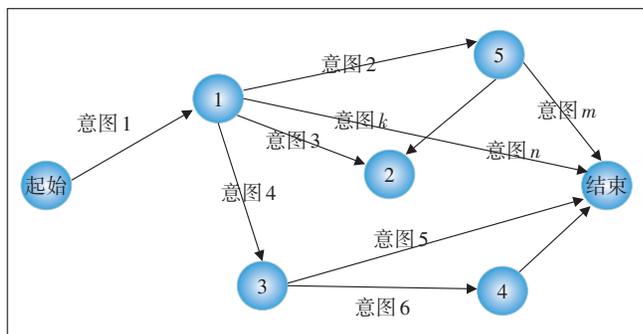


图3 业务场景建模

系统模拟测试用例:某省运营商综合满意度回访,问卷内容为综合满意度、NPS、各服务环节满意度+不满原因追问,成功样本量达到1 000本。

经过测试,达到以下指标。

- a) 机器人电话接通率与人工相当,达到50%。
- b) 机器人问卷成功率为14%,略高于人工(10%)。
- c) 机器人成功样本平均用时6 min,低于人工的7~8 min。
- d) 机器人4条线路,每条线路日均成功量30个,日均成功量共计120单,执行周期从原来13天(4个人工客服)压缩为8天。

4 结束语

本文研究并实现了一种使能5G的AI深度学习训练与服务平台。该平台经通信实际项目应用检验,具备海量、高并发、安全可靠的运行能力,可广泛应用于5G建设与维护领域。

参考文献:

- [1] 王海宇,贾鹏,曹宇诗,等.基于IBN的5G网络管理系统的调度算法研究[J].电子技术应用,2019(10):5-10.
- [2] 尤肖虎,张川,谈晓思,等.基于AI的5G技术——研究方向与范例[J].中国科学:信息科学,2018(12):1589-1602.
- [3] 张嗣宏,左罗.基于人工智能的网络智能化发展探讨[J].中兴通讯技术,2019(63):57-62.
- [4] 梁力维,李荣聪.基于AI的5G基站节能技术研究[J].移动通信,2019(12):32-36.
- [5] 党凯乐.基于神经网络的社交网络信息传播预测算法研究[D].南京:南京邮电大学,2016.
- [6] 朱雪田,张志荣,王旭亮.基于人工智能的5G大规模天线模式配置方法研究[J].电信技术,2019,538(1):23-26.
- [7] 杜瑞.5G背景下人工智能的发展分析[J].中国战略新兴产业(理论版),2019(18):1-1.
- [8] 张薇薇,孙文.浅析5G时代人工智能领域的发展[J].科技传播,2019(11).
- [9] 黄志雄,韦泉,黄少军,等.人工智能在基于SDN的5G传送网中应用的探讨[J].广西通信技术,2018(3):21-27.
- [10] 吴舟,王辉静,李坤颖.基于机器学习的5G系统智能公平调度算法研究[J].深圳信息职业技术学院学报,2018,16(5):37-41.
- [11] 刘召华.人工智能发展的新技术——深度学习[J].山西青年,2016(14):223.
- [12] 吕磊,李文彬.目标关键部位智能识别学习训练方法平台[J].国防科技,2019,40(5).

作者简介:

杨平,毕业于南京邮电大学,工程师,硕士,长期从事无线网的网络规划、设计及新技术研究等工作;裴霁,毕业于东南大学,工程师,学士,长期从事无线网的网络规划、优化工作。